



Exploring a meta-theoretical framework for dynamic assessment and intelligence

by

RAEGAN MURPHY

submitted in partial fulfilment of the requirements for the degree of

Philosophiae Doctor (PhD)

in the

**FACULTY OF HUMANITIES
DEPARTMENT OF PSYCHOLOGY**

at the

UNIVERSITY OF PRETORIA

SUPERVISOR: PROFESSOR D.J.F. MAREE

September 2007

Pretoria



From Dublin tenements and rented one-room housing; from second-rate citizens to beings making forays into meta-theory; two leaps have indeed been made! The second relying wholly on the first. I dedicate this thesis to George and Memory, my parents.



Acknowledgements

The author would like to extend her thanks to the following persons and institutions:

- Professor David Maree who has shown me that patience truly is a virtue
- George, Memory and Garrett Murphy
- J.H Owen and J. Pauw at the Department of Statistics, University of Pretoria, for their assistance with the meta-analysis
- The University of Pretoria for the bursary obtained which enabled me to continue with this study
- Dynamic assessment practitioners across the globe who kindly answered questions pertaining to this field of enquiry

I wonder whether the gentlemen who make a business and a living out of writing books, ever find their own selves getting in the way of their subjects, like me? Gabriel Betteredge from *The Moonstone* by Wilkie Collins (1868)

Summary

Dynamic assessment, as manner of alternative process-based assessment, is currently at a cross-roads chiefly characterised by, at times, vague conceptualisation of terminology, blurred demarcation as to its model and theory status and at times ill-defined fundamental philosophy. As a movement in modern psychological assessment within the broader field of intelligence, dynamic assessment does not present with a coherent unifying theory as such and due to its lack of clarity in a number of key areas its eventual disuse might well be the final outcome of this method and its unique history and methodology. In pursuit of this study's main goal, dynamic assessment models and theories are critically explored by means of a meta-theory largely inspired by the work K.B. Madsen, a Danish meta-theorist and pioneer in theoretical psychology. Madsen's meta-theory is attenuated in order to suit the nature and purposes of this study; so as to better analyse dynamic assessment within intelligence research and assessment.

In its primary aim, this study builds on a foundation of epistemological and ontological considerations within science in general, the social sciences and psychology in particular. In keeping with Madsen's method of meta-theory analysis, the author's predilections are stated at the outset in order to place the progression of analyses of the various models and theories within dynamic assessment. Dynamic assessment and intelligence are discussed and a brief digression into the history of Soviet psychology is offered as it is pertinent to the work of Lev Vygotsky and its subsequent influence within process-based assessment. Theory and model development within science and the social sciences are described from a philosophy-of-science vantage point. Psychological assessment's prime considerations are critically explored and the discussion highlights the role played by the philosophical aspects of mathematics and statistical foundations as leveraging measurement within assessment. Particular attention is paid to the perennial controversy surrounding null hypothesis significance testing and the possible future directions that can be explored by and within dynamic assessment which lends itself to approaches less restrictive than those offered by mainstream statistics.

The obvious and not so obvious aspects within the mathematical, statistical and measurement foundations are critically explored in terms of how best dynamic assessment can manoeuvre within the current mainstream psychological assessment system and how new models of item response theory suited to change-based assessment can be explored as possible manner of handling the gain score issue; itself a paradoxical state of affairs within classical and modern test theory. Dynamic assessment's past has in large part been dictated by mainstream considerations in the areas mentioned and in order to place itself on an alternative path these considerations are critically assessed in terms of dynamic assessment's future path. Dynamic assessment and its place within the broader intelligence assessment field is then investigated by means of the meta-theory developed. It is envisaged that the intuitive appeal of dynamic assessment will continue to garner support from practitioners across the globe, specifically those trained in countries outside the traditional stronghold of Western psychological theory. However, the position taken in this argument is that in order to ensure its survival it will need to make a decision in terms of its future progress: either to branch off from mainstream assessment altogether or to become fused within mainstream assessment. The "best of both worlds" scenario has obviously not worked out as it was originally hoped.

The study concludes with the meta-theoretical exploration of dynamic assessment within intelligence by utilising a small selection of current models. The application of the attenuated Madsenian framework seeks to explore, place and ascertain the nature of each model regarding the ontological and philosophical status of the approach; the nature of the hypothetical terminology, scientific hypotheses and hypothesis system utilised and lastly the nature of the abstract data, concrete data and prime considerations as implicit concerns within the varied approaches. An HQ score is calculated for each such model and is a partial indicator of the testability (verifiability or falsifiability) of the model in question. The models are thus couched in meta, hypothetical and data strata and can be positioned on a continuum of sorts according to which tentative claims can be made regarding the veracity of the approach behind each model.

The study concludes with two appendices; a meta-analysis which was conducted on South African research in the field of dynamic assessment (1961-2002) and which cumulated in a significant effect size evidencing an overall positive effect that dynamic assessment has had as an alternative intervention technique in comparison to conventional or static based assessment models. In order to encourage replication of this study, all details pertaining to the studies included for consideration in the meta-analyses are attached in section 2 of this appendix. Secondly, an informal content analysis was conducted on eleven responses to questionnaires that were originally delivered to one hundred dynamic assessment practitioners and researchers across the globe. The purpose of the questionnaire was to ascertain information on core issues within dynamic assessment, as these fundamental issues were considered as pivotal in the future of this approaches' eventual development or stagnation. The analysis concluded that dynamic assessment is indeed perceived to be at a crossroads of sorts and thus supported the initial hypothesis stated above.

It is hoped that this theoretical study will aid in aligning dynamic assessment in a manner such that its eventual place in psychological assessment will be solidly grounded, theoretically defensible and viable as alternative manner of assessment.



Keywords

Dynamic assessment, intelligence, epistemology, ontology, Soviet psychology, Russian psychology, Vygotsky, philosophy of science, theory, model, framework, meta-theory, psychology, mathematics, statistics, null hypothesis significance testing, measurement theory, item response theory, classical test theory, quantification, meta-analysis, content analysis.

Opsomming

Dinamiese evaluering, as 'n wyse van alternatiewe prosesgebaseerde evaluering, is huidiglik by 'n kruispad wat hoofsaaklik gekenmerk word deur vae beskrywing van terminologie, onduidelike grense ten opsigte van die stand van die model en teorie en soms ook sleg omskrewe grondbeginsels. Dinamiese evaluering, as 'n beweging binne die veld van sielkundige evaluering as deel van breër intelligensie, skiet te kort aan 'n samehangende verenigde teorie en as gevolg van swak gedefinieerde kernareas mag dit gebeur dat dinamiese evaluering, met sy geskiedenis en metodologie, in onbruik verval. Te einde die doel van hierdie studie te bereik, word dinamiese evalueringmodelle- en teorieë krities ondersoek deur gebruik te maak van meta-teorie na die voorbeeld van K. B. Madsen, 'n Deense meta-teoretikus en baanbreker in teoretiese sielkunde. Madsen se teorie is aangepas met die oog op die aard en doel van hierdie studie, naamlik om dinamiese evaluering, binne die veld van intelligensienavorsing en evaluering, beter te kan ontleed.

Met die hoofdoel in gedagte bou hierdie studie voort op die basis van epistemologiese en ontologiese beskouings binne die raamwerk van die Wetenskap en meer spesifiek, die Sosiale Wetenskap en sielkunde in besonder. Ten einde getrou te bly aan Madsen se metode van metateorie-analise word die outeur se voorkeure van meet af aan gestel sodat die vordering van ontleding van die verskillende modelle en teorieë binne dinamiese evaluering geplaas kan word. Dinamiese evaluering en intelligensie word bespreek, gevolg deur 'n kort uiteensetting van die geskiedenis van Sowjet sielkunde, die invloed daarvan op die werk van Lev Vygotsky asook die invloed van Vygotsky se werk op prosesgebaseerde evaluering. Teorie en model ontwikkeling in Wetenskap en Sosiale Wetenskap word beskryf vanuit 'n filosofie van wetenskap oogpunt. Die belangrikste beskouings van sielkundige evaluering word krities ondersoek. Die bespreking beklemtoon die rol wat die filosofiese aspekte van wiskundige en statistiese vertrekpunte speel en hoe dit meting in sielkundige evaluering beïnvloed. Spesifieke aandag word gegee aan die voordurende polemieë rondom nul hipotese beduidenis toetsing en moontlike toekomstige rigtings wat ondersoek kan word deur en in die veld van dinamiese evaluering wat minder beperkend as die hoofstroom statistiek is.

Die duidelike en die minder duidelike aspekte binne die wiskundige, statistiese en metings grondslae word krities ondersoek aan die hand van hoe dinamiese evaluering die beste binne die huidige hoofstroom sielkundige toetsing en meting metodes kan inpas asook hoe nuwe modelle van item-antwoord teorie geskik is vir veranderinggebaseerde evaluering ondersoek kan word as 'n moontlike wyse om die toeneems punt te hanteer; hierdie is opsigself 'n teenstrydige saak in gevestigte en moderne toetsteorie. Die verloop van dinamiese evaluering in die voorafgaande areas was grootliks voorgeskryf deur hoofstroom beskouinge wat in hierdie studie krities evalueer word met die doel om dinamiese evaluering op 'n alternatiewe pad in die toekoms te plaas. Dinamiese evaluering en die plek daarvan binne die groter veld van intelligensiemetings word dan ondersoek deur gebruik te maak van die ontwikkelde meta-teorie. Dit word verwag dat die intuïtiewe aanslag van dinamiese evaluering sal voortgaan om wêreldwyd steun te wen, veral in lande waar daar nie die tradisionele westerse beskouinge gehuldig word nie. Hoe dit ook al sy, die standpunt wat in hierdie studie gehuldig word is dat ten einde die voortbestaan van dinamiese evaluering te veseker, daar 'n besluit geneem sal moet word ten opsigte van die toekomstige verloop, hetsy enersyds 'n totale breuk met die hoofstroom toetsing en meting of andersyds, 'n totale samesmelting daarmee. Dit het duidelik geword dat 'die brood aan albei kante gebotter hê' opsie nie hier haalbaar was nie.

Die studie sluit af met 'n meta-teoretiese ondersoek van dinamiese evaluering in intelligensie deur gebruik te maak van 'n paar huidige modelle. Die aangepaste weergawe van Madsen se raamwerk het dit ten doel om die ontologiese en filosofiese aard van elke model te ondersoek, te plaas en vas te stel; die aard van die hipotetiese terminologie, wetenskaplike hipoteses en hipotese-stelsel in gebruik asook die aard van die abstrakte data, konkrete data en die belangrikste oorwegings as implisiete bekommernisse binne die verskillende benaderings. 'n HQ telling is bereken vir elke model en dien as 'n gedeeltelike aanwyser van die toetsbaarheid (bewysbaarheid en vervalsbaarheid) van die betrokke model. Die modelle word dus groepeer in meta, hipotetiese en data stratum en kan op 'n kontinuüm geplaas word waarvolgens tentatiewe eise gemaak kan word ten opsigte van die betroubaarheid van die benadering van elke model.

Die studie sluit verder af met twee aanhangsels; 'n meta-analise gedoen op Suid-Afrikaanse navorsing op die gebied van dinamiese evaluering (1961 - 2002) wat kulmineer het in 'n beduidende effek grootte wat aanduiding gee van die algemene positiewe effek wat dinamiese evaluering in vergelyking met konvensionele metingsmetodes gehad het. Om die herhaling van hierdie studie aan te moedig word alle besonderhede van die betrokke navorsing in die meta-analise aangeheg in afdeling twee van hierdie aanhangsel. Tweedens is 'n informele inhouds-analise uitgevoer deur vraelyste uit te stuur aan een honderd dinamiese evaluering praktisyns en navorsers wêreldwyd. Die doel van die vraelyste was om inligting te verkry wat handel oor die fundamentele sake rakende dinamiese evaluering aangesien dit 'n bepalende rol speel in of die oorlewing of die stagnasie van



dinamiese evaluering. Die gevolgtrekking van die analise het inderdaad bewys dat dinamiese evaluering by 'n kruispad staan en onderskryf dus die aanvanklike hipotese.

Daar word gehoop dat hierdie teoretiese studie sal bydra om dinamiese evaluering sy plek in sielkundige toetsing en meting te laat inneem op 'n goed gefundeerde, teoreties verdedigbare en lewensvatbare wyse as 'n alternatiewe vorm van evaluering.

Sleutelwoorde

Dinamiese evaluering, intelligensie, epistemologie, ontology, Sowjet sielkunde, Russiese sielkunde, Vygotsky, filosofie van die wetenskap, model, raamwerk, meta-teorie, sielkunde, wiskunde, statistiek, nul hipotese beduidenis toetsing, metingsteorie, item-antwoord teorie, gevestigte / klassieke toetsteorie, hoeveelheidsbepaling, meta-analise, inhoudanalise.



Table of contents

Acknowledgements	iii
Summary	iv
Opsomming	v
Chapter 1 Introduction and background to the study.....	1
1.1 Brief overview of the study.....	1
1.2 Motivation and rationale	1
1.3 Research orientation	2
1.4 Delineation and discussion of chapters to follow.....	2
1.5 An in-depth look at why such a study is necessitated.....	3
1.6 Pertinence to the South African context.....	4
1.7 Dynamic assessment's entanglement.....	5
1.8 The role of science	6
1.8.1 A personal note on philosophical affiliations	7
1.9 A map to the study.....	8
1.10 Conclusion	9
Chapter 2 Research method, philosophical affinities and dynamic assessment within intelligence	10
2.1 Introduction	10
2.2 Literature - prime source of information	10
2.2.1 Platforms searched	10
2.2.2 Questionnaire.....	10
2.3 Model utilised	11
2.3.1 The pervading meta-theoretical framework.....	11
2.4 Laying personal cards on the table: a deck of truth	11
2.4.1 Geographic and socio-political location.....	11
2.4.2 Time.....	11
2.4.3 Schooling within psychology.....	12
2.4.4 Personal experiences within psychological intelligence assessment.....	12
2.4.5 Philosophical leanings - implicit assumptions pervading this study	12
2.4.5.1 First axis - A Jungian typology	12
2.4.5.2 Second axis - A more traditional axis: an informal personal psycho-epistemological profile	13
2.4.5.3 Third axis - A complement	13
2.4.5.3.1 Introduction	14
2.4.5.3.2 The mind-body problem.....	14
2.4.5.3.3 Consciousness	17
2.4.5.3.4 G-dominated vs. multiple intelligence (MI) - dominated leanings towards the understanding of intelligence	18
2.4.5.3.5 Emergence (irreducibility) vs. reductionism.....	23
2.4.5.3.6 Realist vs. relativist approaches towards research.....	26
2.4.5.3.7 Perennial nature/nurture	31
2.4.5.3.8 Static vs. dynamic assessment of intelligence and potential.....	32
2.5 Summary.....	33
2.6 Dynamic assessment within intelligence.....	34
2.7 Dynamic assessment: fundamentals	34
2.7.1 History.....	48
2.7.1.1 Soviet ideology and Vygotsky's manoeuvrability within the system	55
2.7.2 Current trends	70
2.8 Intelligence.....	72
2.8.1 Desperately seeking a definition	73
2.8.2 Approaches towards the study of intelligence	78
2.8.3 Situating aspects of intelligence measurement	82
2.8.4 Omnipresent issues within approaches towards intelligence research	84



2.8.4.1 Levels of description.....	90
2.8.5 Yesteryear and today - how far have we come?.....	95
2.8.6 Cognition.....	97
2.8.7 Non-intellective factors	106
2.9 Summary.....	109
2.10 Conclusion	111
Chapter 3 Delineation of the meta-theoretical framework	112
3.1 Introduction	112
3.2 Science	112
3.2.1 Explanatory mechanisms in science.....	115
3.2.2 Scientific theories	117
3.3 Social science	119
3.3.1 Explanatory mechanisms in social science.....	127
3.4 Psychology.....	129
3.4.1 Explanatory mechanisms in psychology	131
3.4.2 Psychological theories.....	136
3.4.2.1 Concepts, definitions and propositions	138
3.5 Theory appraisal	142
3.6 Towards a meta-theory.....	152
3.7 Summary.....	156
3.8 K.B. Madsen.....	156
3.9 Madsen's meta-science and meta-theory	158
3.9.1 Madsen's definition of science	159
3.9.2 Madsen's definition of meta-science.....	159
3.9.3 Madsen's view on theory	161
3.9.4 Madsen's view on models.....	161
3.10 Madsen's overarching framework: Systematology - a comparative meta-theoretical taxonomy.....	163
3.10.1 The meta-stratum	165
3.10.2 The hypothetical-stratum	171
3.10.3 The data stratum	176
3.11 Considering the whole text	177
3.11.1 The testability of theories - hypothesis quotient (HQ).....	178
3.12 An amalgamated Madsenian framework	181
3.13 Conclusion	182
Chapter 4 The fundamentals of assessment: prime considerations	184
4.1 Introduction	184
4.1.1 The quantitative imperative.....	184
4.2 Mathematical foundation	193
4.2.1 The philosophical implications of mathematics	196
4.2.2 The link made manifest	203
4.2.3 Summary.....	205
4.3 Statistical foundation	206
4.3.1 Statistical issues.....	207
4.3.1.1 Historical contingency.....	209
4.3.1.2 The hegemony of null-hypothesis significance testing.....	216
4.3.2 Summary.....	226
4.4 Measurement foundation.....	227
4.4.1 Elements of measurement.....	229
4.4.1.1 Extensive and intensive measures.....	232
4.4.1.2 Rules of concatenation - the necessary additive structure.....	233
4.4.1.3 Conjoint measurement as fundamental extensive.....	234
4.4.2 Test theory	240
4.4.2.1 Representational measurement: the legacy of Stevens' arbitrary powerful scales of measurement	242



4.4.2.2 Classical test theory - a goal for inference.....	243
4.4.2.2.1 Malleable stability - the gain score issue.....	248
4.4.2.2.2 Generalizability theory.....	249
4.4.2.3 Modern test theory.....	250
4.4.2.3.1 Modern test theory change models - an answer for dynamic assessment's change score problem.....	254
4.4.3 Summary.....	265
4.5 Integrating Madsen - attenuating the framework.....	267
4.6 Conclusion.....	270
Chapter 5 Dynamic assessment and intelligence: exploring a meta-theoretical framework.....	273
5.1 Introduction.....	273
5.1.1 Exploring the framework for dynamic assessment and intelligence: the eight considerations.....	274
5.2 Dynamic assessment and learning potential models conceptualised within the framework.....	275
5.2.1 The Leipzig learning test (LLT) and the adaptive computer assisted intelligence learning test battery (ACIL): diagnostic programmes.....	276
5.2.1.1 A(i) Ontology.....	277
5.2.1.2 A(ii) Philosophy.....	277
5.2.1.3 B(i) Hypothetical terms.....	278
5.2.1.4 B(ii) Scientific hypotheses.....	278
5.2.1.5 B(iii) Hypothesis system.....	280
5.2.1.6 C(i) Abstract data.....	280
5.2.1.7 C(ii) Concrete data.....	280
5.2.1.8 C(iii) Prime considerations.....	280
5.2.2 Assessing the learning potential for inductive reasoning (LIR) in young children.....	283
5.2.2.1 A(i) Ontology.....	285
5.2.2.2 A(ii) Philosophy.....	286
5.2.2.3 B(i) Hypothetical terms.....	286
5.2.2.4 B(ii) Scientific hypotheses.....	286
5.2.2.5 B(iii) Hypothesis system.....	287
5.2.2.6 C(i) Abstract data.....	288
5.2.2.7 C(ii) Concrete data.....	288
5.2.2.8 C(iii) Prime considerations.....	288
5.2.3 Dynamic assessment for students with learning disabilities.....	288
5.2.3.1 A(i) Ontology.....	290
5.2.3.2 A(ii) Philosophy.....	290
5.2.3.3 B(i) Hypothetical terms.....	291
5.2.3.4 B(ii) Scientific hypotheses.....	291
5.2.3.5 B(iii) Hypothesis system.....	292
5.2.3.6 C(i) Abstract data.....	292
5.2.3.7 C(ii) Concrete data.....	292
5.2.3.8 C(iii) Prime considerations.....	292
5.2.4 A model for dynamically assessing intelligence: differences in feedback.....	293
5.2.4.1 A(i) Ontology.....	294
5.2.4.2 A(ii) Philosophy.....	297
5.2.4.3 B(i) Hypothetical terms.....	298
5.2.4.4 B(ii) Scientific hypotheses.....	299
5.2.4.5 B(iii) Hypothesis system.....	300
5.2.4.6 C(i) Abstract data.....	300
5.2.4.7 C(ii) Concrete data.....	300
5.2.4.8 C(iii) Prime considerations.....	301
5.2.5 Dynamic assessment of the level of internalisation of elementary school children's problem-solving activity.....	301
5.2.5.1 A(i) Ontology.....	302
5.2.5.2 A(ii) Philosophy.....	302
5.2.5.3 B(i) Hypothetical terms.....	305



5.2.5.4 B(ii) Scientific hypotheses.....	305
5.2.5.5 B(iii) Hypothesis system	306
5.2.5.6 C(i) Abstract data.....	306
5.2.5.7 C(ii) Concrete data	306
5.2.5.8 C(iii) Prime considerations.....	307
5.2.6 The learning potential test for ethnic minorities (LEM)	307
5.2.6.1 A(i) Ontology	308
5.2.6.2 A(ii) Philosophy	308
5.2.6.3 B(i) Hypothetical terms	308
5.2.6.4 B(ii) Scientific hypotheses.....	308
5.2.6.5 B(iii) Hypothesis system	309
5.2.6.6 C(i) Abstract data.....	309
5.2.6.7 C(ii) Concrete data	309
5.2.6.8 C(iii) Prime considerations.....	309
5.2.7 Swanson-cognitive processing test (S-CPT).....	312
5.2.7.1 A(i) Ontology	314
5.2.7.2 A(ii) Philosophy	314
5.2.7.3 B(i) Hypothetical terms	314
5.2.7.4 B(ii) Scientific hypotheses.....	315
5.2.7.5 B(iii) Hypothesis system	315
5.2.7.6 C(i) Abstract data.....	316
5.2.7.7 C(ii) Concrete data	316
5.2.7.8 C(iii) Prime considerations.....	316
5.2.8 Feuerstein's Learning propensity assessment device (LPAD).....	317
5.2.8.1 A(i) Ontology	318
5.2.8.2 A(ii) Philosophy	319
5.2.8.3 B(i) Hypothetical terms	321
5.2.8.4 B(ii) Scientific hypotheses.....	322
5.2.8.5 B(iii) Hypothesis system	325
5.2.8.6 C(i) Abstract data.....	326
5.2.8.7 C(ii) Concrete data	326
5.2.8.8 C(iii) Prime considerations.....	327
5.2.9 The assessment of learning potential: the EPA instrument	327
5.2.9.1 A(i) Ontology	328
5.2.9.2 A(ii) Philosophy	328
5.2.9.3 B(i) Hypothetical terms	328
5.2.9.4 B(ii) Scientific hypotheses.....	328
5.2.9.5 B(iii) Hypothesis system	329
5.2.9.6 C(i) Abstract data.....	329
5.2.9.7 C(ii) Concrete data	330
5.2.9.8 C(iii) Prime considerations.....	331
5.2.10 Application of cognitive functions scale (ACFS).....	333
5.2.10.1 A(i) Ontology	334
5.2.10.2 A(ii) Philosophy	335
5.2.10.3 B(i) Hypothetical terms	336
5.2.10.4 B(ii) Scientific hypotheses.....	337
5.2.10.5 B(iii) Hypothesis system	338
5.2.10.6 C(i) Abstract data.....	338
5.2.10.7 C(ii) Concrete data	338
5.2.10.8 C(iii) Prime considerations.....	339
5.2.11 Analogical reasoning learning test (ARLT).....	342
5.2.11.1 A(i) Ontology	343
5.2.11.2 A(ii) Philosophy	343



5.2.11.3 B(i) Hypothetical terms	343
5.2.11.4 B(ii) Scientific hypotheses.....	343
5.2.11.5 B(iii) Hypothesis system	344
5.2.11.6 C(i) Abstract data.....	345
5.2.11.7 C(ii) Concrete data	345
5.2.11.8 C(iii) Prime considerations.....	345
5.2.12 The Mindladder model: using dynamic assessment to help students learn to assemble and use knowledge.....	347
5.2.12.1 A(i) Ontology	349
5.2.12.2 A(ii) Philosophy	350
5.2.12.3 B(i) Hypothetical terms	350
5.2.12.4 B(ii) Scientific hypotheses.....	350
5.2.12.5 B(iii) Hypothesis system	351
5.2.12.6 C(i) Abstract data.....	352
5.2.12.7 C(ii) Concrete data	352
5.2.12.8 C(iii) Prime considerations.....	352
5.2.13 The cognitive modifiability battery (CMB).....	352
5.2.13.1 A(i) Ontology	353
5.2.13.2 A(ii) Philosophy	353
5.2.13.3 B(i) Hypothetical terms	353
5.2.13.4 B(ii) Scientific hypotheses.....	354
5.2.13.5 B(iii) Hypothesis system	355
5.2.13.6 C(i) Abstract data.....	355
5.2.13.7 C(ii) Concrete data	356
5.2.13.8 C(iii) Prime considerations.....	356
5.3 Summary of HQ results	357
5.4 Conclusion	359
Chapter 6 Conclusion	361
6.1 Introduction: the value of the study in terms of its coverage	361
6.2 Fundamentals	361
6.3 Theory, meta-theory, science, the social sciences, psychology and assessment	362
6.4 Prime considerations: quantification as imperative	363
6.5 The exploration of dynamic assessment within a meta-theoretical framework.....	364
6.6 Limitations of the study.....	365
Appendix 1 Meta-analysis of South African dynamic assessment research.....	367
Section A Meta-analysis.....	367
1. Introduction	367
1.1 The need for a meta-analysis of South African dynamic assessment.....	368
1.2 A brief tour of meta-analysis	369
2. Method	370
2.1 Two-fold purpose of the meta-analysis	370
2.2. "META - Easy to answer" version III by D. A. Kenny.....	371
2.2.1 Programme structure.....	371
2.3 Meta-analysis programme version 5.3 by R. Schwarzer	374
2.3.1 Programme structure.....	374
2.4 Comparison of the two programmes	375
2.5 Limitations of the programmes and violation of assumptions	376
3. Results	376
3.1 Criteria used for inclusion of studies	376
3.2 First run with outliers included (25 effect sizes).....	378
3.2.1 Results using META by Kenny	378
3.2.2 Results using meta-analysis by Schwarzer.....	380
3.3 Second run with outliers excluded (22 effect sizes).....	381
3.3.1 Results using META by Kenny	381



3.3.2 Results using meta-analysis by Schwarzer.....	383
4. Discussion.....	386
4.1 Kenny's programme	386
4.2 Schwarzer's programme.....	386
4.3 A search for moderators	387
4.4 Limitations of the meta-analysis.....	388
4.5 Implications of the findings for dynamic assessment research in South Africa	388
4.6 Recommendations	389
5. Conclusion	389
6. Excursion into a potentially rich field of investigation - an aside to this meta-analysis	389
6.1 From GeneWays to MemeWays	389
Section B Further information regarding studies used in the meta analysis.....	392
1. Introduction	392
Appendix 2 Content analysis of returned questionnaires	417
1. Introduction	417
1.1 Questionnaire.....	417
1.2. Brief introduction to content analysis	422
1.2.1 Reliability and validity	424
1.2.2 Qualitative vs. quantitative content analysis.....	424
1.2.3 Manifest and latent meaning.....	425
2. Method.....	425
2.1 Sample.....	425
2.2 Procedure	425
2.3.1 Analysis.....	425
2.3.1.1 Coding the content	425
2.3.1.2 Process	425
2.3.1.3 Coding frame.....	426
3. Results	427
4. Discussion.....	434
4.1 What was not mentioned - a further source of information.....	435
5. Conclusion	436
References	437

List of figures

Figure 1 The analogous co-ordinate map illustrating chapter placements.....	9
Figure 2 Relations between various fields of knowing and validity criteria involved (Ettin, 1984, p.217).....	17
Figure 3 Dynamic assessment placed in a two dimensional framework of complexity of task and time limit (Neřka & Orzechowski, 2005, p.124).....	25
Figure 4 Relativism and realism between and within worlds	30
Figure 5 The inseparability of IQ and learning potential and the iron-clad grip of IQ upon dynamic constructs	40
Figure 6 Dynamic assessment and the understanding of development.....	47
Figure 7 Intelligence-dynamic assessment link: the bio-ecological intelligence model and the Feuersteinian dynamic assessment model (bio-ecological model from Ceci, Rosenblum, De Bruyn & Lee, 1997, p.314).....	50
Figure 8 Linkages between schools of thought with emphasis on Soviet systems (Madsen, 1988, p.444)	63
Figure 9 Tharp and Gallimore's (1988, p.35) depiction of Vygotsky's zone of proximal development	66
Figure 10 A possible alignment of IQ within the continuum of successful adaptation as differentially measured	74
Figure 11 Lay and expert opinions about intelligence (Gregory, 2004, p.142; Sternberg, 2004c, 2004e). Concerns about interpretations	76
Figure 12 Greenspan and Driscoll's (1997, p.133) content model of personal competence with attending aspects of importance to dynamic assessment.....	77
Figure 13 Paradigms within intelligence explanation (Li, 1996) with problems overlaid.....	79
Figure 14 Intelligence within two realms	80
Figure 15 Different approaches towards the study of a single issue.....	81
Figure 16 Anderson's (1999, p.5) terrain of intelligence in which three principle axes are delineated according to which placement of various forms of intelligence are situated.....	83
Figure 17 The puzzling increase in IQ and the resultant implications for construct validation.....	87
Figure 18 Jensen's (1982b, p.270) depiction of g's relations to tests, processes and genotypes in the reaction-time paradigm .	89
Figure 19 Eysencks' (1986) model of the relation between biological, IQ (psychometric) and social intelligence	91
Figure 20 From the specificities to the generalities of intelligence assessment - where dynamic assessment can invent its place	94
Figure 21 Loci of intelligence (Sternberg, 1988, pp.4-5).....	96
Figure 22 The cognitive hexagon (Pylyshyn, 1983, p.76): cognitive science's boundary transcendence.....	99
Figure 23 Three conceptions of computer-mind analogies (Fetzer, 2002, pp.xiii-xvii)	101
Figure 24 Simon and Wallach's (1999 in Strube, 2000) criteria of empirical adequacy for correspondence expanded upon by the depiction of the role of bridging efforts.....	103
Figure 25 Neural and mental structures in the sub symbolic paradigm (Haugeland 1997 in Harnish, 2002, p.347).	104
Figure 26 Heritability, intelligence and its calculations (Grigorenko, 2004a).....	107
Figure 27 (i) Placement of psychology as discipline in between varying poles of methodology and areas of concern (Royce, 1973, p.16).....	114
Figure 27 (ii) Trends in psychology with placement of dynamic assessment along axes of epistemology and methodology (Watanabe, 2003, p.131).....	115
Figure 28 Relations among scientific disciplines, (McErlean, 2000).....	121
Figure 29 Permeating tension: how an initial higher-level problem becomes manifest at lower levels	123
Figure 30 The nature of scientific investigation (Coombs, Dawes & Tversky, 1970, p.3).....	132
Figure 31 Utilising Coombs, Dawes and Tversky (1970, p.3) and overlaying central tenets pertinent to the study	133
Figure 32 Cattell's (1988b) depiction of the basis of distinction between theoretical concepts and empirical constructs.....	135
Figure 33 Concept-empirical indicator link via operational definition with example (Fawcett & Downs, 1992, p.28).....	141
Figure 34 Relation between intra and inter-theory correspondence rules and bridge principles	141
Figure 35 The basic difference between inductively and deductively derived theory (Oldroyd, 1986, p.62)	144
Figure 36 Cattell's (1988a, p.17) inductive-hypothetico-deductive spiral.....	144
Figure 37a Continuum of theoretical power (Royce, 1978, p.262).....	147
Figure 37b Continuum of metatheoretical reification (Royce, 1978, p.262)	148
Figure 38 Continuum of metatheoretical reification with additions by Royce (1978).....	148
Figure 39 Dependency of theoretical power upon conceptual-linguistic precision (Royce, 1978, p.268).....	148
Figure 40 Elements in an orienting strategy (Berger & Zelditch, 1993, p.18).....	155

Figure 41 Illustration of ordinary science discourse with examples	160
Figure 42 Illustration of meta-science discourse with examples	160
Figure 43 Madsen's overarching scheme. From Madsen (1971), p.7; (1973), p.698; (1976), p.301; (1984), p.201; (1985), p.3 ; (1987a), p167; (1987b), p268; (1988), p.29; (1992), p504.	164
Figure 44 Madsen's (1988) conception of psycho-physical theory within the ontological meta-theses as part of the meta-stratum	166
Figure 45 (a) Madsen's (1988) meta-stratum (A) Ontological meta-theses	169
Figure 45 (b) Madsen's (1988) meta-stratum (B) Philosophical meta-theses	170
Figure 46 Schematic representation of model functioning, after Madsen, 1988.....	174
Figure 47 Madsen's (1988) hypothetical-stratum	175
Figure 48 Madsen's (1988) data-stratum	177
Figure 49 Diagrammatic illustration of the interrelations between various meta-scientific disciplines	177
Figure 50 Formal analytical view of theory constituents (Ross, 1964)	181
Figure 51 Assessment predicates.....	186
Figure 52 The tension of two realms.....	187
Figure 53 Empirical construct validation and meaning-ladenness of true construct.....	189
Figure 54 The vacillation between theory, measurement and construct validation (adapted from Oberhauer, 2005, p.394)	190
Figure 55 Measurement theory and theory of positive real numbers (Ross, 1964, p.60)	192
Figure 56 Barrett's (2000, p.41) utilisation of Wittman's (1997) illustration	192
Figure 57 Three realms of shared concern; the base and pillars of which may not be as entirely pure as seemingly implied in the upheld pinnacle.....	197
Figure 58 The relation between various worlds and how each world takes a small part of its predecessor with it into the next world with attenuations as to the mathematical foundations (Penrose, 1994, p.414).....	201
Figure 59 The relation between the three worlds utilising IQ as reified yet ill-defined measurable construct.....	202
Figure 60 The theoretical significant difference between two groups either receiving or not receiving mediatory interventions couched within NHST technique (attenuated after Loftus, 1996, p.165).....	221
Figure 61 Information provided by confidence intervals which are not provided by point estimate statistical tests (Reichardt & Gollub, 1997, p.273).....	224
Figure 62 Attitude measures are not geometrical measures and even if they were they are not quantitative by default.....	236
Figure 63 The discipline today and in the future: a possible scenario for maintaining both mainstream and dynamic assessment methodologies.....	241
Figure 64 Rationale behind measurement: ill-conceived notions of quantifiable constructs.....	243
Figure 65 Change of performance in terms of CTT (Wiedl, 2002, 2003)	245
Figure 66 Erroneous equating of true score and construct.....	245
Figure 67 The integration of dynamic assessment into the testing process: Guthke's ADAFI (Wiedl, 2003, p.96)	246
Figure 68 Kormann and Sporer's (1983) overview of change measurement as viewed both classically and alternatively.....	255
Figure 69 Effective ability as consequence of initial and modified ability within the Rasch model (Embretson, 1991b, p.189) ..	259
Figure 70 Relations between fundamental measurement, CTT and IRT(Borsboom, 2005, p.129)	265
Figure 71 The attenuated Madsenian systematology framework	269
Appendix 1 Figure 1 Range of effect size across study number	387
Appendix 2 Figure 1 View of dynamic assessment along a continuum of increasing conceptualisation.....	428

List of tables

Table 1 Assessment and teaching within dynamic assessment: commonalities and differences (Brown, Campione, Webber & McGilly, 1993, p.175).....	44
Table 2 Main features differentiating dynamic assessment from static assessment (Morrison, 2001 p.308; Tzuriel, 2001, p.7)..	45
Table 3 Playing devil's advocate: some not so sinister aspects of static assessment and some problematic issues within dynamic assessment.....	46
Table 4 Placement of dynamic assessment precursors within the broader context of Russian psychological history (Sahakian, 1975)	57
Table 5 Conventional hierarchical arrangement of disciplines, in which physics is considered the ultimate as reductionist explanatory mechanism (McErlean 2000).....	120
Table 6 The relation of concept to theory (Gerstlé, 1989)	138
Table 7 Typology of propositions (Fawcett & Downs, 1992).....	139
Table 8 Classification of the three sub-systems within which theories are produced (Madsen, 1987b, 1988).....	158
Table 9 Level of discourse according to classification scheme from Madsen (1988)	161
Table 10 Framework for model abstraction and function.....	163
Table 12 Main difference between Fisher's P-value based approach and Neyman-Pearson's fixed alpha approach (Huberty, 1993, p.318; Kline, 2004)	214
Table 13 Alternative forms of NHST	215
Table 14 Relation between substantive and statistical hypotheses and how sample size can impact on conclusions (Johnson, 1999; Nester, 1996)	219
Table 15 Common misconceptions pertaining to NHST.....	223
Table 16 Justification of the reasoning pertaining to NHST table 15	224
Table 17 Comparison of schools of statistical inference (Oakes, 1986)	226
Table 18 Structure change dependent on modified ability through successive occasions	257
Table 19 An anatomy of inference (from Wright, 1997, 1999).....	264
Table 20 Highlighted issues within the selected models	276
Appendix 1 Table 1 Schwarzer programme menu options (Schwarzer, 1989)	374
Appendix 1 Table 2 Utilities menu in Schwarzer's programme.....	375
Appendix 1 Table 3 Study numbers and sample classification.....	377
Appendix 1 Table 4 Brief description of the seven studies included in the meta-analysis	378
Appendix 1 Table 5 Meta analysis results using Kenny's programme	378
Appendix 1 Table 6 META data file using Kenny's programme	379
Appendix 1 Table 7 Meta-analysis results using Schwarzer's programme	380
Appendix 1 Table 8 Data file from Schwarzer's programme.....	381
Appendix 1 Table 9 Meta-analysis results using Kenny's programme.....	382
Appendix 1 Table 10 META Data file using Kenny's programme	383
Appendix 1 Table 11 Meta-analysis results using Schwarzer's programme	384
Appendix 1 Table 12 Data file using Schwarzer's programme	385
Appendix 1 Table 13 Effect size cluster analysis	388
Appendix 2 Table 1 Example of selective verbatim response text coding for dichotomous text (question 1, 2, nature vs. nurture debate) and thematic text (question 3)	423
Appendix 2 Table 2 Dynamic assessment's status as scheme, model or theory	428
Appendix 2 Table 3 Dynamic assessment philosophy	429
Appendix 2 Table 4 The role played by intelligence within dynamic assessment procedures.....	431
Appendix 2 Table 5 The role of neuroscience within dynamic assessment	432
Appendix 2 Table 6 Dynamic assessment's future existence.....	433
Appendix 2 Table 7 The role of quantification and measurement within dynamic assessment	434
Appendix 2 Table 8 The role of meta-theory within dynamic assessment	434

Chapter 1 Introduction and background to the study

1.1 Brief overview of the study

The aim of this theoretical study is to construct a meta-theoretical framework for the field of dynamic assessment, as this field is currently experiencing some tension regarding fundamental aspects surrounding theory and practice. The means by which this will be achieved will be based on critical analyses and evaluations of various theories and models developed within the field. In order to develop such an envisaged meta-theory,¹ the historical significance of various trends, models and theories will be assessed in terms of their respective contributions to the current status of dynamic assessment within the broader context of intelligence research.

Dynamic assessment is a manner of assessing individuals in a fluid, process-orientated and flexible way and has much to offer the field of intellectual assessment due to its unique stance on measurement. The origins of dynamic assessment are situated in areas of research in which practitioners were originally given mandates to assess prospective learners within specific cultural contexts. Concepts, ideals and notions associated with this type of assessment were often translated into quantifiable measurements in order to satisfy the reigning paradigm of testing. More often than not, dynamic assessment ideals did not filter through to later editions and modifications of tests and the goal of many early intelligence tests was to categorise individuals based on their performance within the testing situation. Partly owing to industrialisation and expansion, testing of groups of individuals was deemed a more expedient alternative to assessing on an individual basis, thus allowing for more people to be tested within a shorter time span.

Unfair discrimination and biased use of many intelligence tests gradually led to the growing dissatisfaction with these static and product-bound instruments and in an attempt to find a suitable form of assessment, dynamic assessment has offered partial solutions to the problem-riddled field of intelligence assessment. However, "partial solutions" is perhaps currently the most apt description of the status of this field and what it can presently offer. The beginnings of dynamic assessment are widely spread in terms of geographic origins, disparate in terms of initial reasons as to its use and culturally diverse in the contexts within which it works. How has the origin and development of dynamic assessment compared to mainstream intelligence research in terms of theoretical contribution, practical enhancement of current tests, reduction of bias and generalisability? Does the field of dynamic assessment lack a coherent and sound theoretical base? As with most theories within the realm of science, as it is practised both within the social and natural milieus, tenuous foundations and fuzzy concepts hamper the progression of ideas and theories. What is the theory and framework behind dynamic assessment? Ironically, one aspect of dynamic assessment research for which consensus has indeed been reached is that there is as yet no one particular framework in which this form of assessment can be adequately housed.

Multifarious as the field of dynamic assessment may at first glance appear, it may be possible to connect various trends and researchers within this field by investigating the underlying historical influences, basic premises upon which theories are based and contemporaneous personalities within intelligence assessment. In attempting to link, compare and integrate various dynamic assessment trends, a more in-depth understanding of what dynamic assessment is, how it came to be and the nature of its future course of progression can be sought. Seeking coherence and consistency in a field as disparate as dynamic assessment may well prove fruitful in its future endeavours in the intelligence arena and its subsequent acceptance into larger academic circles within psychology.

1.2 Motivation and rationale

In order to advance within the field of dynamic assessment (in terms of this form of psychology being more generally accepted), basic issues have to be readdressed. Many of these fundamental issues are considered as accepted, established and standard sub-areas within the discipline, some such areas being measurement as well as Western ideas and notions of what it is to study psychology and conduct psychological assessment. The very notion of what it means to measure is confounded by institutionalised acceptance of what is considered measurable and what is not. Of course, it is not only dynamic assessment which labours against these notions, but also traditionally accepted forms of assessment which, without critical thought for the philosophy of measurement, have taken ever greater strides in theory and model development without first having addressed basic issues. If fundamental and core issues can at least be addressed then perhaps a sounder footing can be established in which all forms of assessment can take root, in particular, dynamic assessment. This study attempts to disentangle current preoccupations with certain issues in the field and to objectify, explore and more fully understand why it is that this field is experiencing problems in terms of greater acceptance within the broader assessment profession. The statement by T.J.

¹ "Meta-theory" and "metatheory" are used interchangeably throughout the literature, although the majority of writers utilise the latter spelling.



Watson's words "first look backward in order to look forward" (in Williams, 1997) is a sentiment which is upheld throughout this study.

Overarching, all-encompassing, fully inclusive, bridged meta-theories may be pie in the sky in terms of the veracity of such models (and what they purport to lend to the ever-burgeoning field of assessment) and what in fact can be practically offered by such models to the general practitioner within psychological assessment. Fragmented data,² seemingly incompatible foundations, blurred epistemological axioms, cross-pollinated methodological measures, the growing magnitude of perceived novel contributions, the increasing morass of incoherent theories and lack of containment within just one field of psychology, namely intelligence assessment, seems to overpower the individual researcher and practitioner to the extent that as long as assessment can deliver the "goods" within the ambit of moral and ethical codes, a practice can survive long enough to benefit those who assume that the field is well entrenched in consistency and stability. Perhaps it is just as well that those who seek to benefit from assessment remain blissful in their ignorance of what in fact emanates from within the perceived "ever-progressing" encampment of solid psychological science. In order to shed some much-needed light and to perhaps aid in the retention of such a perceived situation within psychological intelligence assessment a study such as this, it is hoped, may prod the more sensitive among researchers and practitioners in a direction which may hold the fruits of a venture yet to begin.

1.3 Research orientation

Although not intended as a philosophical treatise on dynamic assessment, the philosophical imperatives according to which studies such as this are conducted should be delineated at the outset. The "lens" through which literature is read, understood and integrated within a greater context of understanding, influences (albeit to a limited extent) the manner in which information is comprehended. As should be mentioned at the start of this discourse, the author in no way purports to be a philosopher and being mindful of the expansive literature and history concerned with philosophy regrets the simplicity with which certain basic philosophical issues are discussed and addressed as pertains to dynamic assessment and the greater purview of intelligence research. The specific philosophical stances will be discussed and expanded upon later in Chapter 3.

At this juncture it may be prudent to expand somewhat on the intended stance or view from which the literary landscape will be surveyed. The author has leaned very heavily on the concepts, ideas and models used within the works of one K.B. Madsen of Danish origin who himself has been influenced by a host of Scandinavian researchers (1968) and who, among others, co-founded the fundamentals of a theoretically attuned psychology (Baker, Hyland, Rappard & Staats, 1987). He was most notably influenced by Sigmund Koch who pioneered comparative studies of scientific theories within psychology (Koch, 1959; Madsen, 1987). It seems almost fitting that the foundations upon which this study rests emanate from geographically and philosophically disparate areas, much akin to the very disparities which make up the history and progression of dynamic assessment. Mainstream psychological theory and practice (if such a notion in fact exists) offers much by way of origins and development but it is noteworthy that Western ways of practicing what is considered a psychology has helped fuel the coalescence of non-Western practises. Grandiose models and schemas that encompass what may seem to be almost everything will never quite see the light of day - reality in all its complexities simply will not facilitate such visionary philosophical ponderings. However, much in the same vein as Joseph Royce (1973; 1987), perhaps it is more felicitous to build "minitheories" or microtheories (Brand, 1997) which represents a tentative start to a life-long endeavour (which incidentally occupied a considerable amount of Madsen's professional career).³ Madsen's methodology and choice of locating theory within the grander culmination of multitudinous strata offers a guide towards systematising the overflowing outpouring of data and method. That this type of theorising is not to everyone's liking is an unavoidable fact of life within academic research and no excuse is made or proffered in terms of warranting receipt of sympathy that might be elicited upon such an apology. The method is unashamedly grandiose. The conclusions are open to controversy at worst and debate at best.

1.4 Delineation and discussion of chapters to follow

The nature of this theoretical discussion warrants an extensive overview of the history, development, status and future of intelligence research and more specifically dynamic assessment. In order to accomplish this task a critical and evaluative rendering of texts is of prime importance. *This chapter* serves as basic introduction to the study that is to follow. *Chapter 2* discusses the research method involved as well as the fundamental philosophical affiliations from which this study can be viewed. It includes a discussion on dynamic assessment fundamentals as well as dynamic assessment history along with a detour into the reigning Soviet ideology and how this played into the works and theories of Lev Vygotsky and concludes with a brief look at intelligence assessment. *Chapter 3* entreats the reader to carefully study the philosophical underpinnings of

² On many an occasion it has been lamented that the addition of yet more data to a field already bursting at the seams is often construed as data gathering at the expense of pursuing theory, data being construed as a means towards an end; a state of affairs which could do with some measure of synthesis (Howard, 1998; Royce, 1987).

³ Perhaps this might contextualise the nature of such a task (but not in any way excuse its obvious limitations and flaws).



psychological science as it is encompassed within social science research and the broader field of scientific knowledge acquisition. Also included is the introduction and discussion on the meta-theoretical framework utilised according to which theories and models of dynamic assessment within intelligence are to be studied. This chapter also discusses the influence of historicity, socio-politics, concepts, theories, schemata and models within social science research. *Chapter 4* addresses the fundamentals of psychological assessment and includes sections on the mathematical, statistical and measurement foundations and concludes with how these foundations form and are informed by intelligence assessment specifically dynamic assessment and details the manner in which these concerns fit into the larger attenuated Madesian framework. *Chapter 5* revolves around the comparison of dynamic assessment theories and models within intelligence assessment and seeks to explore the developed meta-theoretical framework for dynamic assessment and intelligence. *Chapter 6* concludes the study with summaries, conclusions and recommendations. Two appendices are attached which include Appendix 1 section A and section B as well as Appendix 2. *Appendix 1 section A* concerns itself with a meta-analysis conducted on South African research within dynamic assessment. *Section B* details the extensive codings used within the meta-analysis and is included for any future replication studies which might be conducted on the studies. *Appendix 2* discusses the content analysis conducted on completed questionnaires sent out to dynamic assessment practitioners across the globe.

1.5 An in-depth look at why such a study is necessitated

At the outset it must be stated that this treatise may be considered an affront to much of what is currently routinely accepted as being "true" and the intent of this study is not to uncritically attack any point of view. It is in the hope of scientific and scholarly discourse and debate that the author does not fall into the trap of being critical but rather engaging in creative attempts to offer partial solutions to the questions being asked (Horn, 1979). The author does not seek to create a "theoretical superstructure" (Purves, 1997) but merely a tentative exploratory framework. Dynamic assessment research currently exists within a framework informed by the history of intelligence⁴ research but has as its main driving force theories and ideas⁵ emanating from Vygotsky in the 1920's and notions of assessing potential derived largely from findings of immigrant childrens' intelligence results on traditional (static) test batteries in the 1950's. However, there are other theories and ideas that have contributed to the development of dynamic assessment and these theories will in turn be evaluated in terms of contributions to the field.⁶ The *aspect or research problem* that is to be studied in this conceptual and theoretical study is the contextual and theoretical framework within which dynamic assessment is currently housed⁷ or the edifice upon which it is based and how it has been influenced by intelligence research within psychology. In an attempt to more fully understand this foundation and improve upon it, the construction of a newly revised framework will be attempted to further solidify and inform the progress of dynamic assessment theory and in turn practical assessment research.

The context of the social reality (phenomenon or problem area) constitutes the reality of intelligence research. This realm is further refined by a novel approach to intelligence research, namely dynamic assessment research. Intelligence assessment

⁴ Although Binet's thoughts on intelligence (encompassing the notion of assessing whilst learning) is currently common knowledge within assessment, it is still of interest to note that his original theoretical ideas were "nowhere to be found" in the translation by Goddard when it was brought over to the United States (Resing, 2001). Only in 1973 with the comprehensive treatment of Binet's work extolled by Wolf (1973, in Sarason, 1976) did his work and basic philosophy become entrenched in assessment theory (Sarason, 1976). It is of interest here precisely due to the loss of this basic theoretical underpinning. Why his theoretical notion was not carried through remains speculative, especially since the Stanford-Binet has had such an overwhelming influence on test development especially in the United States with particular influence on the Wechsler-Bellevue Scale (Boake, 2002). One can only surmise what the present-day scenario would have been regarding the understanding and measurement of intelligence, had his views "filtered" through. Also of interest is the early continental European influence within dynamic assessment, as Binet was French. Later, Russian ideas on child development were to reveal similar sentiments in terms of assessing the child whilst engaged in the learning process. Standardised tests were banned from the Soviet educational system in 1936 partially as an attempt to escape the notion of classification (the idea of classification bore all too strong a resemblance to an autocratic tsarist regime - a regime which the governing party of the day sought avidly to avoid) (Valsiner, 1988). The history of intelligence testing is loaded with political and ideological sentiments especially regarding the "divine right of the King" prominent throughout Europe up till the late 19th century and manifesting particularly in countries such as Britain and the greater Soviet Union. "In more recent times the corresponding theory that supports a class structure is the belief in the innate superiority of the educated upper classes, with its corresponding overemphasis on the genetic components of intelligence" (Wortis, 1987). It was this very issue of class distinction that was partially responsible for the abolishment of standardised testing in the Soviet Union with the resultant emphasis being laid upon instructional assessment of children, i.e. assessment via learning processes and hence a precursor to the dynamic assessment of children.

⁵ It is imperative to note that as preliminary reading suggests, ideas and notions (such as Vygotsky's Zone of Proximal Development [ZPD]) cannot always be interpreted as "theory" as is commonly accepted in scientific circles. Dynamic assessment research at times produces conclusions which may or may not adhere to the scrutiny of what constitutes a theory or model. For instance, in the span of just one article by Sternberg and Grigorenko (2001a), Vygotsky's ZPD was at once described as being a 'notion', a 'concept' and a 'theoretical notion'. Likewise Feuerstein's approach is referred to as a 'paradigm' - is it worthy of paradigmatic status at this stage as stipulated by Kuhn (1962)?

⁶ A post-modernist may ask why the need to impose coherence on a system is necessary in the first place. The fact that gaps are found in many areas within dynamic assessment theory and practice may not necessarily be construed as a negative aspect from a post-modernist point of view (Mautner, 2000). However, as will be highlighted in due course, in order for this field to progress within the current structure in which it is located, these gaps (which are poor in information regarding theory and practice) are areas which at this stage detract from dynamic assessment's progress and hinder its development.

⁷ By this is meant the theories, models, philosophies, axioms, meta-contexts, constructs, schematas, frameworks, typologies, methods, concepts and ultimate goals behind various intelligence theories and models. Ultimately theory is built from concepts, variables, statements, formats, axioms, propositions and hypotheses (Turner, 1982).



takes place within a multi-layered context consisting of cognitive and non-cognitive variables (Sternberg & Grigorenko, 2002), distal and proximal influences (Feuerstein, Rand & Hoffman, 1979), socio-cultural variables (Vygotsky, 1978), impinging contextual factors such as the continued nature/nurture debate (Fienberg & Resnick, 1997; Miller, 1989), the socio-economic environment (Murphy & Davidshofer, 1998), biological predisposition research (Grigorenko, 2004a; Quartz & Sejnowski, 1997a), developmental models (Van Geert, 2000) as well as historical/political/social/cultural and philosophical forces which inform and are in turn informed by intelligence research; theory generation, meta-theorising and the progress of social science research (Shweder & Fiske, 1986) amidst many other factors which have shaped and continue to shape current trends in intelligence research. Regarding the importance of historicity, Louw (2002) recently emphasised the need to critically assess history when interpreting the position of psychology within society and this same sentiment rings true of dynamic assessment research, "...*history* provides a more powerful starting point for an analysis of the position and functioning of [dynamic assessment] in contemporary societ[y]" (own emphasis and substitution) (p.2). The above-mentioned context of social reality is embedded within the study of both individuals and groups within society. Literature within this field points decisively at this level of analysis, often not venturing beyond the context of "the group" (such as classroom, student lecture group, hearing-impaired learners, grade-school children, mentally-retarded individuals etc). Organisations and communities will undoubtedly benefit in the long run, but the focus within the area of dynamic assessment research has thus far confined itself to the study of individuals and sub-groups (Carlson, 1995; Haywood & Tzuriel, 1992; Lidz & Elliott, 2000a; Van der Aalsvoort, Resing & Ruijsseenaars, 2002).

The proposed study is theoretical in nature and as such, the above-mentioned factors do not *directly* impinge on the outcome or conclusions reached in the study. However, these factors play a major role *indirectly* in constituting the reality within which practical research takes place. In an empirical study these factors would necessitate in-depth analyses and an obviously closely integrated study. A theoretical study by nature a level above practice in terms of concept definition already encompasses these variables or factors in its theorising or model building. It is in essence a meta-theoretical approach towards the study of dynamic assessment framed within the progress of intelligence research (Berger & Zelditch, 1993b; Denishoff, Callahan & Levine, 1974; Fiske & Shweder, 1986; Hempel, 1970). In constructing a new theory or model, these factors would be considered in the process as major contributors.

1.6 Pertinence to the South African context

South Africa, although a fledgling democracy and despite great strides that have been made in a variety of contexts since 1994, still harbours an enormous amount of untapped potential in its peoples. Schooling in South Africa has unfortunately lingered under great strain and stress and as a result many students are not equipped to handle the educational environment within the tertiary context. The various results from dynamic assessment studies (Murphy, 2002; Murphy & Maree, 2006) has shown that the validity and reliability of this form of assessment is tenable as a means of predicting academic success in the tertiary environment within limits, as well as performing a function aside from the educational context (such as language disorders and hearing-impaired research) (Pena, 2000; Pena, Igllessias & Lidz, 2001).

The social, legislative, cultural, economic and moral circumstances in the country are legally⁸ founded on equality and opportunity and these ideals are automatically subsumed within various sectors of society, education assuming a leading position. In order to redress the past circumstances in this country the above-mentioned factors play paramount roles and are perhaps more overly sensitive to discrimination due to the tainted past. Basic health, low-cost housing and education are areas that are receiving more attention from government and corporate citizenship across the board in South Africa. This study seeks to research further the theoretical framework in which dynamic assessment is housed so that further practical research can take place within educational assessment and in so doing allow more people greater opportunities to grow and contribute to society by means of a better education. This is, needless to say, a long-term goal for which to strive.

Past methods of assessment of individuals remain an area of focus. Dynamic assessment research can aid in this endeavour. However, a stale-mate has been reached in the area of dynamic assessment research across the globe in that dynamic assessment cannot adequately progress upon a theoretical footing that is tenuous to say the least and in order to construct assessments, engage in educational enrichment studies and in general proceed with new batteries, a theoretical model which seeks to synthesise various approaches within the intelligence field needs to be investigated first.

Elliott (2001) states for instance that the possible reason as to why dynamic assessment has not reached as wide an audience as it should, may be due to the methodological and procedural difficulties encountered when attempting to identify what exactly learning potential is, "one problem lies in the many different tasks and modalities that can be used to assess this construct" (p.186) and this is of particular concern to this study as the ill-defined notion of hypothetical and empirical construct poses many

⁸ It must be noted that since the inception of the new-found democracy in South Africa in 1994, equality in its broadest sense is accepted legally. However, the pervasiveness of inequality in many areas of life is still evident. Reality thus tempers the notion of equality for all, notwithstanding this reality however, constitutionally all are equal in the eyes of the law.



hurdles. Note that there is as yet no consensus within intelligence research as to the exact definition of intelligence either (Anderson, 2005a). This is of course a main issue both within the broader intelligence field and dynamic assessment, hence the need to delve into ontological and epistemological questions. Elliott (2001) maintains that lack of explicit delineation of dynamic assessment's purpose may also well impede the progress of the development of this sub-field. Sternberg and Grigorenko (2001a) firmly put forward their reasons as to the limited utilisation of dynamic assessment, namely that current conceptualisations of dynamic assessment culminates in ill-defined conceptualisation, operationalisation and the subsequent analysis of the construct. Perhaps the most enlightening statement made by the aforementioned authors is that "the reason that dynamic testing may have remained merely promising is that we have incorrectly conceptualised and operationalised what it tells us, resulting in scoring that gives us a misguided conception of what individuals can do" (pp.162-163). However, Haywood (2001a) asserts that Sternberg and Grigorenko (2001a) themselves betray a lack of understanding of what exactly dynamic 'assessment' is when they refer to it exclusively as dynamic 'testing', as it binds Sternberg and Grigorenko (2001a) to the psychometric tradition with all the attached issues of standardisation.

Kaniel (2001) reiterates the state of conceptual confusion in which dynamic assessment presently finds itself,⁹ maintaining that the very term has become indistinct and that it becomes more "fuzzy" as one proceeds towards implementation. "The blurring of the identity of dynamic assessment is manifested in three areas: (a) conceptual confusion, (b) approaches that are not differentiated and (c) failure to produce an effective intervention program from the assessment" (p.112). Kaniel continues to highlight the areas of investigation hitherto neglected by many dynamic assessment researchers, such as individual differences,¹⁰ up-to-date integration of neuropsychological research, the advances within cognitive, personality and motivation research and concludes that this only reflects a partial list. Noncognitive aspects such as motivation (the core construct within the model proposed by Grigorenko and Sternberg 1998; Sternberg and Grigorenko, 2001; 2002) and attitudes affect performance even on mediated tasks - an aspect that is not duly considered as priority, as lack of transfer is normally associated with lack of potential within dynamic assessment research, an erroneous claim at best¹¹ (Kaniel, 2001). Ultimately, Kaniel concludes that dynamic assessment perpetuates the mislabelling of children just as static tests do due to the infrequency of dynamic assessments thus resulting in static scores, scores which "perpetuate labelling to a unidimensional score" (p.217).

Resing (2001) brings to the reader's attention that, contrary to the view expressed by Sternberg and Grigorenko (2001a), not all learning potential tests are uniform in format, in other words there is no one specific format for the assessment of learning potential as varying forms of this dynamic approach highlight varying aspects of the potential of individuals. Moreover, Resing (2001) makes mention of the seldom-stated aspects of the different phases of the diagnostic process through which clinicians and their clients proceed. Learning potential assessments, like any other assessment tool, must reflect cognisance of what the aim of the assessment is as well as the answers that are sought from diagnostic queries. Is the goal remediation, prediction or classification? Depending on the aims, not only should alternative forms (static vs. dynamic) of assessment be considered, but so too should the formats (graduated hints, test-teach-test etc) in which the assessments are conducted. Predictive assessments are not necessarily the best format for the assessment of potential, as the nature of the test is not geared up for this nature of inquiry.

A redress in terms of basic philosophies and progress within intelligence research is necessary now more than ever. This study aims to enlighten the progress of dynamic assessment research as a context within intelligence research and takes cognisance of both dynamic assessment's philosophy as well as intelligence research within psychology as a discipline, the growth of testing and the contexts of key players in the field of dynamic assessment which have influenced the direction of research and philosophical grounding within this test movement.

1.7 Dynamic assessment's entanglement

There is currently a heated debate ensuing in academic circles about the status of dynamic assessment within the realm of psychology. In fact this debate has been an ever-present one since the publication of Feuerstein's groundbreaking 1979 book on the dynamic assessment of retarded performers and resurfaced in 1992 when Frisby and Braden severely criticised a number of aspects within dynamic assessment research. During the period between Feuerstein's book and Frisby and Braden's article there were numerous debates within the field but none reached fever pitch until 1998 when Grigorenko and Sternberg (1998)

⁹ And the rest of psychology as a discipline (Jordaan, 1989; Lichtenstein, 1980); the problematic issues evident in dynamic assessment can be seen to be symptomatic of the entire field.

¹⁰ Individual differences as they manifest in test results are becoming more pertinent in the cognitive development literature, and as dynamic assessment is more attuned towards emphasising individual differences, it would make sense to further this area of research. In fact Byrnes (1996) places heavy emphasis on such differences when discussing cognitive development in his book. Variation within cognitive architecture is due primarily to genetic differences but for the most part this architecture is similar across people but experience and environment also account for differences in cognitive abilities (Stillings, Weisler, Chase, Feinstein, Garfield & Rissland, 1995).

¹¹ However, Resing (2001) does not consider motivation and personality characteristics as part of the concept of intelligence, even though she acknowledges their importance.

published their lengthy article, once again critiquing the field. There has once again recently been an upsurge in criticism within the field. Proponents of dynamic assessment recently engaged in a peer-reviewed debate about the issue of dynamic assessment's integrity and it is pertinent to expand upon somewhat as it further qualifies the need for this issue to be redressed.

The 2001 *Issues in Education* journal dedicated volume 7 number 2 to this debate. Of critical importance is the response garnered from a host of astute researchers within the field of dynamic assessment to the article written by Sternberg and Grigorenko (2001a). The core of the issue rests with the confusion pervading the field both between professionals and seasoned researchers as well as between professionals and novices to the field (Sternberg & Grigorenko, 2001a). Dissension among the ranks also adds to the confusion addressed in the article. Progress within the dynamic assessment field has increased in terms of the number of published articles and results from disparate sources, yet the one common and uniting feature about this mass of research is the lack of coherence and lack of identifiable theory and common thread running throughout the discipline. In fact the dynamic assessment "trend" lacks the very ingredients that most theories in science and social science require for it to be accepted as theory at all. A question that can be raised is whether or not the psychometric approach towards understanding intelligence can be referred to as a science. Rust and Golombok (1992) ponder this very issue and respond by highlighting the fact that social science and natural science research cannot be equated in terms of what constitutes science and the progress thereof. These contentious statements will receive due deliberation in chapters 3 and 4.

Sternberg, among others, as a leading proponent of intelligence research, has tackled many burning issues within the field and has subsequently received due criticism of his own (Fernandez-Ballesteros & Calero, 2001; Gerber, 2001; Haywood, 2001a). However, his critique has come at a time when it is desperately needed. "Some data on dynamic testing are mixed. Some of the data are positive - especially when they are collected by the originators of a given approach" (Sternberg & Grigorenko, 2001a, p.161). Furthermore, it has been intimated that dynamic assessment has yet to reveal its "promise" of a revolution in testing and that not all dynamic testing can assume the role of performing "miracles" of measurement with the zone of proximal development (ZPD). Another major point highlighted by Sternberg and Grigorenko (2002) is the fact that not only are general issues within the dynamic assessment field debated, but even basic issues¹² have yet to receive clarification. It is argued that one of the main reasons why the field is in disarray is due to the sound and fundamental conceptual framework which it lacks. Another seemingly glaring backlog within the broader intelligence field is the fact that the field has not in essence moved beyond the conceptual confines of intelligence as construed by Spearman (1981)¹³ (originally in 1904) (Sternberg and Grigorenko, 2001b). A refurbished and renovated or remodelled view of dynamic assessment within intelligence is necessitated. These issues are unwieldy and warrant research at a grand scale. However, it is not the goal of this study to re-invent the field of intelligence, but merely to add clarity and focus to one prominent and continually growing sub-field within it. It is the above-mentioned scenario that has served as the motivating factor and thus spurred this author to conduct this enquiry into the underpinnings of dynamic assessment and to develop for it an integrated and conceptual framework for future studies. That a unified theory as such is to be developed is not the aim nor focus within the framework development and the two ideas are not synonymous (Cooper & Shallice, 1995). Unified theories (as they relate specifically to the cognitive sciences) seek to explain the larger picture; "if a theory covers only one part or component, it flirts with trouble from the start. It goes without saying that there are dissociations, independencies, impenetrabilities, and modularities. These all help to break the web of each bit of behaviour being shaped by an unlimited set of antecedents. So they are important to understand and help to make that theory simple enough to use. But they don't remove the necessity of a theory that provides the total picture and explain the role of the parts and why they exist" (Newell, 1990, p.18). It is with this in mind that a meta-theoretical framework is warranted and may prove helpful.

1.8 The role of science

Science as a concept and word originates, from among other root words, the Latin "scientia" meaning "to know" and science's main duty is to solve issues that pose some sort of problem to society and in so doing become familiar with the issue at hand. In essence, it is "to know". Another point highlighted above, is the fact that these issues occur in everyday life and this cannot be more stressed than it is within the social sciences as many issues affect the ordinary citizen in society. In essence, any basic scientific exercise that seeks new knowledge and information is in fact contributing to the advancement of science, no matter how obscure the data might seem. The point is any new knowledge will, at some time or another, contribute to science, but the question should really focus on how great the impact will be. The specific context within science here is of course dynamic

¹² Basic issues here relate to the definitions of dynamic tests and dynamic assessments as well as to what constitutes dynamic assessment and what does not. In their rebuttals to various researchers' responses to the original article, Sternberg and Grigorenko (2001b) state that the "mutual conflict, in itself, is not worrisome. What is worrisome is the certainty with which each expert touts his or her own particular point of view in the name of 'fact'" (p.253).

¹³ As Sternberg and Grigorenko (2001b) point out, "Would any other field in psychology, or science, for that matter, have pride in using a theory that dates back close to a century?" (p.257). Similar sentiments are echoed by Suzuki, Ponterotto and Meller (2001) who lament the lack of change as it pertains to multicultural assessment. Cf. also Neisworth and Bagnato (1992) in which intelligence tests are in general criticised due to a lack of definition of what constitutes intelligence as well as the lack of theoretical basis for most intelligence tests. Carroll's (1996) reverence for Spearman's intelligence research is justifiable yet he does note the current preoccupation with theories well past their sell-by date.



assessment research within intelligence research within the broader field of psychology and some proponents within psychology will most assuredly consider the discipline a science. By attempting to "rescue" dynamic assessment research and make it more palatable to practitioners everywhere, the field needs to be grounded in a coherent and theoretically defensible framework. It was Arthur Eddington who once stated that an experimental result should never be trusted till it was confirmed by theory. It is precisely this notion that will serve to guide this study. The advancement of dynamic assessment as a science can only take place once this bedrock of theory and concept is properly cemented within the field, and as such, this study can most certainly advance science in this regard and these issues will be discussed in chapter 3.

1.8.1 A personal note on philosophical affiliations

In order to clarify at the outset the aims of the study, it is considered prudent to unambiguously and clearly state the case of the author's own intentions and inclinations within the sub-discipline of psychological assessment, namely dynamic assessment within intelligence, within the broader discipline of psychology, within the yet more encompassing scope of social science research and ultimately the broadest of all schemes: science and as is stated here, "before undertaking any psychological study, one must assume a basic belief in the specific nature of life, which is a philosophical exercise" (Brennan, 1982, p.138). This is reminiscent of Winch's ideas concerning the nature of understanding in its broadest sense "for any worthwhile study of society must be philosophical in character and any worthwhile philosophy must be concerned with the nature of human society" (1970, p.3). This advice is taken to heart. After all what is a PhD but a degree of *philosophy* which in this case happens to have as focal point psychology (Royce, 1970)? Like Dawkins (1999), the author wishes to convey the unabashed feeling of surety in her firm predilections for certain views which are not always in vogue (certainly not currently) and in order to accomplish this the tone of the thesis "is not conciliatory or apologetic - such is not the way of an advocate that sincerely believes in his case" (p.v).

Regarding the notion of science per se, a preference for more positivist stances (which includes a host of subsidiary philosophical branchings such as instrumentalism) is ceded as underlying affiliation, with a preference for the works of such historians and philosophers of science such as, among others, Francis Bacon; John Locke, David Hume; August Comte; the philosophers, mathematicians and logicians associated with the Vienna Circle; John Stuart Mill, Carl Hempel, Ernest Nagel, Charles Peirce, John Dewey and the numerous natural scientists who have uncovered and yielded to the public world a wealth of useful information.¹⁴ Doubtless, these early views are perhaps more restrictive as a framework from which to work, but the point here is not that firm advocacy of these approaches cannot be made malleable by refining of these originators' works but more the point that these are the roots from which personal affiliations emanate and also, these aforementioned luminaries did not always agree on many aspects, resulting in contentious and often heated debates (Outhwaite, 1987). Conceptual shifts within the discipline are an ever-present occurrence in which there are shifts away from justificational epistemologies such as logical positivism, a move away from classical rationalism, a subsequent re-appraisal of classical realism, a move away from determinism to the acknowledgement of complex systems and much else besides (Mahoney, 1989). There are countless virtues in these reappraised movements and surely befitting of growing trends of newer ideas but as Miller (1985) states, at times the discipline is reminiscent of an "intellectual zoo" (p.40). However, the lure of the original philosophies is an enduring one, much can be done to tailor it, but the foundation is cemented and for a reason perhaps. The author's acknowledgement of theoretical psychology¹⁵ as a subdiscipline within the field may also at times seem contradictory to her affiliations with positivist schools of philosophy as the positivists' leanings were certainly not in the direction of theoretical speculation (Kukla, 1990b; Slife & Williams, 1997).

Regarding psychological movements and schools of thought, the author is less than enthusiastic about the future of the discipline of psychology in its current form or should that be forms? Which is perhaps the point. There is no need for fear of the loss of one's favoured area of interest and research, for these can be accommodated quite freely within other disciplinary areas. The existence of psychology departments the world over is testament to the discipline's firm positioning in academia (at least this is how it is seen from the lay public outsider point of view). That ponderings of the splintering and continuing fragmentation of psychology as a discipline is negative and not altogether helpful is acknowledged by some as aiding in the downfall of this discipline. This is considered an unfair allegation, and an affair seemingly within the grasp of theoretical psychology given the fact that the "founding fathers" (James and Wundt) were mostly theoretical in orientation (Slife & Williams, 1997). Fragmentation can also be viewed as a consequence of "disciplinary boundary maintenance that is called into play by personal, social, and institutional forces" (Stam, 1998, p.70) and not merely as a problem in need of a solution. The continuing search for knowledge, whether to aid in general or to simply satisfy human curiosity might not be value neutral but the methods should be (subjectivities taken into account). The physical sciences are replete with sub-disciplines evidencing ever-increasing fragmentation, yet physics still manages to progress along a firmer footing than psychology (McNally, 1992); so fragmentation

¹⁴ Ignoring for the moment those darker periods in history during which much science has indeed had dreadful consequences (and currently still evidences), but then again, it is not the method of discovery which is being critiqued here, but the all-too-common denominator called the human factor.

¹⁵ Not necessarily equated with philosophy (Slife & Williams, 1997).

per se cannot be proffered as sole reason for the lack of similar progress within psychology. Furthermore, the author firmly believes that all manner of statistical treatment of data can confer much understanding of numericised data, but ponders the often explicit denial of the lack of fit that these manipulations of data sometimes evidence thus resulting in a mere plethora of data and analysis bringing the discipline no closer to explanations of behavioural functioning and phenomena than was the case fifty years ago. Why is this? Why is there such staunch resistance in seeking alternatives to the methodological study of social concerns? Enter the various “isms” of which there is also an unfortunate plethora. One might state that the author cedes no leeway for any action at all; what with critiques of how social science is now practised and how it is purportedly being rectified. But is it being rectified at all? The author’s views, just as with any other view, is of course open to much critique of its own, but an attempt is made to aid in directing, at least for dynamic assessment, a potential path to follow which may allow for greater acceptance of what this method of assessment has to offer the ever-burgeoning field of intelligence research.

These preferences are really of course just that: preferences. Simply ignoring the wealth of views emanating from a number of alternative perspectives would be tantamount to a heresy of sorts! So many wonderful and immensely inspiring ideas flow from the works of opposing schools of thought, that it becomes, at times, very difficult to attempt to draw a line demarcating this preference from another. Also, at times, is the yearning to draw from alternative views in contradiction to the views espoused by the aforementioned.¹⁶ However, wild eclecticism is to the author’s mind not a viable manner in which to study phenomena in general. Fusing ideas from various views into a coherent and defensible treatise is one thing, but willy-nilly picking and choosing from what ostensibly looks good is not good practice. Constructionism, deconstructionism, relativism, feminism, Marxism and countless many other “isms” pour forth into “enlightened” current literature but the author questions the real progress that much of these movements in fact proffer as. As Grace and Ferreras (1998) maintain “although logical positivism may be untenable from a strictly philosophical view, postmodern and postpositivist philosophies have yet to articulate a more effective methodology for psychology” (p.68) where qualitative methodologies have often been paralleled with the latter and quantitative methodologies paralleled with the former; a state of affairs in need of rectification (Michell, 2005). It is always good practice then, to posit ones own affiliations and leanings regarding the topic at hand, for much energy will be spared in trying to defend the argument. The author vehemently criticises much of what the aforementioned schools have to offer the social sciences, not their methods per se, but rather the application of their methods to the study of the social sciences in general and often lays the blame for the state in which social science finds itself squarely at the feet of those social scientists positing critiques of their own about the methods they so carefully endorse! Lest the author of this treatise be accused of much the same thing, the defense of affiliations is stated at the beginning.¹⁷

This study, then, is neither a study in the philosophy of science nor social science, neither is it a study in the philosophy of measurement nor a study on the subject of intelligence per se. It is a study on how such philosophies impinge on the placement of dynamic assessment within intelligence, for it is hypothesised that these very foundational issues lie at the root of many problems facing the future of dynamic assessment. It is “back to the drawing-board” time, time to stop and think about what it is that dynamic assessment is trying to achieve within the greater realm within which it is subsumed, whether or not it likes this enclosed area within which it works and how best to sustain it in its present form, or to at least attenuate it so that it remains in existence albeit in different form. It must be noted at the start of this thesis that although the application of the meta-theoretical model is developed through chapters 2 - 4, continuous comment will accompany the chapters in terms of applications of certain views to dynamic assessment. This is done so as to present a coherent flow of thoughts throughout the study and to remind the reader that each chapter is fully intertwined within the study but needs to be discussed within its own context.

1.9 A map to the study

The study takes a number of argumentative turns throughout the ensuing discussions and a map may be helpful as an aid in placing the various chapters into context (see figure 1 below). The study is an exercise in:

- “Exploring” - hence no definitive conclusions are reached, however, firmly supported statements are upheld throughout the study which consistently wend there way through the exposition. Preceding chapter 3 are epistemological and ontological issues which form the bedrock of theory development and hence impinge on the movement of the discipline as a whole. Certain issues are discussed as they pertain to intelligence assessment (chapter 2)

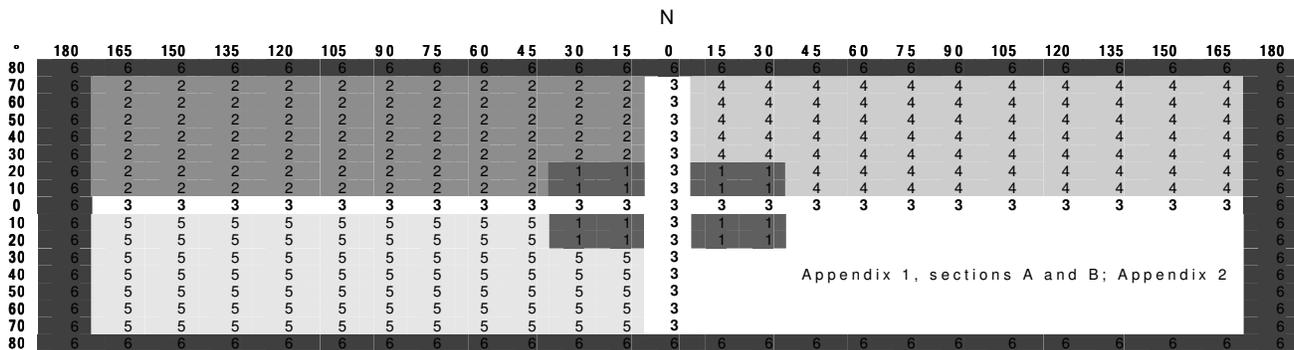
¹⁶ For instance, as will be evidenced in chapter 2, the author has leanings towards a more reductionist approach to the study of human behaviour and as such prefers inter-theoretic reductions (if such theories are not competitor theories) but takes exception to the strict reductionism offered by evolutionary psychology in explaining away human behaviour as an example (Maiers, 2003).

¹⁷ Although there is a firm affiliation for the scientific method (with attenuations of course) the author also acknowledges that science, as with any other knowledge-gathering enterprise is open to all sorts of recriminations. The author has learned that nothing is completely known, nor will it be completely known in the foreseeable future, but there is another belief which resonates with this sentiment and that is the firm belief that, in time, many things will become completely known. But for now, there is always doubt, even in and of scientific products, be they natural science or social science products.



- “a meta-theoretical framework” - in order to address meta-theory, the first step is to address theory (chapter 3); within theory, concerns such as science development, social science progress, psychological theory development and theory appraisal are addressed (chapter 3). Lastly, a meta-theoretical framework is attenuated and deployed for the purposes of this study
- “for dynamic assessment and intelligence” - a major artery within psychological assessment is the life-supporting role played by the quantitative imperative. The manifestation of this quantification plays out in the functioning of null hypothesis significance testing, supported in part by mathematical foundations. Dynamic assessment is placed precariously between poles of differing assumptions (chapter 2 and 3) and statistical and measurement issues plague this manner of assessment in deterring it from a progressive path forwards (chapter 4). Dynamic assessment and intelligence can only be discussed once the foregoing foundations are laid down. Utilisation of the chosen meta-theoretical framework culminates in the discussion of the placement of various dynamic assessment models and theories within intelligence assessment (chapter 5) concluding with chapter 6.
- Two appendices yield new information. Firstly, South African data are used to determine an effect size from a meta-analysis conducted on dynamic assessment studies and secondly, questionnaires are analysed for pertinent themes which address the cadre of queries posed in chapters 2 - 5. Paucity of original data has relegated this section to appendix status

Figure 1 The analogous co-ordinate map illustrating chapter placements



Key:

This two dimensional co-ordinate earth-based map illustrates the positioning of the various chapters. Zero degrees longitude and latitude are the two axes along which *chapter 3* is positioned. The meta-theoretical framework forms the two axes throughout the study and serves to keep the structure in place. *Chapter 1* occupies the least co-ordinate space as it is merely introductory and is thus located around the centre of the map serving as point of origin. *Chapters 2 - 5 as well as the appendices* occupy the four quadrants as they build the structure of the study. *Chapter 6* concludes with all inclusive chapter information thus framing the co-ordinate map. Shading merely aids in visualising the chapter sections. The *Appendices* quite rightly deserve chapter status but due to quality of data were chosen to reside in the fourth quadrant

1.10 Conclusion

An overview of the study has been presented in which the motivation and rationale has been discussed and chapter demarcations have been delineated which detail the nature of the study. An in-depth review of the current debates within the dynamic assessment literature has been highlighted to further support the need for a study such as this. Chapter 2 proceeds with an investigation of the author’s implicit philosophical underpinnings and affiliations as these pertain to the field of assessment, after which an exploration of a meta-theoretical framework for dynamic assessment and intelligence will be discussed.

Chapter 2 Research method, philosophical affinities and dynamic assessment within intelligence

2.1 Introduction

This chapter presents a brief overview of what is considered to be vitally important issues when discussing theories and models within dynamic assessment and intelligence. The meta-theoretical framework developed by Madsen (1988) is the framework employed in this thesis. A major area of concern within his overarching scheme is the amount of attention paid to basic, fundamental philosophical positionings with which theorists align themselves. This framework will be discussed at length in chapter 3 but it is necessary at this point to align this study with the concerns voiced by Madsen and Jordaan (1989) and that is to strive for explicit formulations of implicit assumptions pervading theories and models or even points of view. Although the author is not developing a theory but rather assessing other theories it is nevertheless deemed important that implicit assumptions are rendered manifest at the outset of any study.¹

2.2 Literature - prime source of information

For the sake of clarity the reader is briefly engaged in the process undertaken to elicit sources from online databases. The following macro-search encompasses platforms, publishers and individual on-line journals (the list of individual journals is not stated herein, due to the fact that a substantial number of these journals are already indexed by the platforms mentioned below). Only a few individual journals are searched; these being journals that are only available on-line and hence not necessarily indexed by traditional platforms.

The search terms combine the need to include the status of theory within psychology and dynamic assessment research, which itself is subsumed within the field of intelligence research. Overlaps in the content of the number of "hits" were expected but due to the nature of the indexing criteria as specified by different vendors, this frequency is not at all high. Thus, the search is divided into the above-mentioned categories and included keywords/phrases captured within the "title" and "keywords"; as well as being listed in the "title, abstract and keywords" as these formed part of the refined searches as they now included a greater scope of research. Platforms searched follow below:

2.2.1 Platforms searched²

- Bookfind which includes Premier Service as well as Bookwise; Cambridge scientific abstracts; Current contents at Ovid; EBSCO Host which includes Academic Search Premier and Master File Premier (2000+); ECER (1969 - 2002/12); ISI web of knowledge which includes Science Citations Index, Social Science Citation Index and Arts and Humanities Index (2000+); ERIC (1966 - 2002/12); Emerald; HighWire press; Ingenta; Ingenta Select; InfoTrac - expanded academic ASAP; Kluwer Online; Philosopher's Index; Oxford Journals; ProQuest; Psyche; Psycholoquy; PsycINFO (1887+); Sage; Science direct; Springer Link; SwetsWise; Wiley Interscience and Wilson's humanities abstracts

Three local South African databases were scoured for information and research into the topic, namely the Nexus³ National Research Foundation (NRF) database, the Human Sciences Research Council (HSRC) database as well as SABINET; the local South African database indexing and amalgamating higher education and academic institutions publications.

2.2.2 Questionnaire

Most of the information contained within this research was gleaned from published sources as well as from interviews conducted with practitioners in the field. Questionnaires were delivered to one hundred dynamic assessment practitioners across the globe. More information concerning this exercise can be found in Appendix 2 which delineates the entire process from questionnaire

¹ Although not yet discussed, the author is, by commencing this chapter, attending to what Madsen refers to as the meta-stratum of concern. The issues pertinent to this meta-stratum are considered paramount to the scientific endeavour of model and theory development, especially within the social sciences which do not possess the luxury of more constrained systems as found in the natural world (before a retort is offered to this statement, the author wishes to remind the reader that the study of human behaviour will most likely prove the most perplexing of all complex systems seeing as we are a part of the system we wish to study; an inherent and inescapable limit within the system thus resulting in a system which cannot *ipso facto* be studied). Not to mention the fact that unlike physical systems which are more often than not uniform in nature, human behavioural systems are not (Feist, 1995).

² Of particular note within these databases are PsycINFO, ERIC, ECER, EBSCO host and ISI web of knowledge which are the main sources for the topic at hand.

³ Not the most reliable database and certainly not up to date.



conceptualisation through to end results. However, due to the very low response rate, it was deemed more appropriate for this study to append the exercise rather than have it as a focus and topical point within the body of this thesis.

2.3 Model utilised

This theoretical study seeks to advance dynamic assessment theory within intelligence research and in order to explore a meta-theoretical framework which will suit such an endeavour, analyses of existing theories, models and mini-frameworks will be studied with the aid of a model of comparative criteria as originally set out by K.B. Madsen of which more will follow below.

2.3.1 The pervading meta-theoretical framework

Of the few models (Royce, 1975) that expand on the meta-theoretical advances within psychology far fewer are evident for psychology's sub-discipline of intelligence assessment and none are evident within the further submerged field of dynamic assessment.⁴ A comparative overview encompassing various strata of influential forces will serve as model according to which assessments of theories, models and schemas will be studied and conclusions deduced from such comparisons. Although references other than those of Madsen are consulted and fused into his models, Madsen's understanding of the broader field upon which the theories and models are constructed and subsequently play out was deemed more in keeping with the meta-theoretical notions and research predispositions of the author. In Chapter 3 Madsen's views on models, theories, meta-theories and overarching philosophy of science will be looked at in-depth; yet there is a two-fold reason as to the necessity of warranting this chapter. Firstly, the author's implicit assumptions regarding certain key philosophical issues play out within the broader scope of dynamic assessment within intelligence and secondly, these views are subsumed within Madsen's framework of meta-theoretical psychology.

The "fit" then, between methodology and scheme and likewise between Madsen and author should not be underestimated as the author too has leanings, philosophical preferences and ideas of her own which will ineluctably colour the process of investigation and conclusions reached at the finale. That this remark should so early on make its presence is rather telling of how strong a consideration it is that prior beliefs, notions and ideas pervade any work within the realm of science. Without opening the floodgates to a heated debate between the empirical-realist vs. relativist stance on how science is practiced, let it be known at the outset the various philosophical positions of this author.

2.4 Laying personal cards on the table: a deck of truth

In order to contextualise and frame the study within a time and place a brief excursion into the author's philosophical leanings is warranted as these leanings impinge on the study consciously and seep in throughout unconsciously.

2.4.1 Geographic and socio-political location

The author was born and has always lived in Pretoria, South Africa. A climate of tolerance and democracy followed a long period of repression and malcontent. A history replete with misunderstandings, intolerance and chaos ended after considerable struggle within totalitarian regimented governance. Although anchored to the past, the emerging socio-culturally multilayered society within which the author works has undoubtedly influenced most life experiences and in areas none more so evident than within intelligence assessment. The confluence of Western ideals has at times met equally with resistance, appreciation and integration in ways which are at times more difficult for westerners of other industrialised nations to understand and with which to get to grips.

2.4.2 Time

All times are unique and interesting. To say otherwise would indicate a time and place in which nothing much happens in the way of progress or lack thereof in any area. The most that can be said of this period in human history is that we simply have a larger repository of historical facts on and from which to lean and gather information. Much of it is obsolete (phlogiston, ether, Freudian psychoanalysis⁵) (Skinner, 1986) and much is unnervingly prescient to occurrences at present⁶ (how not go about assessing for a reified construct such as intelligence). This work may well be relegated to the bin of obscurity in the year 2205 entitled along with other works as "research of the time" or work in keeping with the known facts at the time. How to "understand" outside one's own time is indeed a rare gift or luck as the case may be. Suffice it to say that apologies about what has not yet

⁴ As far as the author is aware at the time of writing this thesis, 2005. What may come to light of day or what might have been missed reflects one or both of two conditions: lack of thorough searching and constraints due to time. The former contributing more weight as erroneous research (controllable error) than the latter.

⁵ Interestingly enough though, Freud did attempt to redress his psychoanalytical thoughts and attune it towards a more natural science approach (a neurological reductionist one) (Wilson, 1996).

⁶ To add cliché to sentiment recall Santayana's saying about those being doomed to repeat history! (Cole & Valsiner, 2005).



occurred will not do. Strides in neuroscience, theory development, measurement techniques, quantum computation and behaviour in all its flavours will impinge on what is written now and although speculation as to what waits for us in years to come might stimulate and provoke controversy, the author wishes to leave such ponderings to philosophically inclined science-fiction writers.⁷ The great equaliser after all is time and as such it will tell.

2.4.3 Schooling within psychology

Beginning with Wundt's laboratory in Leipzig in 1879 and following a trajectory all too familiar within Western psychological academia, the author has been exposed to a host of eminent key players, rather too heavily focusing upon American mainstream psychology perhaps.⁸ That the role of such mainstream psychology should ever be underestimated would be a misfortune indeed but in order not to be blind sighted by such powerful testimony to such an edifice, a tolerant and incisive look at what emanates from across the globe might well yield fruits which if not tasted will start to rot. The roots of dynamic assessment are not always commensurate with the hitherto received notions of what it means to conduct assessments but as was hinted above, time will tell whether other psychologies will repatriate with their own territories or make their winding way into a confluence of mainstream Western psychology. Our global emancipated future (if the present is anything to go on) will surely be the result of less divergence in many areas of life. But who is to know. The fact that in 2005 dynamic assessment may either be confined to anecdotal footnotes within future psychological assessment texts or finds a voice large enough to engulf other smaller theories (or models) is testament to the precipice it now faces. The author's literary wonderings as far a field as Russia, Israel and the United States and as locally diverse within South Africa has led to the discovery of well written documentation abounding in areas within psychology over-and-above that with which undergraduate students are confronted in South Africa.

2.4.4 Personal experiences within psychological intelligence assessment

Confronted with intelligence tests that necessitate speedy and accurate responses, in a climate of anticipation and stress and being cognisant of a future which depends on these very answers on which one is about to deliberate (not to mention the consternation at being placed below your fellow peers who aim to outdo everyone else on the pretext of being socially very accommodating) is a snapshot of what a testing situation may be like for some. To have this experience repeated for the next eight or so years and then to have finally "made it" into an area of work in which eventual comfort is found is rather more telling of our cultural understanding of what intelligence is than what it in fact truly is. Not that one is yet protected within the confines of working life, as you are threatened from all corners regarding performance goals and assessments that will "help" you focus that little bit more. Once you have managed to secure a worthy fulfilling role within your working life, you are routinely assessed in many other areas of life. Old age assessments have now become a lucrative source of mutual beneficence (or so it seems) in which as a wiser yet "slower" individual you are now bombarded with yet more tests of well-being and global functioning assessments.⁹ From cradle to grave (Suen, 1990)¹⁰ one is confronted with tests of all sorts. Do they really make a difference and how? That they do is not in question, how they do so, however, is.

2.4.5 Philosophical leanings – implicit assumptions pervading this study

As things in life rarely remain the same (if even genes randomly mutate what chance do you really have?) so too does this sentiment apply to one's philosophical leanings and much the same as that well worn pendulum-analogy illustrates, a slow but progressive tick-tock is all that represents growth and development within any area but specifically within matters philosophical. To unequivocally state that one is such-and-such a person or that one has this-or-that leaning towards these-or-those issues is tantamount to stretching the truth somewhat as nothing is really set in stone. However, to perhaps venture a general leaning towards such-and-such a view is not altogether unfounded. That certain creatures do not change their stripes nor spots should nest this particular notion. Doing away with metaphors for the moment, it is necessary that philosophical positions or a so-called "philosophy of man" (Madsen, 1971) be delineated at the outset of a study dealing specifically with other researchers' leanings. Not to do so would seem hypocritical. There are many frameworks from which to choose and most offer similar axes and coordinates as to where one should comfortably place oneself. On three dimensions then, this matter will rest.

2.4.5.1 First axis – A Jungian typology

Nicely entrenched within this typology of types is the concise generality of its notion. Although generality can be construed as either a hindrance or a progressive stance on the matter of personality classification, the author feels at home in using this as

⁷ With whom I share affinities.

⁸ A common trend across the globe (Slife & Williams, 1997) which is not a criticism just an observation. However, as time moves forward, there is an increasing awareness of non-Western influences in psychology cf. Pillimer and White (2005) in this regard who focus on developmental psychology within the larger domain of science.

⁹ Our ever-increasing older population (mainly within industrialised nations) (Cohen, 2005) is proving to be "quite the sample" to study, what with all the benefits emanating from this potentially rich source of income.

¹⁰ Literally.

generally descriptive of her stance pervading this entire treatise. On the attitudinal type - introversion; on the psychological function - thought is paramount, followed by intuition, sensation and followed lastly by feeling; in totality the personality type which best describes the author is introvert with thinking and intuition (ITN). This rendition of events is simply the assessment made by the author in a very informal manner.¹¹

2.4.5.2 Second axis – A more traditional axis: an informal personal psycho-epistemological profile¹²

The more classical epistemological categories include such philosophical affiliations as realism, idealism, empiricism, rationalism and intuitionism (Royce, 1964, 1970a, 1973; Royce, Coward, Egan, Kessel & Mos, 1978¹³). On a rather mundane linear hierarchical scale, the author acknowledges her leanings in the following manner:

1. Rationalistic
2. Intuitive (metaphorism)¹⁴
3. Empirical
4. Realistic
5. Idealistic
6. Authoritarian

Cognisance is taken of the fact that although the author has positioned herself as an introvert in the Jungian typology this might seem contradictory to her subsequent positioning on the intuitive scale as evidenced above, as this is often regarded as a mark of extraversion (Eysenck in Madsen, 1987). This can be justified by stating that socially introversion is more evident yet within professional research the latter (extraversion) is more evident.

2.4.5.3 Third axis – A complement

Numerous dichotomies¹⁵ (legitimate or otherwise; Cole & Valsiner, 2005) pervade philosophical and psychological constructs, however the following seven contentious issues were decided upon as they play a large role in how the products of intelligence and assessment research endeavours flavour resultant conclusions. In particular for the purposes of this study and for this author, that these implicit assumptions guide any investigation is testament to their historical contingency and morally situated nature (Kristensen, Slife & Yanchar, 2000; Slife & Williams, 1997). The fact that some of these issues are not always suitably dealt with in the intelligence and dynamic assessment literature is telling of two possibilities: either that some of these to-be-mentioned issues are not considered problematic within the respective fields of study or that such issues are not instrumental in any purported advances to be made within these respective fields. That these issues are usually missing (or consciously avoided) further entices the author to state her proclivities in this regard as they are considered essential. The development of science in general, the social sciences and psychology in particular regarding theories and models is a primary concern for philosophers of science who seek to understand various conceptions of understanding, in other words such philosophers are concerned with epistemological and ontological issues which lead to different choices regarding various aspects including choice of assessment in this particular instance (Schraw & Olafson, 2003).

Essential to any research endeavour is a set of basic assumptions that are implicitly or explicitly held by the researcher (akin to a theory-laden approach) and these assumptions, depending on the specific context of study, influence if not the outcomes, at the very least the “feel” of the study. This is obviously evident when reading works of individuals as the following simple example will suffice: reading the works of Jensen (1994) as well as Valencia and Suzuki (2001) puts the reader into different frames of mind, at times resulting in unease due to simultaneous acceptance of basic philosophies which underpin both expositions, and as these two “sets” of writers both make sense there is a conceptual and philosophical struggle between accepting one over the

¹¹ That a radically divergent outcome should result is not anticipated.

¹² Taken directly from Joseph R. Royce but utilised in a somewhat different manner.

¹³ Upon first glance these dimensions appear very similar to those of Jung. This seeming similarity was affirmed as Royce makes mention of Jung in his 1959 address in American scientist, although not directly in conjunction with the dimensions as stated.

¹⁴ Although the two may at times be synonymous, it is noted that semantic confusion does exist about the exact difference embedded within these two concepts (Royce, 1973) although Madsen (1971) uses the two terms synonymously.

¹⁵ That dichotomies exist at all is somewhat perplexing given the serious advancement within science and philosophy in recent decades. Why many continuums have to be construed as either all-or-nothing itself gives one reason to wonder. Perhaps this is a manifestation of deeper behavioural inner workings of the ever-evolving brain. Or perhaps it is really just a reflection of “the intellectual history of our society”; a Cartesian dualism left-over perhaps creating “separation rather than inclusion”? (Bateson, 1995). Is this tendency one of epistemological or ontological origin asks Rose (1995, p.201)? Here we go again - yet another dichotomous ‘either-or’ sentiment! Gould (1998) refers to this human tendency as a “propensity for division by two” (p.30) and is so widespread because so much in nature can be classified into two categories and thus is really just based on good observation. Also somewhat akin to the mathematical principle of the excluded middle, against which intuitionist mathematician Luitzen Brouwer ruled (Clapham, 1996) but more of this in chapter 4. It is interesting how the tendency to dichotomise manages to seep into the groundings of mathematics which results in the question of how “true” our mathematical groundings really are if they are given to the whims of our evolutionary heritage of looking at the world in a certain manner. This too has a bearing on how we fashion our tools of assessment and how we regard what we regard to be correct and true.



other. The question may be asked as to whether one really needs to choose between them in the first place – perhaps holding two supposedly opposite views would seem contradictory in some way.¹⁶ For the time being remaining consistent is of greater importance. Not only are these issues important but the influential ties between them all is clearly evident as one works through the mind-brain problem right through to a discussion on static and dynamic forms of assessment. These issues are not isolated ones and often discussions within philosophy texts touch on each of these issues whether superficially or in-depth when detailing certain aspects pertinent to only one of these topics. In sum, these topics were chosen due to their pertinence to the topic at hand (dynamic assessment within intelligence) and because they overlap in content and subject matter. The following is neither a treatise nor a comprehensive survey of the terrain of issues involved, but allows the reader further insight into how the author's leanings colour the study of dynamic assessment within intelligence and it should be noted that some of these issues are not mutually exclusive but intertwined in many ways. These include:

- The mind-body “problem”¹⁷
- Consciousness
- G - dominated vs. multiple intelligence (MI) - dominated leanings towards the understanding of intelligence
- Emergence (irreducibility) vs. reductionism
- Realist vs. relativist approaches towards research
- Nature/nurture
- Static - dynamic assessment of intelligence and potential

2.4.5.3.1 Introduction

From the very start of investigation into the surrounding environment¹⁸ (within and without) splits, divides and opposing beliefs have been the stuff of knowledge-gathering and inference and like all other knowledge-gathering activities humans' propensity to continually divide systems of belief is and always has been omnipresent. Is it really necessary that a belief be located on one of two areas on a continuum? (Jencks, 2000). Is there not a manner in which one can move beyond the “inadequacies of dualism?” (Shakespeare & Erickson, 2000, p.190). It is reasoned that this unfortunate set of events (an evolutionary product of the human brain?) at time colours and alters somewhat the views actually espoused or endorsed. Such dichotomies pervade all these issues (from the mind-brain to nature-nurture controversies for instance and even construals of how intelligence is to be measured in both verbal and non-verbal manners; Flanagan & McGrew, 1997) and to seek alliance with one or the other view is rather unfortunate. In the hope of not being pedantic, the author then too, has to align herself to some degree with various views that are located on just such a prototypical continuum. Cognisance is taken of this narrow-minded construal of events. There are always two sides to every issue and in order to support a preferred view it would be quite easy to merely cite those in favour of the author's leanings. Of course this is no good as academic debate and the discussions below include “for-and-against” views in order to properly debate the reasons for the choices made on each issue. As Williams (2003, p.9) cautions whilst referring to theoretical psychology as a discipline:

we have been content to skate along the top of such issues, not getting into them too deeply, and not realizing that we all must take some position on these issues, and that we do so usually in an implicit manner without the benefit of careful and penetrating examination and contemplation.

2.4.5.3.2 The mind-body problem

That this issue is still problematic is telling of the status of received views concerning physiological findings (evidence) when transplanted into a philosophical context. Foregoing any attempt at an answer because the issue might be perceived as untenable and fundamentally undecidable (Bickle, 2003b; Koch, 1981) is not an option this author wishes to adopt. A strict type-type identity theory¹⁹ is perhaps the seemingly closest alliance that can be made philosophically in terms of a view endorsed by the author; as it is believed that such an isomorphism (one-to-one mapping of mental to neural; or sameness of structure²⁰) really does exist (Stillings, Weisler, Chase, Feinstein, Garfield & Rissland, 1995; Royce, 1959). However, due caution is

¹⁶ Unfortunately it often happens that questions within philosophy are answered by yet more questions. A frustrating phenomenon encountered within the study of philosophy (which is why this study is cognisant of philosophical viewpoints but is itself not a philosophical treatise).

¹⁷ Inverted commas have been inserted as this is not perceived as a problem in the conventional sense and is quite anathema to the way the state of things appear to the author. For there simply is no problem evident. That this is contentious is not in question.

¹⁸ Reference is made here to the very beginnings of humanities' obsession with understanding ourselves and the world in which we inhabit. Needless to say, science (as officially acknowledged and designated as such by an empirical method) is of course included.

¹⁹ Also globally referred to as the psychobiological approach which states that the “mind” is a collection of brain processes and that behaviour is both controlled by and is a product of the environment and the central nervous system. Compare this to strict behaviourism which states that there is no intermediary mind as such and that behaviour really is just a stimulus-response system (severely criticised by, among others, Chomsky, 1959) and that of animism which states that the immaterial mind controls the body (Bunge, 1985; Butterfield, Slocum & Nelson, 1992).

²⁰ Although indirect and not, at this stage, related to a comprehensive theory of intelligence (Newman & Just, 2005).



attached to the interpretation of research results emanating from nascent developments in this domain (Gardner, 1986). To unequivocally remark that this isomorphic mapping is true and always the case would be an absurd proposition as identity theory itself was a reaction to the behaviourists' repudiation of inner mental events. It is itself vulnerable to presumptions it cannot explain away (Lycan, 2003) relying too heavily on human biology as ultimate explanation of events across time and space. However, to deny the leaning in this direction would be fallacious. The mere fact that terms such as "mental" and "physical" have to be delineated within this take on the mind-brain problem already colours the stance proffered. The author's leanings are towards the unification of the two concepts and that mind/brain are synonymous. Perhaps as Damasio (2003) has stated, "mind" is the creation of brain²¹ the image of body implanted within the brain as neural networks in order to aid its survival. Cultural impingements either facilitate or hamper the adaptive functioning of the brain in the author's personal view. Time will tell whether cultural adaptations have in some measurable way aided in survival or militated against it (Diamond, 2006).

The year 1739, when Hume wrote of humans' personalities being "bundles of perceptions"²² (Hume, 1981), is indeed quite remarkable when one considers that this idea is still debated some two hundred and sixty six years later (Kukla, 2001). As an early inductive positivist (Howson, 2001b), Hume's assertions that all complex perceptions (further divided into impressions and ideas and the occurrence of the resultant associationism which forms the basis of all intellectual operations) are built up from simpler perceptions and can be analysed into simple components is testament to his atomist approaches to the study of human ideas and sensations (he was influenced by Newton's inductive reasoning) (Capra, 1983; Delius, Gatzemeier, Sertcan & Wünschler, 2000; Leahey, 2000; Mautner, 2000; Midgley, 2000; Oldroyd, 1986; Porter, 2004; Worrall, 2002). He thus inspired future logical positivists (Newton-Smith, 2001a; Ray, 2001) in which they recast his ideas into what is commonly understood as analytical (*a priori* experience-independent truths) which could not be proven invalid (Hempel, 1983; Turner, 1967) (see chapter 4 in which the discussion on mathematical axiomatisation was shown to be seriously flawed as formal mechanism of logic and in which the realm of the *a priori* as Platonic is discussed) and synthetic arguments (*a posteriori* experience-dependent truths) which are contingent on experience and so can be validated (Ernest, 1998; Fetzer, 1993, 2002; Heyes, 1989; McGuire, 1989; Quine, 1993; Turner, 1967). However, this strict materialistic monism²³ may be misleading (it being a more general label of several materialist oriented theories of mind; Churchland, 2000) since type-type identity theory as advocated does not necessarily analytically reduce mind to brain (in other words no independent status is granted to mental states although they are not denied; Eccles, 1989). Rather, it maintains that the two should be reduced to an equivalent class of property classifiable under different vocabulary (Maslin, 2001) which, as Watson (2005) states, may be the very problem and as such, "many psychological concepts, because they originate in language coined before the rise of science, are now outmoded" (p.42).

Perhaps our vocabulary is conceptually biased and inadequate to contain what is now understood to be the more correct meaning (Hall, 2000). Encompassed within the mind-body issue is a subsidiary issue of nativist and cognitivist views on language capacity and propensity.²⁴ After all, the nativist assumes a physicalist notion of language acquisition as opposed to the learning acquisition model posited by the cognitive learning school (Karmiloff-Smith, 2000). That language is unique to homo sapiens-sapiens however (as intimated by the nativists) is at this stage far from being resolved (Mason, 2005). This is evidenced by recent research detailing the region in the macaque's brain which controls jaw movements (serving as a direct homologue to Broca's speech area in the human brain). This finding is cited as evidence against novel neural pathways evident in only humans (Scientific American editorial, 2005; Petrides, Cadoret & Mackey, 2005).

The ceaseless debate and ongoing wrangling between opposing schools of nativist/cognitivist thought has however shed light on what was in the past an insurmountable and intractable problem. The jury has as yet to deliver their verdict on the matter once and for all, though it would seem that the nativist position is firmly entrenched in its place for the time being (Casti, 2001). Language, as developed within human beings, seems at this juncture in scientific history to indeed be a product of genetically mapped instructions. Although the exact details as to a "mechanism" (the Chomskyan view, Vosniadou, 1996a) within the brain has yet to be incontrovertibly detailed and accepted (it would appear that neurological underpinnings of this language mechanism is in fact a diffuse product of various coalescing features of the plastic brain; Casti, (2001); as well as a distributed as opposed to a unitary mechanism; Szentágothai (1989)). Firmer understandings of these and other issues can better be advanced utilising more empirical reductionist accounts. So-called "bridge principles"²⁵ or mapping relations (Looren De Jong, 1995; McGrew, 1997) are necessary in order to link psychological concepts with their physical counterparts and at each bridge

²¹ Damasio (2003) takes this idea from Spinoza's The Ethics Part II and is an idea dating back to the mid seventeenth century (Mautner, 2000).

²² Curiously, compare this to Binet's "bundle of tendencies" statement referring to humans' complex make-up, quite contrary in basic philosophy to what Hume proposed (Wolf, 1973 in Sarason, 1976).

²³ Perhaps the term "naturalist" would be better utilised here as opposed to monist.

²⁴ This hierarchy may be faulty to some as it is not necessarily the case that the language acquisition controversy can be neatly placed within the mind-body debate. I have, however, chosen to locate it here for the flow of the argument.

²⁵ Meehl (2002) states that bridge laws are themselves only postulates or an "operational link" and by inference it might be stated that these principles too are at the mercy of theory appraisal as with any theory.



on this “intertheoretic reduction relation” the disparate vocabularies will hopefully be connected (Bickle, 2003a; Kukla, 1995a).²⁶ As Lichtenstein stated over twenty years ago, cognitive psychology could find an ally in psychobiology as the mind-body identity theory was widely accepted by the two serving as connecting link (1980). Explanatory unification is the resultant future consequence with the hypotheses, postulates and laws of the one theory being reduced to those of the other theory (Churchland, 1986;²⁷ Turner, 1967) (such inductive realism will be discussed in chapter 3 along with intertheoretic reduction at both inter and intralevel theory explanation, see section 3.6). “The bridge laws establish connections between the two theories’ predicates, providing grounds for the explanation of the upper-level theory and for the revelation that its entities are nothing but combinations of lower-level entities. Thus, in principle at least, the lower-level theory can, allegedly, replace the upper-level theory without explanatory or ontological loss” (McCauley, 1998, p.613). This bridging principle notion is highly contested though and it would be best to consider counter arguments so as to allow the debate to continue in a fruitful scientific manner (Howe, 1996). This type of theoretical unity concerns the amalgamation and explication of a number of observations via a limited set of explanatory mechanisms, referred to as causal-mechanical (Schouten, 2003) where the same explanatory phenomena can be utilised to explain aspects at different levels of analysis (Finlay, 1997).

Of course this take on inter-theoretic reduction is a nomological-deductive one²⁸ (Ettin, 1984) (and draws its main architecture from the logical positivists such as E. Nagel for instance and assumes that higher and lower level theories can be commensurate²⁹). It stands in stark contrast to models presupposing that such reduction between lower and higher level theories will be incommensurate³⁰ as exemplified by Feyerband and Kuhn for instance, even though there are differences in the meanings attributed to incommensurate theories utilised by these two philosophers (Khalidi, 2001) or will founder due to a fundamental loss of information due to “translation” errors; i.e. when terms of the one theory are translated into the terms employed by the other (McCauley, 2001; Shweder, 1986). For it might be the case that not only do the different approaches utilise different languages but they may also refer to different phenomena (David, Miclea & Opre, 2004). Kukla (1995a) also cautions the reader to the fact that psychological reductionism as discussed above is based on “sheer conjecture” unlike some natural science counterpart examples of intertheoretical reduction (p.213); but unification via reductionism does not necessarily entail eliminativism (Barendregt, 2003) as already highlighted. In a manner, bridge principles can be equated with correspondence rules used for theory construction but there is as yet no systematic manner devised in which such rules can be sought in psychology between theories proffered and the data observed (MacKay, 1995). That this stance on mind-brain is the last word or that it has in any manner sufficed as adequate is of course a premature deduction (Gray, 1989). What is being emphasised here is the leaning towards this rendering of what constitutes the mind/brain. The crux is whether the theoretical explanation at both the neurological and psychological levels can be logically reduced to a single theory; and biting, Maddox (1998) states that it was inevitable that psychology was to become the handmaiden of neuroscience utilising “fancy names” (p.278) such as cognitive science to demarcate this change. Doubtless, his view can be contested. The fact that mainstream philosophy of mind remains indifferent to the findings of neurosciences in general (Bickle, 2003b; Mundale, 2001) is problematic and a main reason attributed to this indifference is the premise of reduction, a method followed within the neuroscience enterprise. It is doubly perplexing in fact when one considers that William James himself proposed physiological mechanisms underlying memory and behaviour as far back as his 1890 treatise on “Principles of Psychology” (Rosenzweig, 1998).

²⁶ Compare this to Hempel’s (2000) bridge principles or connective statements which are necessary if reduction from one area of study is to be made to another. Although not quite the same meaning is implied here, Wilson (1999) posits four bridges or links that can be and are being spanned across the social and natural sciences divides, namely links emanating from the disciplines of cognitive neuroscience, human behavioural genetics, evolutionary biology and environmental sciences. The first two are of particular importance within the area of intelligence research and are ever-expanding areas of focused research efforts in this realm within psychology.

²⁷ Churchland (2000) however states that bridge principles will never be found because they will not be bridging anything as such, seeing that naïve psychological explanations are invariable wrong; “as the eliminative materialists see it, the one-to-one match-ups will not be found, and our common-sense psychological framework will not enjoy an intertheoretic reduction, because our common-sense psychological framework is a false and radically misleading conception of the causes of human behaviour and the nature of cognitive activity” (p.489). Most philosophers, states McCauley (1998) are of this opinion where “high level” theories proffering extreme reductionism within developmental studies for instance are not yet within our reach (Schlesinger, 1995).

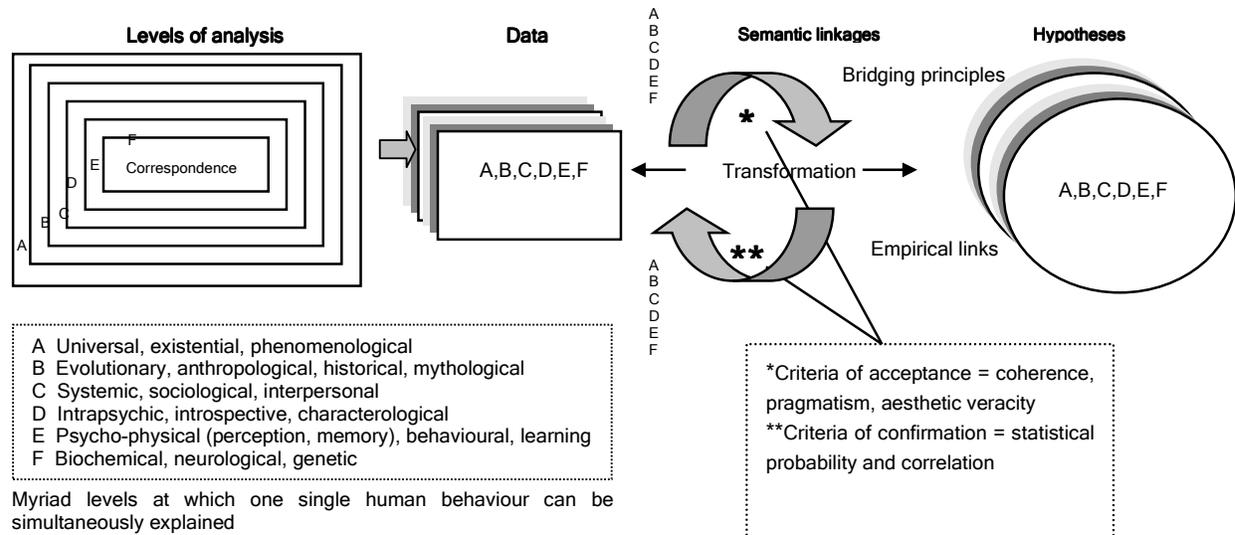
²⁸ Schouten (2003) states that one cannot necessarily equate deductive-nomological and causal-mechanical reduction systems of explication.

²⁹ In terms of meaning variance as well, for just as various same-level theories may utilise words in a different way so too might lower and higher level theories. The invariance of meaning was considered by the logical positivists as a given, but later advocates such as Feyerband, Kuhn, Hanson and Toulmin advocated that meaning is in fact variant across theories (Kordig, 1971). One can see the utility value of ascribing the same meaning to the same terms used within various theories; i.e. when physicists talk about gravity it is assumed that the meaning inherent in this term is invariant (even though the concept itself may undergo change; Chown, 2006). Can the same be said for psychology and assessment? The word intelligence for instance can hardly be said to be meaning invariant. Of course it might well be that a set of facts or data can (and should) be explicable by more than one theory (Copi, 1972; Faust & Meehl, 1992; Lamal, 1988) and hence the need for theory appraisal criteria.

³⁰ Of course there is no “winner” as such between competing theories as these theories are operating at different levels and should be viewed as co-evolutionary rather than as competitive (McCauley, 2001). The one discipline (neuroscience) does not seek to eliminate the other (psychology) but seeks to work together (Bechtel, 1988; Van Strien, 1987). Moreover, theories need not be revolutionary in nature to supersede mainstream theories (normal science period activity) but can, as Xiang, Anderson and Barker (1998) state reflect slow accretive changes permeating the realm of research.

Following from this train of argument and indeed logically consistent with the author's leanings in this regard is the parallel issue of dualism. If the aforementioned is doubted on any grounds, it is reiterated that no such dualism is warranted in order to explain functioning at any level and although not within the purview of current scientific progress, it is considered merely a matter of time before much else besides the mind/body problem receives even greater clarification. That there exists no such concept as "thought" within the brain reflects a naïve view but that the concept of thought should be placed in any category other than the category to which "neuronal activation" belongs is the essence of the type-type identity theory. As consciousness is often seen as the pivotal mystery within the mind-body debate (Ludwig, 2003; Nagel, 1981), it is this issue on which attention is next focused. Etting (1984) offers a succinct overview of the "levels of description" analysis and it is here diagrammatised in figure 2 as it illustrates much of what has so far been discussed.

Figure 2 Relations between various fields of knowing and validity criteria involved (Etting, 1984, p.217)



2.4.5.3.3 Consciousness

The statement that "social, psychological and cognitive sciences remain stuck with pre-scientific words and concepts" (Watson, 2005, p.44) may be contentious but a ring of truth resonates as Watson states furthermore that 'consciousness' among other concepts is in fact an imprecise term to describe what we think to be whatever it is consciousness is. As Blakemore (2000) emphasises in this regard, "so far we are not getting on very well" (p.221), an echo of Zimmer's (2005b) contention that such terms "have a way of slipping around in the semantic mud" (p.288). Consciousness is a term first used in English in 1601 by none other than Francis Bacon (Barrow & Tipler, 1996). That 'consciousness' may or may not be an idea worth pursuing or experimentally validating conveys the unease with which researchers are currently burdened (Plotkin, 1997; Singer, 1999). The idea that the concept and its innumerable meanings and interpretations (Deacon, 1997) are intractable does not warrant researchers denying themselves the pleasure of in fact deciphering all that it entails, however the fact that homo sapiens-sapiens might not be *able* to deliberate beyond a very rudimentary level on this issue, is telling of how enclosed we are within our own system (or within our own heads/brains/central nervous systems).

That we might never be able to escape this prison of confinement and hence do justice to a more fully explanatory model has not escaped philosophers and neuroscientists alike (Dennett, 1995; Pinker, 1997). This line of thought is rather disconcerting to say the least and will not do if humanity is to proceed along a course of self-discovery. Perhaps then, a change of vocabulary or concept category is necessary, but we will not be able to deter the fact that consciousness (in whatever terms it could be described) will haunt us. Neural correlates; a one-to-one isomorphic mapping; psychophysical parallelism;³¹ a ruthless reduction (Bickle, 2003a; Michell, 1999; Wakefield, 2001) of mental to physical may not always be a palatable exercise for some. The

³¹ Considered as a prejudice of our time by Gödel and Wittgenstein (Wang, 1995).



extreme of this approach is evidenced by the writings of Penrose (1994)³² (Horgan, 1994) for instance in which the very matter which allows for consciousness to emerge is what makes it untraceable - an epiphenomenon which is the cause but itself cannot cause anything. Our "conscious thoughts reveal to us so little of what gives rise to them" (Minsky, 1988), but it is believed by this author to be the best of the current repertoires of engirdled approaches on offer as this view is underscored by Ramachandran and Blakeslee (1999) whose sentiments echo those already alluded to, "I do think there is a new way to study consciousness by treating it not as a philosophical, logical or conceptual issue, but rather as an empirical problem" (p.228) and as has already been evidenced just such reductionistic goals have been achieved within a psychological framework (Kandel & Squire, 2000).

The view that physicalism is only good for explication at the functional and structural levels (Chalmers, 2003) is construed as an affront to the experimental approach, for in trying to shed light on a matter as difficult as this, any attempt from this side of the debate is considered a step forward. Consciousness³³ has always been enveloped within some sort of framework attempting to add credence to the concept and as such has travelled alongside structural (componential), functional, behavioural,³⁴ phenomenological, cognitive (computational) and systemic emergent psychological movements which pepper the historical progression of psychology (Maree, 1995; Spiker, 1989) and has been studied from a plethora of viewpoints and angles.

That this empirical problem can be studied from a evolutionary stance is perhaps an even better programme as it allows researchers to understand the gradual (not teleological³⁵) progression of how and why this state of affairs (consciousness) arose in the first place (Damasio 2003; Dennett, 2003), much akin to the value that intelligence confers on behaviour (Johnston, Partridge & Lopez, 1983). In this regard behaviour and consciousness is viewed as an emergent phenomenon (Horgan, 2000) or property of the brain which allows for a more accommodating adaptation to the environment. Although precisely how this is done at the neuronal level is not yet known, (Grossberg, 1997), which is a notion espoused by the functionalist school of psychology advocated in particular by William James³⁶ (Maree, 1995). This is a start towards a solution, an effort that will garner many specialists' views, opinions and research results for many decades to come.

2.4.5.3.4 G-dominated vs. multiple intelligence (MI) – dominated leanings towards the understanding of intelligence

A more sensitive, contentious, emotionally overwrought, discordant and dissenting and at times adversarial issue is hard to locate within the annals of intelligence literature and it is for this very reason that at the outset of a study such as this, it is considered prudent to particularise the author's affiliations regarding *g*.

The issue of *g* as presented briefly here is of course an issue more complex than when it first made its appearance in scientific journals in 1904 (Gould, 1997a;³⁷ Spearman, 1981). Moving beyond the original confines of *g*-related research, newer models now exist which, although fraught with their own problems, offer a less restrictive conceptualisation of what *g* in fact entails (Bowman, Markham & Roberts, 2002) and a simple straightforward avowal of *g* is certainly not what is intended when leanings are said to go one way or the other. As performance-based tests become increasingly used in many settings, *g*, as originally conceived of during the 20th century, will play a less dominating role in assessment (Sternberg, 2003a) not to mention the fact that *g* is not unlike the definitive feature of intelligence. That it poses no clear definition (Horn, 1988) but to conform to current dominating views of *g* as understood within the psychometric community, one could perhaps state that *g* or GMA (general mental ability) "is a general information-processing capacity and is extracted as a general factor (the first unrotated factor) from a battery of specific ability tests" (Ones, Viswesvaran & Dilchert, 2005, p.431). It represents "decontextualised reasoning ability" (Brody, 2005, p.492). *G*-supported research and its rivals periodically swing from the descendant to the ascendant (Stankov, 2005b) and context (research climate, funding, general views permeating the field) by and large dictate when and where it is

³² Who explicitly states that he is *not* identifying consciousness with quantum mechanics per se; rather "it is my view that solving the quantum measurement problem is a *prerequisite* for an understanding of mind and *not at all* that they are the same problem. The problem of mind is a much more difficult problem than the measurement problem" (1994, p.331).

³³ In whatever terms it has variously been described throughout the ages and throughout cultures.

³⁴ A framework where it was out-and-out denied.

³⁵ Arguments referring to "causal explanations in which some of the antecedent conditions are motives of the agents whose actions are to be explained" (Hempel & Oppenheim, 2000, p.64). Evolution as such is not going anywhere for instance, "it" is not moving towards a "greater" or "better" design of the brain. If it so happens that the environment were to somehow favour those with smaller brains then it might well occasion that smaller brains become the order of the day; human brains did not enlarge *because of* anything, those humans evidencing larger capacities were favoured and as such reproductive success enabled generation of those individuals' genetic material (Gould, 2000a).

³⁶ Very prescient, as this is often considered a rather modern idea.

³⁷ As much as this author is a fan of Gould's prodigious output it cannot be ignored that his original 1981 edition of "The Mismeasure of Man" itself came in for much criticism (cf. Carroll, 1995 and his legitimised criticisms; Deary & Smith, 2004; Jensen, 1982a (partly in an attempt to defend himself from uncritical attacks from Gould)). Gould's philosophy concerning human ability and growth is a like-minded concern and as this study has as its main subject dynamic assessment this further supports the authors leanings too in the direction of human change and growth. However, alternative philosophies as espoused by eminent scientists such as E.O. Wilson and A. Jensen for instance also cannot be ignored and it is with the latter that this liberal philosophy is attenuated (cf. Jensen's 1982a rebuttal to Gould's "attack").

emphasised. A distinction needs to be drawn between what one considers a general g as implied by an all-encompassing, ever-present general level of functioning that can be explained at even the most rudimentary and naïve folk psychology point of view through to the very strict mathematical psychometric g envisaged and “found” by Spearman³⁸ and held up as evidence of a unitary element representing all there is to intelligence (Detterman, 1982; Irvine, 1987; Sternberg, 1991) and in onto the “microanatomic or biochemical features of the brain [which are] involved in the heritable component of g ” (Meehl, 1998, p.3). Spearman’s envisaged g as evidenced from his tetrad correlations was a Platonic attempt at revealing hidden structures of the intellect (Brody, 2005); the Platonist realm of investigation is important in this thesis for this very reason and is discussed later in chapter 4. G as evidenced by statistical decompositions and g as expressed by less abstruse means (academic vs. ordinary discourse) are both considered hereunder and within the view taken by the author. It is perhaps no co-incidence, or at the very least it should come as no surprise that Spearman is also credited as having partially developed the theory of true score and error score along with Yule, especially as the nascent development of correlations played a pivotal role within both g research and the true score model of mental testing (Crocker & Algina, 1986; Fischer, 1996; Ghiselli, Campbell & Zedeck, 1981). Assessing for a trait considered unchangeable (Spearman’s g) via a means of stability testing (Spearman’s test theory) can perhaps be considered two aspects tying up on tightly knit conception of what intelligence is and how best to measure it (Dennis & Tapsfield, 1996; Kyllonen, 1996). Is it possible to conceive of a past in which intelligence was not considered a unitary concept and for which subsequent alternative statistical techniques were devised to test for such multidimensional traits not regressable to g ? Fluid and crystallised g -based models and three stratum models all hark back to the ever-persistent g or some form of general factor³⁹ (Carroll, 1997b; Keith, 1997; McGrew, 1997). In the words of Deary (in press) “ g is a highly consistent and an interesting finding, g is substantially heritable, g matters in life”; concurring with McCallum and Bracken’s (1997) non-verbal intelligence tests which are founded upon the measurement of pervasive g . Note that what is being stated is not that fluid g is in fact psychometric g as is frequently asserted as well as argued for and against (Anderson, 2006; Benga, 2006; Blair, 2006; Demetriou, 2006; Kaufman & Kaufman, 2006; Voracek, 2006) but that this may be the case; this issue is far from any clear resolution. Would the establishment now be saddled with similar nagging issues if g did not haunt research results? This thought experiment may well manifest in an approach to change-assessment predicated on fluid g . Till such time avails of such techniques, it seems likely that mainstream assessment will have to continue its somewhat dubious role of foundational mechanism, upon which rests change-based philosophies of assessment. Currently, though, g -based and the more intuitively appealing multiple intelligence-based leanings are equally represented in the literature (Hopper & Hurry, 2000; Stankov, 2004).

That g has played the role of harbinger of the hereditarian-influenced views is evident especially since Pearson was granted a Galton professorship (and completed the work Galton had started on correlations; Hughes (2000); Kevles (1997)). Galton was considered the founder of formal testing (he coined the term “mental test” but seldom used the word “intelligence” and never formally defining it either), was the inventor of statistical correlation⁴⁰ and is considered the father of differential psychology (Ittenbach, Esters & Wainer, 1997; Jensen, 1998b; Millar, Millar, Millar & Millar, 1996; Nunnally, 1978). He was Darwin’s cousin and thus firm believer in attempting to mould intelligence to the same framework utilised in evolutionary mechanisms described by Darwin (no date; original 1859; Crocker & Algina, 1986; Kevles, 1997; Madsen, 1988; Sahakian, 1975).

Fusing statistical⁴¹ ideas from Quetelet (1849) who imported the idea of the law of error from astronomy into the social sciences (Gigerenzer, 1991; Lazarsfeld, 1977; Stewart, 1990; Stigler, 1999) and his cousin’s ideas on the favoured transmission of hereditarian traits alongside Locke’s associationism⁴² and ideas regarding sensation as key to knowledge as well as Fechner’s psychophysical ideas (thus introducing probability-based modelling and inference into psychology; Stigler (1999) and adversely

³⁸ Who later adopted a rather less severe strict g -view and opted for the specifics as well as general intelligence (Nunnally, 1978) but it must be noted that Spearman did indeed discuss a “hierarchy of intelligence” and included in this hierarchy specific intelligences, although the underlying global intelligence was evident to him from the correlations be obtained between four school subjects thus leading him to ascribe g as the reason behind high correlations between any two tests (Spearman, 1981; Thorndike, 1997).

³⁹ These models do not advocate g as such but the strata converge back on g as initial layered structure. As they say, you can run but you can’t hide! G is there...somewhere.

⁴⁰ Albeit in crude form. Galton was not only a statistician but explorer, anthropologist, inventor and of course eugenicist (Pickover, 1999; Simonton, 2003).

⁴¹ The history of statistics within the social sciences is quite a colourful one which will be discussed at length later on in this study (as it pertains to intelligence assessment). Of note here is that before Spearman utilised his approach to derive g , Wissler’s 1901 pioneering study was the first to utilise the correlation coefficient method and preceded him in so doing (his study did not find robust correlations between basic processes and performance in tertiary education courses; Detterman, 1979); although Pearson had already evolved the mathematical formulation for the technique originally thought of by Galton who supported Pearson in his work (Carroll, 1997a; Du Bois, 1970). The term correlation coefficient was first used as such in a paper by Krueger and Spearman in 1907 (Du Bois, 1970). The psychophysicists of the early twentieth century (for instance at Titchener’s founded laboratory in the United States) were not convinced of the efficacy and relatedness of the more developmental approaches favoured by Binet and to strengthen the blow delivered to intelligence testing in the early days, correlation coefficients yielded nothing substantial in terms of how simple psychological processes were related to intelligence (Brody, 1992). That correlations were fundamental in the construction of the research domain that is intelligence research cannot be over-emphasised (Detterman, 1991).

⁴² Compare to Hume’s ideas on sense-impressions in the brain which does constitute mind. From the very start of intelligence assessment the ideas and philosophies of Locke and Hume were very evident. Locke, of course, having conceptualised a framework of modern empiricism (Baker, 2001; Kukla, 2001). E.L. Thorndike was an early pioneer in the study of associationism which itself served as prelude to the behaviourist movement (Benjafeld, 1993).



influencing proceeding generations in terms of what it means to “measure” in psychology; Michell, 1997, 1999; see chapter 4), Galton came forth with what at the time can be seen to be indeed revolutionary ideas (however incorrect and questionable they has since proven to be) (Brody, 2004; Huysamen, 1980; Stankov, 2005a; Thorndike, 1997). Once again this unfortunate dichotomy forever separating issues along a supposed manifest divide often skews what beliefs belie certain staunch views. Dichotomies too pervade the sphere of personality research evidencing a grave injustice to an area of investigation which is already suffused with intractable problems which it has yet to sort out (but will most likely not, due to the nature of the subject matter which is perhaps better left to less contrived modelling of human behaviour within the behavioural genetics realm). Take for instance the numerous dichotomies associated with pathological source traits where one is assigned an index score allotting to a “side” on the theoretical (not empirical) continuum (Smith, 1988). According to the manner of answers, the degree to which items are endorsed and the pattern of repeated endorsements one is labelled as “low bored depression” if one has validly endorsed items eliciting thoughts about one’s behaviour that are considered “relaxed, considerate and cheerful with people” and quite the opposite in the case of “avoids contact and involvement with people, seeks isolation, shows discomfort with people”. The very first question that arises from this situation is how consistent is such a representation? This bears on its veracity as reflective of the “true personality”. There is as yet no definition for personality quite apart from the fact that we presume to measure it nor does there appear to be a similar g -type notion associated with personality (Stankov, 2005).

Time can often be a great equaliser and as such lower-performing individuals when confronted with subjects at school, may blossom later on in life much to the chagrin of higher-performing cohorts; reminiscent of Plotkin’s (2003) view on intelligence and genes; “intelligence is caused by genes; the consequences of intelligence, however, cannot be reduced to genetic explanation” (p.81). The interplay of biology and environment results in either a dampening of what genes might harbour or a veritable explosion of growth, the capacity of which is enticed by a novel and accommodating environment. Notwithstanding the positive changes occurring throughout life, the situation may equally often present with lower-performing individuals who will forever remain so. Such nihilism is not unfounded but perhaps another term will be better utilised and shall henceforth be referred to as realism. It goes without saying that should the author have ever felt inclined to espouse such dark ponderings such as a deterministic written-in-your-genes scenario a study dealing with dynamic assessment would never have been attempted. Dynamic assessment by nature and definition assumes (not just espouses) and presumes a basis from which change within an organism is evident. That this propensity to change never existed is anathema to the very core philosophy of this approach. Change is and always will be a possibility, this is not in question and that dynamic assessment is a vehicle or tool with which such change is guided is also not in question. What may be in question though is the very basis or predetermined layer of instructions evidenced in every organism that has ever existed on the planet.

A pre-wired written set of protein instructions forms part of such a basis and however malleable the phenotype (measured variance accounted for by the genetic differences between subjects; Lynn & Vanhanen, 2002 and what IQ strictly is, i.e. a phenotype; Jensen, 1998b) may be, it is questionable how plastic the genotype will prove to be.⁴³ The systematic decomposition of phenotypical traits to genetic and environmental traits⁴⁴ is of concern to heritability-of-intelligence supporters (Grigorenko, 2004a). Current biometric genetic theory utilises mathematical relations underlying componential measures assumed to be a continuous distributed trait, in that the effect of individual genes combines to form the genotype (McArdle & Prescott, 1997) which with appropriate genetic targeting research can result in phenotypical modifications (Silva & Giese, 1998). To pronounce that the speed of neural conductivity (the chronometric approach⁴⁵) is intelligence is questionable; that it may be correlated with intelligence and cognitive ability is not (Reed, Vernon & Johnson, 2004).⁴⁶ Many well-utilised intelligence batteries encompass mental speed tests (Danthir, Roberts, Schulze & Wilhelm, 2005) although it is noted that reaction time and psychometric speed research are not at all synonymous. To reify and raise intelligence as a construct (entity) to status not befitting it is likewise erroneous and very misleading and at times tows the line between justice and injustice (Ackerman & Beier, 2005; Bardis, 1985). It is an activity which Coulter (1997) accuses cognitivists of practising when construing mental and experimental aspects as neural. To likewise make manifest a mathematical construct such as g (Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg & Urbina, 1996) too may be misleading, yet this is not to gainsay the role that the general intelligence factor plays in more often than not engulfing lower-levelled hierarchical specificities (Daniel, 1997). This is particularly so when one considers the tools utilised to make g manifest, namely factor analysis, perhaps the key statistical technique utilised in intelligence research (Keith, 1997). The first general factor is almost always g which, after rotation, is still manifest as main source of dominant variation within other rotated factors; so in essence it does not disappear (Ree & Earles,

⁴³ This is so rapidly becoming an incorrect postulate and unfounded axiom of truth; the facility and tenacity with which gene sequences are currently being explored runs counter to the notion that one is forever “fixed” from birth. That this fixedness will remain so for very long is debateable. The question now arises as to what humanity will do with this information.

⁴⁴ Phenotype is the result from genotype-environmental interaction (Waldman, 1997).

⁴⁵ Which utilises easy tasks but has been evidenced to be correlated with intelligence the more difficult the task becomes; Beaujean, 2005 which is similar in nature to the usual findings on intelligence tests where more difficult tasks correlate to higher degrees with IQ than less difficult tasks with concomitant increases and decreases in the heritability quotient.

⁴⁶ More specifically such evoked potentials are often associated with psychometric intelligence (Eysenck, 1988) (as opposed to other conceptions of intelligence!)

1996). Problematic in early ventures attempting to discriminate biological and neural patterns via behavioural or “mental” assessments was the lack of fit between methodologies directed at different levels (Bigler, 1994). Progress on many technological fronts within the neurosciences is allowing for a more inclusive approach towards the study of intelligence from local levels of brain functioning (Duncan, Seitz, Kolodny, Bor, Herzog, Ahmed, Newell & Emslie, 2000; Willerman, Schultz, Rutledge & Bigler, 1994).

Belief in many types of intelligences is a comfort not afforded by other more *g*-dominated intelligence theories (which may view multiple “intelligences” more as forms of “talents” than separate intelligences per se and was propounded by E. L Thorndike over eighty years ago; Greenspan & Driscoll, 1997; Neisser et al. 1996) but to rest-assured within the boundaries of multiple intelligences can be misleading. Surely this take on intelligence allows all-and-sundry a place in the sun without leaving anyone remaining in the shadows, as it is hoped that no individual shall be devoid of even the slightest hint of a propensity towards some kind of intelligence.⁴⁷ A more altruistic appeasement within intelligence assessment will be hard to find. Is it the duty though of intelligence research to find a place for everyone? Or is it to further explore and explicate on the notions of what intelligence may possibly be and how it can most fruitfully be assessed and brought in to work for humanity in numerous beneficial ways? Just as intelligence as a concept is not a reified thing but a construct measured and made manifest through statistical methods, so too can one refer to *g* as just such a construct - the product of factor analysis; a construct which multiple intelligence theory does not deny but which seriously questions its explanatory importance. Moreover, this general factor seems to dilute across a range of intelligence-related tasks (Chen & Gardner, 1997; Ittenbach, Esters & Wainer, 1997; Sternberg, 2004c; Walters & Gardner, 1986). However one chooses to explain *g*, the point is that it seems to be evident in numerous assessments; assessments which are, some may advocate, really just the same type of measure and that it is not *g* one is measuring but ensuring the measures that are used converge to same construct as before, namely *g*. That it has proven to be difficult to obtain group factors that are not dependent on *g* regarding heritability of genetic traits within the psychometric model of course does not necessarily indicate a similar paucity when viewing the problem from an information processing approach (Plomin & Neiderhiser, 1991). This leads the argument to areas of methodological inquiry and the degree to which various levels of inquiry are able to handle these aspects. The author affirms her support for the existence of the notion of *g* (in its psychometric and biological correlates guise) despite contrary “evidence” which has become evident in the literature over the last thirty years. The context in which this research is conducted has unfortunately tainted *g*-supported research when in fact this area of study can very accurately and scientifically be researched and defended. Jensen’s (1998b) comprehensive portrayal of evidential arguments against *g* theory is briefly summarised below. Reproducing summaries in a study such as this is not a practice one could reflect on as advancement of the discipline, but the counter arguments offered by Jensen are tantalisingly succinct and support and further buffer the notion of the existence of *g* as psychometric construct.

Summarised Jensen arguments against *g* and interspersed authorial opinions

Verbal arguments

1. The specificity doctrine - in which it is argued that intelligence is really only what has been learned in specific contexts and that these contexts are biased towards the ruling majority view. Currently in vogue among liberal circles and certainly a view point to advocate when writing proposals for funding. This sentiment is harsh, yet rings true
 - a. *Intelligence as learned behaviour* - in which it is argued that intelligence is resultant from stimulus-response type scenarios. This extreme approach fails however to account for correlations between dissimilar test batteries (which *g* underlies) and such views invoke “learned behaviour in terms of its generalities” argument. This is not compelling least of all because there really is no way of proving this. Strict behaviourism is also unable to account for interitem correlations. Novel tasks are equally novel to an array of intellects and past learning can hardly be cited as evidential if none have been exposed to such novelty prior to testing
 - b. *Contextualism* - often taken in tandem with relativist thoughts on what intelligence supposedly is or represents. As defined within this approach, intelligence can only be understood from the context which houses it but this goes no way towards offering reasons as to the disparate scores within one context. In other words, differential psychology is null and void which we know not to be the case

⁴⁷ Which brings this argument down to the very fundamentals of how “equal” humans are in terms of intelligence (as a construct measured at a specific period in time with a specific instrument). Jensen’s (1980) “egalitarian fallacy” springs to mind, a fallacy he maintains incorrectly assumed equality in all respects concerning any trait or ability across all groupings and before this view is turned into a misconstrued one tainted with negative “socio-biology” ideas, the author cites Dawkins on the role of nature, which “is neither kind nor cruel but indifferent” (2004, p.11).

Mathematical and statistical arguments

These arguments assume a more robust mantle in terms of their sophistication and argumentation and cannot be as easily dismissed as has been done with those above.

1. Guilford's "structure-of-intellect model" - recalled by Jensen as a pseudoscientific attempt at discrediting Spearman's g . G in fact underlies his numerous abilities which he erroneously suggested did not exist. *A priori* presumptions surrounding the non-existence of g resulted in applying methodology specifically attuned to not locating g
 - a. *Zero correlations between abilities* - Guilford's conclusion reached about non-significant correlations between cognitive tests has since been revised as a number as problematic issues have since surfaced including sampling error, restriction of the range of talent and measurement error. No scientifically valid evidence presents when Guilford's research conclusions are scoured for "evidential" claims
2. Sampling theories of Thorndike and Thomson - Thorndike's rendering of g did not gainsay the views of Spearman entirely (Thorndike reached these conclusions before Spearman's 1904 publication) but considered g as manifest linkages made between what, at the time appeared to be understood by Thorndike to be "bonds" or synaptic connections. Different test items would result in various bonds' being activated and varying tests would activate similar bonds hence the connection between the tests. People would thus have different bonds and inherently more or less than others. Hence, a connectionist understanding differed from the unitary understanding of explaining away g . Thomson was sought after and was offered a position by Thorndike and his main contribution to the g controversy consisted of his mathematical and statistical formulation of Thorndike's contention. Of importance in this argument is Thomson's convincing support for the notion of an extractable factor, such as g , but to infer the notion that this represents a unifying mental construct cannot be logically nor scientifically deduced as necessary concern. Thomson's rendering of sampling theory as underlying common factors cannot however be falsified. (Assuming for the moment that one subscribes to the tenet of falsificationism; see chapter 3).
 - a. *Cognitive process theories* - relating specifically to elementary cognitive processes which form part of many cognitive tests assessing for speed and accuracy of information processing as loading highly on g . Jensen's stance on such elementary tasks research and its implication for g is understandably contentious
 - b. *Behavioral repertoire theory* - cannot be accused of discounting g but it's main advocate, Lloyd Humphrey, construes g as general intelligence resultant from Thomson's sampling theory but utilised phenotype as unit and not genotype. Jensen does not argue against this as he cites this theory as not denying g 's veracity but merely offering an alternative angle. G , according to Eysenck (in Jensen, 1998b) cannot be defined in terms of its consequences (underlying structure) but rather defined as that which results in such underlying structure. This harks back to the nature of the statistical construct which g appears to be in terms of this argument (g , as is known, is also considered as more than a theoretical construct)
3. Cattell's theory of fluid and crystallised abilities - factoring fluid and crystallised intelligence in support of two general factors in contradiction to a general factor has led to much research into the nature of fluid and crystallised intelligence. Yet, in essence, gf accounts for gc variance. The correlation between g and gf has resulted in conclusions affirming their "one-and-the-same" status. More large-scale research is however needed to attest to these findings
 - a. *Width and altitude of intellect* - E.L. Thorndike's conceptualisation of the breadth of knowledge (information known) and height of knowledge (complexity or difficulty of items) were originally conceived as similar to $gf:gc$. Not surprisingly they are in fact correlated which accounts for some constant underlying both width and altitude
4. Multidimensional scaling and Guttman's radex model - which, as with others above, has been incorrectly cited as not supporting g , merely due the label's omission. The spatial rendering of the closeness of various tests to one another represents their relations and suggests analogous g -like characteristics.
5. Gardner's seven "frames of mind" and mental modules - Gardner's theory of multiple intelligences is an instance of intuitively appealing framing of intelligence theory within the larger discipline but as yet remains unconvincing in its evidence to buffer its claims. Abilities and intelligence are not synonymous constructs and the empirical evidence surrounding the claims are scant. Thus it cannot be leveraged against g
 - a. *Modular abilities* - Jensen is particularly incisive in his arguments regarding the neurological basis of modular abilities stating once again that g is a statistical construct accounting for positive correlations among tests and need not be invoked as supportive for neurological unity. That g may or may not have a physiological basis is another question altogether. A third order factor resulting from individual difference research and neurological modular bases of intelligence are completely different aspects of investigation
6. Sternberg's componential and triarchic theories of "intelligence" - is cited as supportive of g and is in fact contrary to what popular understandings may allude to. G is in fact part of this theory when more closely scrutinised.

Differing methodologies mandate other techniques not in keeping with other models, for instance the information processing approach vs. the psychometric approach towards the study of intelligence. Spearman did not contend that g was a biological modular capacity located within neural substrates, nor did later psychometricists (Glymour, 1997), notwithstanding research which attests to g 's location within a single cerebral location namely the frontal lobe and it's other main explanatory role, that of working memory (Deary & Smith, 2004; Hambrick, Kane & Engle, 2005). Many British contenders vie for the explanation of why g will for



quite some time be an aspect of intelligence assessment (Brand, 1985). Perhaps it is timely to venture forth what might be considered an unpopular sentiment and to echo the notion of what has already transpired above, that the term “intelligence” is an outmoded and ill-defined concept with which to pursue research on this front (Jensen, 1994). Practitioners and researchers are saddled with somewhat of a conundrum: intelligence testing assumes that what is being tested is intelligence but intelligence is so multifarious that in order to measure anything one needs to know what one is measuring, and in so doing the whole is broken down into componential parts and re-integrated back into the whole again, leaving one at times none-the-wiser for what has transpired. Hence, long before testing is a consideration it would make sense to first have some idea as to what the instrument will measure (Hunt, 1994).

Perhaps one can reduce the concept of intelligence to an equivalent of neural efficiency or as a product of networks of cortical areas (Davidson & Downing, 2004; Eysenck, 1986; Lashley, 1930;⁴⁸ Newman & Just, 2005) or some such physicalist account - as has been done with the terms “consciousness” and “mind. If this appears too over-simplified an approach at the very least the methodologies may well be complementary (Neufeld, 2002) in their common strivings towards doing justice to understanding more of what the concept “intelligence” has to offer. To reduce intelligence to biological and neural levels is rather similar in nature to the quandary in which the study of consciousness is found, and attempts through the ages to do so have often met with little success (Matarazzo, 1992; Schafer, 1982) even though comments such as those by Callaway (1985) evidence contrary sentiments “the fact that intelligence in humans has a biological basis seems too obvious to merit serious discussion” (p.223). Perhaps when an extrapolation is made from mind-consciousness-intelligence through to “psychology” a similar sentiment accompanies the attempt to try and explain what in fact is meant by psychology - another outmoded word unable to wrench itself away from embedded meanings which has since its formal founding as a science been unable to stand alone as a method of investigation into what makes humanity human.

2.4.5.3.5 Emergence (irreducibility) vs. reductionism

This particular aspect is of great importance as the issue is wholly intertwined with the issues of consciousness, nature-nurture, intelligence and mind-body problems. Causal inference as a goal of science is very closely aligned with the reductive imperative,⁴⁹ which seeks explanations as its most mechanistic level (Ryan, 1970; Von Bertalanffy, 1970). Radical reduction may well lead to answers to questions focusing on the very specifics of an issue and were it not for this scaling-down of reality, little would be accomplished in the understanding of how things work at a very limiting level. For instance, treatments for Alzheimer's patients benefit in profound ways due to work conducted at this reduced level (Lozano & Kalia, 2005; Rolls & Treves, 1998; Schmiedeskamp, 2000; Zimmer, 2005a). Bunge (1985) offers supportive arguments for the case of reductionism within psychology emphasising that the case of reductionism is not really eliminative as such (Barendregt, 2003) but rather one of integration.⁵⁰ Reduction of mental to neural does not mean that psychology becomes physiology; rather

- Neuropsychology contains information not only about the neurophysiology of behaviour but also information on the neural systems and functions including both environmental and genetic influences on behaviour
- Traditional psychological findings guide neuropsychology and neurophysiology; for instance by understanding the perceptual system and how it relates to and is related to the whole organism
- Neuropsychology unlike neurophysiology does not study humans detached from their social environments⁵¹
- Neuropsychology is not a branch of neurophysiology but rather a parallel endeavour into the understanding of human behaviour via data from the biological branch.

From what has transpired above, it is hardly surprising that the author affiliates herself with a more reductive side to explanations of things unknown but this does not necessarily imply that a strict hierarchy of science is a given (Schouten, 2003), even though this sentiment was core to the logical positivists. Psychology has tended to monitor the behavioural level only but that this level should be studied in isolation from the neural and genetic levels is questioned (Merzenich, 1987; Parisi, 1996). Ideas and notions as to what constitutes behaviour (including growth and change - the subject matter with which these areas

⁴⁸ An instance which illustrates psychology's lack of momentum in certain areas of research and human understanding - once again affirming the notion that the field is too undefined as a whole and should seek to fragment before synthesis can occur. A contentious, unpopular and sentimental view perhaps, but a view which necessitates a constant re-evaluation of where psychology is currently and more specifically where psychology positions itself within intelligence assessment.

⁴⁹ In agreement with Segerstråle (2001) and in contrast to what is often taken to be the case, reductionism within psychological research and especially intelligence research does not imply biological determinism.

⁵⁰ The author maintains that integration here refers to specificities within psychology which are amenable to reduction via bridge principles and does not advocate that psychology as a discipline will become unified as such (Krech, 1970; Von Bertalanffy, 1970). Various areas are open to reduction but the entire arena of psychology is not and should not be seen as striving for unification as all the multitudinous agendas are simply too vast.

⁵¹ A new journal entitled “Social neuroscience” which will publish its first issue in 2006 “attempts to explain the psychological and neural basis of social and emotional behaviors in humans and animals. It is a new multi-level integrative analysis approach, rather than solely biological or social” and will feature articles “that examine how the brain mediates social cognition” and looks at the “role of the central nervous system in the development and maintenance of social behaviors” (<http://www.social-neuroscience.com/>).

concern themselves) could be profoundly altered in view of new evidence from other levels of description. This notion was being pondered within psychology as early as 1930 (Lashley, 1930) so can hardly be considered an enlightened view. Evolutionary psychology, for instance, seeks to describe human functioning from three integrated and mutually influential bases; namely, the neurophysiological, cognitive and adaptive bases (Fikes, 2001). Hierarchies of emergent intelligent functioning via the evolution of different brain regions are tantalisingly parsimonious (Pascual-Leone & Johnson, 2005). Seeking to explain behaviour from a biological position which considers humans' past and evolutionary cognitive development is laudable on grounds that we are, essentially, biological beings and more specifically our cognition is historically based within situated human collectives (the emphasis on anthropological models of mediation and cognition) (Toren, 1993). In fact Van de Vijver and Jongmans (2002) have already indicated the potential for a biological correlates approach within a dynamic assessment paradigm with Lidz (1981) having indicated the need for inclusive neuropsychology within assessment more than two decades ago. However, the resounding call from Budoff (1987a) and Entwistle (1987) that biologically based attempts at intelligence measurement (also concerning educational assessment) is doomed to failure due to intelligence being socioculturally influenced is heeded (Fodor maintains similar sentiments regarding psychology as a discipline; Crosson, 1985). Also, neurocognitive perspectives on learning theories are receiving more attention from within the education domain (Anderson, 1997). Researchers had, as early as 1949, considered the biological and neurophysiological foundations underlying learning ability noting that improvement in learning ability might be due to changing metabolic processes underlying neural activity (Eysenck, 1986). The interdisciplinary nature of some literature attests to psychology's contributions to neuroscience and vice versa (albeit not quite to the same extent). Yet far fewer educational insights are attended to in the works of neuroscientists (Friedman & Cocking, 1986) - an area lacking within the discipline and especially dynamic assessment, for it, as an applied version of psychology, would benefit tremendously from educational insights interwoven into neuroscience texts. What are the neurological benefits derived from instruction? This area of concern has yielded studies but it would be of interest to consider the evidence in lieu of dynamic assessment specifically as this manner of assessment is attuned to the transference of guided instruction principles. Viewing children's ZPD and neurological predisposition as moulding to the learning situation could provide a rich source of proving or substantiating the isomorphic claim of reducibility.⁵²

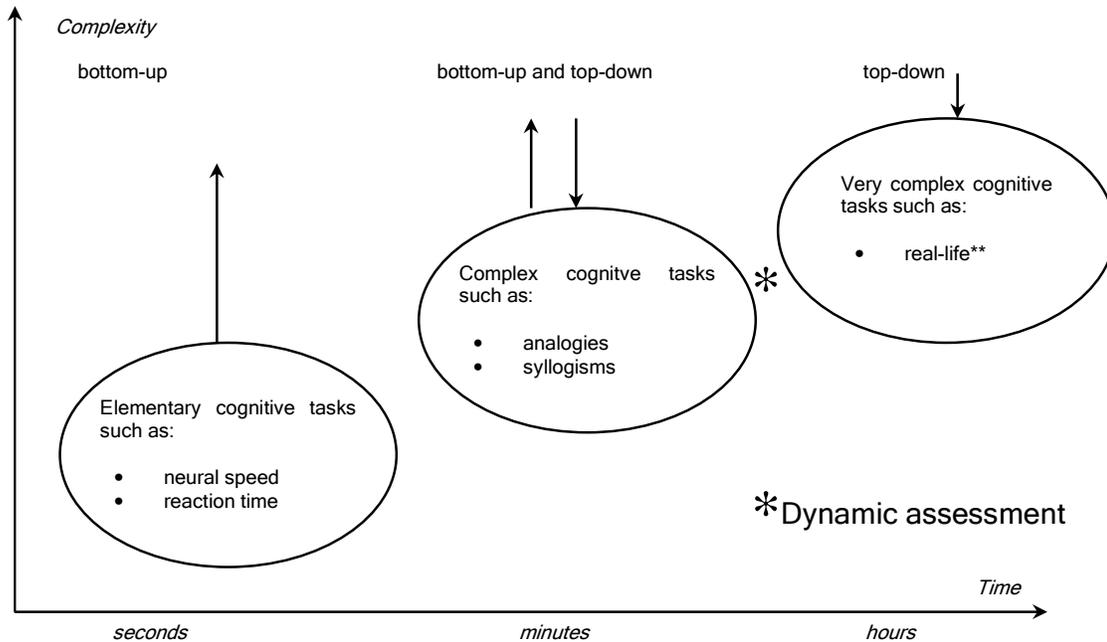
The idea of learning ability and its biological measurement is really not at all that new a concept! One cannot ignore the plethora of research into activity-dependent influences on the brain in which both external and self-modifying behaviour continues unabated throughout life (Smith & Katz, 1996) and neither can one casually dismiss results within the area of polygenic research (Haywood, Tzuriel & Vaught, 1992; Jensen, 1997). Dynamic assessment espouses such malleability so why not enter into reductionist research as complementary activity? Stott (1987) cites a philosopher's contemplations about intelligence thus: "the mind is what it does. Give a mind something to feed upon; give it something new to do; and it becomes a different mind" (p.217). The idea of a changing brain⁵³ brought about by environmentally impinging variables is starkly reminiscent of a plastic mind; a notion to be echoed a few decades later. Thus modifiability within dynamic assessment is almost perfectly attuned towards and should be quite amenable to more physiologically oriented research endeavours. The author maintains that the denial of anything emergent and contingent cannot be equated with reductionism although Gould (2004) firmly contends that in order to accept emergence from reduction it is necessary that simplified explanations at lower levels of explanation - read finer grained levels - need to be linearly additive so that one thing leads to another which leads ultimately to the whole. This, according to Gould, does not always work as explanations do not always proceed along linear lines. Historical contingency then is also a "bugbear" for reductionists, with facts accounting for much of the explanation after events but never occurring in the same way again thus making it almost impossible to form any sort of prediction of events). Scruton (2004) cautions against vulgar reductionism in which "the thing explained is identical with, reducible to, or 'nothing but' the thing that explains it" (p.29) and Sternberg (1996a) although lauding the explosion of biologically-aligned approaches towards the understanding of intelligence has issued a cautionary statement stating that it has yet to prove its link to straightforward applications in the intelligence assessment arena especially for instruction and learning outcomes. Nevertheless, the hope of seeking some sort of consilience between subject areas and methods may be in vein as is often the case when speculative areas are considered from more reductionist viewpoints. The narrowly focused jargon and area of expertise becomes so defined that the technicalities move beyond comprehension (each researcher's specific field that is) and is a view long espoused as overly simplistic especially since its modern inception in the works of Snow and Leavis (Schaffer, 1997). Suppes (2000) implies that by reductionism (at least in language) one moves towards increasing specialisation and divergence in place of the convergence (or unification of science; Bechtel, 1988) supposedly expected of this reduction exercise. Furthermore, such continuous reduction often serves no utilitarian purpose as is the case with the "reduction of computational mathematics to computability by Turing machines" (p.480) in which such reduction is irrelevant to most problems encountered in computation. But what other model is there? Ray (2001) maintains that the "core of the reductionist's faith rests on the assumption that it will always be possible to reduce all empirical statements to more basic statements with clear-cut observational consequences" (p.250). Arguing that reductionistic enterprises are the final answer to methodologies is absurd (Seager, 2001) and by leaning towards reductionism in no way implies rejection

⁵² How one is to go about this exactly is another matter altogether! Consider this recommendation a matter for the future.

⁵³ The author unabashedly nullifies the existence of mind and equates it with brain.

of emerging properties of systems (dynamic assessment, if to be placed on a reduction-emergent continuum, would most likely be placed nearer the emergent side). But in order to grapple with issues subsumed within ever-growing areas of specialisation, reductionism offers a place to start. Cognitive developmental frameworks exist which take into account information processing models and neuroscience which aid in the integration of theories of cognition and developmental theories (Goldman-Rakic, 1986;⁵⁴ Parent, Normandeau & Larivée, 2000). Figure 3 illustrates the level at which intelligence can be described for instance depending on the type of questions asked, nature of investigation and tasks required to do so. Dynamic assessment finds itself varying between complex cognitive tasks and very complex cognitive tasks (Nęcka & Orzechowski, 2005).

Figure 3 Dynamic assessment placed in a two dimensional framework of complexity of task and time limit (Nęcka & Orzechowski, 2005, p. 124)



** almost nigh impossible to test. How does one “test” for intelligence across and within life? Which tasks would be necessary to study (Lohman, 2005)? What version of intelligence theory would be utilised? Emphasis could be placed on neural speed, efficient processing of elementary tasks, biological processes and tasks successfully navigated in cultural and other environmental contexts. Bottom-up investigations are replicable but tasks are far removed from everyday reality; the converse holds for complex cognitive top-down approaches in which tasks are often beyond traditional psychometric practice and definition (Pretz & Sternberg, 2005). The question itself could well occupy many pages but suffice it to say that if complex cognitive tasks pose problems to intelligence assessment (Wenke, Frensch & Funke, 2005) there is little hope of securing answers to the former within the next few decades at least.

⁵⁴ “The pertinence of neuroscience to cognitive psychology and education is so totally direct and obvious that articulation of this relevance would seem hardly necessary” (Goldman-Rakic, 1986, p.233).

The brain is, after all, an emergent organ in the sense that global behaviour cannot be successfully predicted from the workings of specific sections within it. Yet, to properly understand the role that various areas play in this global behaviour, reducing many parts to smaller parts is deemed necessary but this is maintained only with due regard for the fact that the particulars within social systems and those of natural systems are not of the same logical type to begin with (Ryan, 1970). This issue becomes very problematic when discussions centre round social behaviour; for instance, in predicting behavioural outcomes in society at large, decision-making within economics and “laws” governing mass behaviour (often made sense of by natural science models in quite effective ways; Ball, 2005; Wilson, 1999). Reducibility as a method breaks down once systems are studied especially systems involving higher cognitive functions, especially studying within larger systems encapsulating the biological individual, the social and the cultural levels of description which are in many ways inseparable (Berry, 1996; Kriegler & Skuy, 1996; Rogoff & Chavajay, 1995).

The model of investigation is perhaps at fault and as such science as a method should be revamped or at the very least its theories of knowledge (Gergen, 1987a). There is no such model available just yet⁵⁵ (Gergen, 1987b) and so we need to proceed along the lines of working with the model that is available. A model, very much a natural science one, is a root cause of this malcontent in some circles and the rationale behind this is not difficult to understand. Wilson’s (1999) dictum of consilience⁵⁶ between the seemingly dichotomised areas of natural and social science investigation pleads for a reconciliation in which a discipline such as psychology can become fused with the studies of biology, economics and social theory in order to explain human functioning (although “demonstrating examples of consilience is not equivalent to proving the unity of all knowledge” cautions Kendler, 2002, p.501). Within this mammoth task, states Wilson (1999, p.226)

“lies the micro-to-macro problem, the ensemble of processes by which the mass of individual decisions are translated into social patterns. And beyond that, framed by a still wider scale of space and time, is the co-evolution problem, the means by which biological evolution influences culture, and the reverse. Together these domains - human nature, micro-to-macro transition, and the co-evolution of genes and culture - require the full traverse from the social sciences to psychology and thence to the brain science and genetics.”

How are we to cope then in finding answers to life’s problems, to intelligence assessment issues if all we have to go on is a model derived wholesale from natural science investigations? As Pauw (1993) states, when investigating natural science phenomena one does so from one level only - that of the investigator investigating the phenomena; however, social scientists have to contend with two levels of rule formation and understanding, that of the investigator and that of participant too; reminiscent of Winch’s logically inconsistent triad in which the three paragons (“core concerns”) of science (rationality, agency and scientific methodology) are incompatible primarily due to the “socially variable nature of rationality” (Gergen, 1987a; Turner & Roth, p.6) within the social sciences. Change at societal level will need to be instituted, changes in programmes at academic levels will need radical revamping and a new model for humanists will need to be developed. Where is this model and what does it look like? Philosophy of science has traditionally taken attempted answers and solutions to the puzzles facing natural science methodology and similarly attempted to adapt these programmes to the social sciences (Slife & Williams, 1997; Turner & Roth, 2003). That fragmentary parts of a system of social science philosophy may be found strewn across literature is evident but to date there is no successful wholesale transplant. In the meantime, what are researchers supposed to do? Hence the need to recognise the problems inherent in reductionistic studies and to conform to newer emergent views, especially within intelligence assessment. But this study has not as its aim a new model - just the adjustment of theories within the larger model as is.

2.4.5.3.6 Realist vs. relativist approaches towards research

The purpose of this study is not to detail the debate between realists and relativists within science and less so with their respective roles within psychology. However, one view that succinctly states the matter is that of Dawkins (2004) and his statements regarding the relativist problem within science. A “strand of half-baked philosophy” going by the “name of cultural relativism” (p.17) is perhaps harsh but is nonetheless considered a view towards which the author leans. Relativism has been seen by some as a viewpoint supported “by those impatient with the burden of sceptical argument” (Scruton, 2004, p.32). In essence the relativist view can be captured in this sentence: “the world as it is, and the world as it is perceived” (Harré, 1988, p.18). Relativist and critical views concerning social aspects of human life shed light on issues better dealt with from varying

⁵⁵ A suitable topic for another discussion perhaps.

⁵⁶ Originally William Whewell’s term in his 1840 treatise “The philosophy of the inductive sciences, founded upon their history” (Gould, 2004; Jencks, 2000; Mautner, 2000; Whitt, 1988). Whewell’s fusion of Kantian and Baconian views on the progress of science went out of favour in its own time (being heavily criticised by J.S. Mill) as it did not decree an exact line of progress nor did it acquiesce to Mill’s sentiments of exact empiricism (Brody & Capaldi, 1968; Oldroyd, 1986; Turner, 1967). However a return to Whewell after Einstein’s reformulation of Newtonian physics heralded a renewed look at the chances of any theory being successful, with Popper’s critiques in this regard being particularly emphasised (Agassi, 1981; Brody & Capaldi, 1968; Worrall, 2002). The Newtonian and Einsteinian conceptions of gravity and space-time curvature do not evidence similar results in experiments and hence Kuhn’s scientific revolution in this sense is applicable yet the two systems are mathematically similar when gravity is weak and thus physicists often work within both paradigms when the situation so necessitates (Thorne, 1995). Likewise within the social sciences can the same be said?

points of view. This is not contended as it has been often shown that certain social aspects are often explicable within the particular context in which the occurrence takes place. However, to place the progress of a science upon a relativist footing does not auger well for its progress simply due to the lack of progress that will result.⁵⁷ As Harré (2001) states, even though our knowledge of the world is not entirely independent of human concepts and constructions, the instruments which we use to measure and determine aspects do not behave in this manner unless “bolted to the world” (p.100). For instance De Jong, Bem and Schouten (2004) in similar vein, refer to an infinite regress of linear inferences (A results from B but B resulted from C and C resulted from D and so on), referring to the discordant debacles concerning the distinction of theory and data; can the two exist independently? Data needs a theory to guide perception (akin to the theory-ladenness view⁵⁸) and theory is informed by observation and so on towards a regressive spiral or infinite regress (more of this in Chapter 3). If all is relative and observations yield to relativist inferences then a formal discipline will be unable to flourish (it simply has nothing to hold on to); perhaps the realist take on the relativist notion of “nothing real exists” should be that “nothing unreal exists” (Krauss, 1997). In its strictest form, relativism dictates that all observations are relative over cultures, times, paradigms, views, practices, standpoints and styles (McAllister, 2001). However uncontested the acceptable forms of relativism are (a hammer is a tool relative to the function it performs), issues that centre on truth or epistemological relativism are more contentious and any inferences about human behaviour within social inquiry from a relativist framework is severely criticised by objectivists (realists) (Wagner, 2003).

Science as formally practiced and endorsed by formal communities is itself a relative system of rules to be followed, a view espoused by Kuhn (Doyal & Harris, 1986). The researcher cannot escape this inevitable spiral of relativism but the spiral has to stop spiralling at some point if only to allow some vague sense of what “really” exists and how best to pursue knowledge of this alleged truthful existence of “something”. Popper’s falsificationist⁵⁹ rebuttal allows for science at the very least to come closer to the truth in a piecemeal fashion (Bohm, 1997; Braybrooke, 1965; Gould, 1994, 2002; Hawkins, 1997) or “stepwise nature of progress” (Gould, 2003, p.205) in which failures are necessary preludes to success. Even though the version of truth is still relative or not known (Hilderbrandt & Oliver, 2000) this falsificationist approach is perhaps too strict an approach for psychology which, as yet, is “nowhere near that ‘ideal Popperian’ stage of theory testing” (Meehl, 1997, p.415). However science also progresses in piecemeal fashion via the principle of instance confirmation (a principle of induction which states that the more often a theory is evidenced as being true the greater the probability of its being true, a conditional probability⁶⁰ defined by successful confirmations or not; Chalmers, 1999; Howson, 2001b). Yet as Green (1995) states, given all the data in the universe a theory is either true or not and cannot be associated with probabilities, but this is the point of inductivism, we do not possess all the data in the universe - this only rests on a probability foundation and is thus not certain (Harré, 1988). Any finite set of observations can be explained away by an infinite number of hypotheses; Silvestri and Kose (2003) and thus one is not compelled to accept the conclusion (Weyant, 1987). Popper’s falsificationist progression of science does not always proceed as such within practice however and cannot thus be held up as the only criterion of scientific progress (Shadish, Houts, Gholsen & Neimeyer, 1989) and it cannot be ignored that each “falsification of a conjecture is simultaneously a confirmation of an opposite conjecture” and vice versa (Gardner, 2003, p.13). Kline (1998) has stated that Popper’s own theory is itself difficult to test! Knowledge, via the system of falsificationism is thus always provisional at best; a sentiment reflected in early statistical reasoning, upheld by the statistician Ronald Fisher, to be discussed in chapter 4 (Gigerenzer, Swijtink, Porter, Daston, Beatty & Krüger, 1990). This social constructionism of events and realities bodes well for more inclusive understandings of what reality is or is not and does not preclude areas of study as unscientific (Stam, 2001). Yet the social constructivists do need to offer adequate measures to remain consistent in their renderings and interpretations of constructs such as intelligence for instance (Borsboom, 2005) and educational theories (Collins, 1996). The seepage of constructivist thoughts into neuroscience also colours this field to a point where cogent arguments are offered to illegitimise such constructivist concerns stating that constructivism has failed in its long line of attempts at discrediting nativist accounts which accord the mind some power in establishing its own ideas without recourse to external interventions (Bickerton, 1997; Black & Greenough, 1997; Sporns, 1997), after all much cortical development takes place without external events evoking signals (Kennedy & Dehay, 1997).

⁵⁷ The author is keenly aware of the fact that the framework employed in this study has relativist tinges (Madsen being influenced by Kuhn for instance) but this relativist stance within the study merely serves to place and contextualise theories within their varying histories. The black-white relativist/realist stance is obviously not as clear-cut as outlined above, and cognisance is taken of the varieties of approaches between the two poles (Capaldi & Proctor, 2000).

⁵⁸ Incidentally, Meehl (1997) refers to this issue thus “theories entail observations, not conversely”(p.393). But consider the choice of research methods and analyses which already imply and apply psychological assumptions (Ridgway, 2000). More on this in chapter 4.

⁵⁹ Gould’s (1997b) lyrical notion for negative results or falsificationism is “nature’s apparent silence or nonacquiescence to our expectations” (p.123). Closely related to negative findings or falsificationism is the issue of unpublished research results which confirm by way of falsification claims made or hypotheses conjectured as being true; see appendix 1 in which the “file-drawer” problem is briefly discussed.

⁶⁰ Bayesian probability is what is being referred to here; the posterior probability of an occurrence is expressed in terms of the product of the likelihood of this occurrence and the prior probability associated with it (Howson, 2001a). A finite set of probability is firstly distributed among various hypotheses after which Bayes’ theorem can be applied (Brown, 2001; Gigerenzer, Swijtink, Porter, Daston, Beatty & Krüger, 1990) although this method has been criticised as being overly subjective (Percival, 2001; Salmon, 2001; Worrall, 2002), Popper maintains that it is the search for highly falsifiable or improbable hypotheses which steers science whereas Bayesian subjectivist conditional probability maintains the opposite! (Sober, 2001).

The representationalist school of test theory is one such model that constructivists adhere to when considering issues such as intelligence measurement (Borsboom, 2005). Theories of mental test scores are also verificationist-attuned not necessarily amenable to falsificationism due to the nature of their construction (see chapter 4). Realist, relativist and instrumental notions of the concept “intelligence” impinge directly on the assumptions imbedded in models of test theory and cannot be laid aside as mere speculation. If psychometric theory is to continue to adapt and acquire for itself a garb of modern progressive thought, attention needs to be turned towards the underlying philosophy of science predicating these theories (Borsboom, 2005). Realist conceptions of intelligence assume that intelligence is an identifiable construct which finds its construct counterpart in reality via test scores; in other words it is causally relevant and Borsboom (2005) is of the opinion that latent variable theory is just such an implementation of a realist account of test theory. Borsboom (2005) asks the pertinent question of whether intelligence has to exist for theories of intelligence to exist. Answers in the affirmative attest to realist interpretations of intelligence and those answering in the negative attest to anti-realist, instrumentalist, social constructivist, empiricist and logical positivist approaches (see chapter 3 section 3.2.2 for more on the instrumentalist vs. realist accounts of science progress and thought). These philosophical “ponderings” are not merely a distraction in the pursuit of psychometric validity, but are seminal to the work being conducted in the field. Researchers including Barrett (1998, 2000, 2001, 2002, 2003); Borsboom (2005) and Michell (1997, 1999, 2001, 2004, 2005) give credence and a voice to the larger issues plaguing issues of accounts of psychometrics within the philosophical realm.

Can one not tackle the issue thus with an example of colour and its perception? The colour blue might appear as purple to beings from planet Zog and reddish to beings from planet Goz and appears blue to humans on planet Earth, so then, the colour is perceived differently. However, what is being contested here is not the perception of the colour but that the colour blue can be measured at all. Blue lies at a specified wavelength on the electromagnetic spectrum (480 millimicrons⁶¹) and this is the case no matter who or what the perceiver is.⁶² Is this notion relative? No. Do Zogians and Gozians view the 480 millimicrons in different ways? Yes perhaps.⁶³ Different cultures may characterise different colours in different ways due primarily to cultural and category variations, but as to whether such *a priori* characterisation affects elementary colour perception is not clear (Sacks, 1995). It is acknowledged that cultural contexts do indeed greatly influence perception, memory, inference and the way language is used (Hofstadter, 2000). Indeed some might say that cultural contexts actually construct perception and memory (Nisbett, 2003; Nuckolls, 1998) much akin to the now little credited Whorfian language hypothesis (Reber & Reber, 2001; Ross, 2004; 2005). The role of culture in thinking is testimony to how influential culture as construct can be (Altarriba, 1993; Huysamen, 1980) especially in the early years before biological maturity (Cattell, 1963) and is important when mentioned later on in this study regarding the inadequacy of Piagetian cognition and affiliated stage development, at least as this pertains to change based IRT models. We are all human beings and as such all vulnerable to more or less the same sorts of things, we see in the same way, hear in a similar manner, smell and taste in more or less equal fashions (contact lenses, enticing hotdogs and the sounds of the Beatles would not attract so many people if the case were otherwise). Are these sensations relative? Perhaps to prehistoric man, the Beatles might come across as a cacophony of notes, yes; it is conceded that this is indeed relative. But its relativism is predicated on the fact that we all hear. This cannot be doubted. All ears hear. Is this disputed? Can reality then not be predicated on a similar argument of the fact that we can all observe? Yes, but we may observe “different” things, *but we observe* and it is this predicate to which reference is made when considering that science should be construed as realistic and not relativistic. The devil is in the details one might add, and as such these minor issues cannot be ignored by the realists. Meehl (1986, p.322) states in this regard “there is no reason for us to have a phobia about the word ‘truth’. The idea that you shouldn’t ask whether a scientific statement is true, separate from the anthropologist’s or the Hogo Bogo’s belief in it, because you can’t be absolutely certain, is a dumb argument”.

After all social constructionist and critical movements do not necessarily have as their antithesis notions of anti-realism (Stam, 2001) as both positions can be accommodated but the point at which agreement must be reached is often problematic. Evidential weight attributed to empirical findings differs across studies, times and locations. However, the fact that so many innovations surround humankind attests to this formal method’s (science) hierarchical progression from worse to better explanations and understandings of how things work (Mouton, 1993b) even though “science” itself may be construed as yet another religion (Slife & Williams, 1997; Wagner, 2003) or system of knowledge (“a sociological phenomena open to manipulation by power structures”; Brand, 1998, p.66) or “style” of insight and discovery (Gould, 2001b). It is no better than any

⁶¹ Reber and Reber (2001); Sternheim and Kane (1991).

⁶² Dennett (1993) dismisses this line of argument and refers to it as not only elementary but false. This line of argument rests on the nature of the reflectant surface thus altering the light as perceived by ourselves. But once this path is trodden upon, the author contends that the relativist argument is invoked much to her displeasure as it is now known that this type of spiralling argument is solipsistic.

⁶³ Colour construction by the brain can be approached in a top-down manner and although it holds sway as relativist counterargument, it cannot logically argue against the overwhelming influence of the bottom-up approach which evidences that colour need not be influenced by socially constructed knowledge at all and in some cases of colour blindness, colour can be generated by magnetic stimulation of certain cells (known as V4 cells which are specialised for responding to colour) (Sacks, 1995). So is colour constructed? Yes and no. Perceived colour may be relative but its existence as vibrating wavelength is most certainly not. Unless of course 480 millimicrons comes across as green in another dimension evidencing altered laws of physics. But the argument should rest here for the time being.



other method (Chalmers, 1999; Wilson, 1999), even an ideology of sorts (Shames, 1987a). The method is not to blame nor the knowledge that science produces but it is society that decides upon the use of that knowledge (Dennis, 1997). “I take it that the scientific method, of which so much has been heard is hardly more than the native method of solving problems, a little clarified from prejudice and a little cultivated by training” (Lewis, 1959, p.87) is yet another informal definition of what it means to practice the “art”, to which Lewis adds as an aside that measurement yet be a tool only for science and not to be defined as science itself, an aspect to take note of. This aspect of construing techniques as tools is relevant to the techniques and models employed within statistics where tools are not theories but instruments to these theories (De Leeuw, 1994). Weekly peer-reviewed journal articles posit findings and discoveries and follow “forward” moving trajectories in the sense of gathering more data to either support or refute hypotheses (although subversives have often cited instances in which the “progress of science” such as it is, really is quite arbitrary and is a view now commonly referred to as the “science wars”; Hertz, 2002; Nelkin, 2000; Radick, 2005; Segerstråle, 2001). The rebuttal to this is that the very core of this process is itself fraught with inconsistencies, biased reviewers, political interference and of course dictated in large by those with power and wealth.⁶⁴ It is at times considered mere institutionalised attitudinal beliefs (Brown, 2001b, 2001c; Campbell, 1986; Cornwell, 2004; Jungk, 1960; Nester, 1996; Pestre, 1997; Shames, 1987a; Stevenson, 2000) and is tinged with a critical psychology colouring of what it means to engage in scientific research, knowledge for beneficence or exploitation⁶⁵ (Austin & Prilleltensky, 2001; Fox, 2000; Sampson, 2000)? The social constructionism of mathematics and statistics has also been an issue of concern (Ernest, 1998; Hughes, 2000), an area which would seem least vulnerable to relativistic leanings but is of concern as both these foundational subject areas of are of prime concern to this study. Specifically with regard to measurement in psychology, Michell (2000) boldly states that in order for psychology to be given credence for its formalist stance as science, measurement was deemed necessary for political reasons. It was necessary in order for it be given admission to the scientific community and to benefit directly and indirectly from its commercial success as quantitative (see chapter 4). Scientific proofs⁶⁶ cannot be deduced from logical calculus alone as science (as has been shown by critical psychology) is value-laden (Mouton, 1993b) and intuition is often the “given” as starting point within logical argumentation, a given being a primitive which rests on its own assumptions as being unequivocal (Kukla, 2001). An argument, which Wittgenstein employed by maintaining that human perception would never be replicated within a formal rule-based system because any factual statement about the world would need to be derived from another factual statement and so on *ad infinitum*, hence there were no primitives as such that could serve as a foundation for cognition (Horgan, 2000). One has to start somewhere and failing to locate a starting point leads down a path of never-ending regress. Yet, this naïve folk psychology idea of a given, an understanding itself has run the gauntlet of expedient explanation; and “folk psychology has already been pushed way beyond its limit” (Wilson, 1999, p.223). The issues surrounding primitives or axioms are discussed in chapter 4. How is one to evaluate the “self-evidence” of a primitive postulate if it is impossible to prove such a primitive? (Hempel, 1983). The intuitive grasp for primitives is one point of departure within the constructivist school of foundational mathematics which will be discussed in chapter 4.

Relativists do not question empirical research because they fear that inferences are not “true” but because the conclusions are not final (and never can be if one is to extrapolate this argument to its fullest, for all times and places). Axioms would thus find no place within this philosophy as even accepted predicates are not objectively true but this is not to say that relativists are anti-realist in the sense of not acknowledging unobservables (Achinstein, 2001) as it is their interpretation that differs. Relativism has been construed as an edifice of policy that is mounted when moral and political issues are germane to a topic at hand and not as a logical issue as forms part of the traditional philosophy of science (Harding, 2001). That the philosophy of science is relativist in the first place is where the spiral starts to spin out of control. After all the relativist’s stance is relative too surely? Hence, his relativism is absolute and that, according to relativist premises, is not possible (leading to logical contradictions within the system) (Mautner, 2000; Robinson & Groves, 1999; Scruton, 2004), to which the realist may well reply “can we get on with it please?” In relation to the nature/nurture debate which follows next, Wilson (1999) contends that cultural relativism, in a manner of speaking, was able to rescue legitimate biological-basis-of-behaviour theories (by dragging misconstrued social Darwinism out of the mud) into a respectable area of study but only at the cost of “turning against the idea of a unified human nature grounded in heredity” (p.204).

The author has no legitimate gripe with the fact that issues are relative in their broadest sense, however, the argument rests: the realist take on science might not ultimately⁶⁷ be the best philosophical stance to endorse, but it’s as close as we are going to get

⁶⁴ Part Five of the edited volume by Gholson, Shadish, Neimeyer and Houts (1989) deals specifically with the social factors involved in the psychology of science.

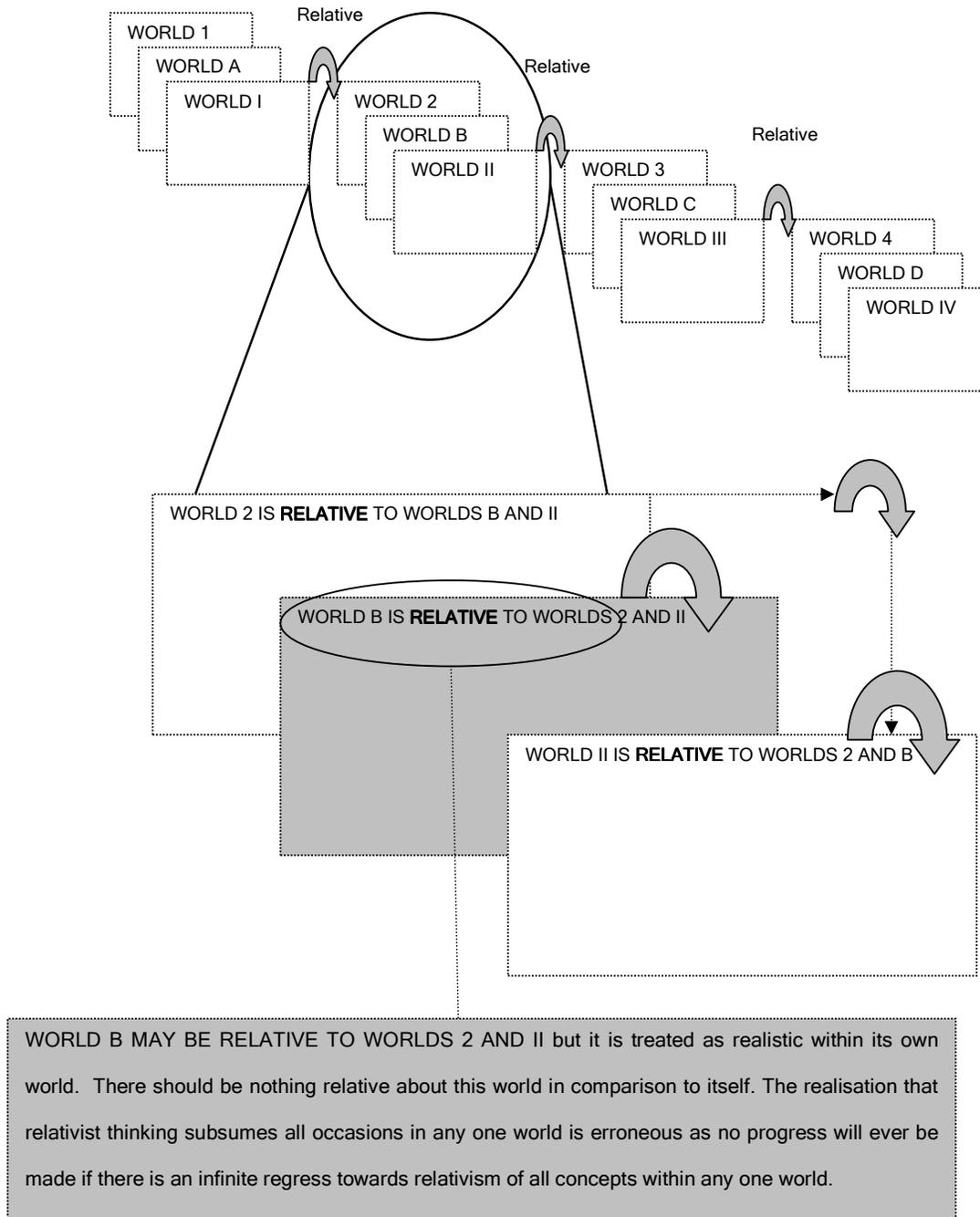
⁶⁵ An issue (like others contained elsewhere in this section) which has particular significance for intelligence assessment and dynamic assessment. Are practitioners needed to classify or assist? What has become of our aspirations to aid in the development and growth of individuals if classification is always the end result? More pertinent to the intelligence assessment issue is what happens when science and social policy collide; more often than not science has to bow to political agendas when all scientists know that this should not be the case (Brown, 1994).

⁶⁶ In the strictest and purest form of mathematics however, nothing can be substituted for proofs in the establishment of truths not even correspondences with reality (Jaffe & Quinn, 1993). Proof in mathematics is analogous to the experiment in the natural sciences; a manner and method of knowing the truth (Butterworth, 2000).

⁶⁷ The author uses the word “ultimately” as it is only with the passage of time that these and related issues might be further resolved.

to one. It is imperative that a position is taken on such an issue (Wagner, 2003) whilst acknowledging the inherent disadvantages within such a chosen approach. Critical psychology's initial reaction to positivist science as originally engaged in the early twentieth century in terms of maintaining the status quo (keeping the marginalized at bay by further supporting the need to do so through scientific research) is acknowledged as such (Snyman, 1993) and its relativist leanings are accounted for by reactions to the ways in which science was practiced. Critical psychology per se reflects no one specific area of investigation, it being an approach towards studies which seek either knowingly or unknowingly to continue to adhere towards discriminatory practises (Nightingale & Neilands, 1997; Prilleltensky & Fox, 1997; Richardson & Fowers, 1997) and in a way, this study's guiding light really is dynamic assessment's philosophy and stance regarding human intelligence and growth with which the author shares a close affinity. Cernovsky (1997) makes explicit his regard towards intelligence assessment relating to this issue. Perhaps a more apt (yet less circumlocutory) title would have included "critical" alongside "exploration". And it is upon *this premise* and general notion that support for this author's opinions is given. A radical relativist might well throw in the towel round about here. An illustration might serve to illuminate this stance on the relativism/realism debate in figure 4 below.

Figure 4 Relativism and realism between and within worlds



2.4.5.3.7 Perennial nature/nurture

Thankfully the debate (or showdown) between nature-nurture⁶⁸ supporters no longer represents the stark and opposing views on the matter as once was the norm (Fischer, 1980; Plomin, 1989, 1997; Rose, 2000). The reconciliation now assumes a more or less substantial portion of both contributions to the end-product of human behaviour including cognition (Klvington, 1986; Plomin & Petrill, 1997; Sternberg & Grigorenko, 1997) resulting in the more-or-less uncontested notion of epigenetic (genetic/environmental) rules through which much of human behaviour is guided (Greenough, Black & Wallace, 1994; Wilson, 1999). Nature or nurture has become nature and nurture; nurture through nature and even nature through nurture (Lloyd, 1995). Yet the idea of transfixed intelligence levels presumed unalterable (yet malleable to a degree via home environment and mediated by latent processes eliciting both continuous and discontinuous development) and predictive of later ability is still pervasive in lieu of many studies'evidence (Cardon & Fulker, 1991) which attest to the limit of environmental factors resulting in the malleability of IQ (Locurto, 1991). Yet, both environmental and native contributions are inherently constraining in how development progresses as certain biases exist both internally (neuronally) and externally (environmentally) (Johnson, Bates, Elman, Karmiloff-Smith & Plunkett, 1997). The oxymoronic situation of nature's hard wiredness and the brain's neuronal⁶⁹ or synaptic plasticity (Squire, 1986; Von der Malsburg, 1987) poses an irony of sorts but when cognisance is taken of both phenotype and genotype influences on the developing organism it can be seen that hard wiredness and plasticity co-occur; an aspect grounded within dynamic assessment. The study of gene-environment interaction pivots the notion of gene sensitivity to environmental effects, in other words how the environment moderates genetic effects (Asbury, Wachs & Plomin, 2005). However this debate takes on a decidedly more subdued veneer when subsumed within intelligence assessment and race issues (Gordon & Lemons, 1997; Valencia & Suzuki, 2001). That debates no longer ensue per se between these two seemingly opposing views is not a true sentiment, as the focus has now shifted onto the quantifiable differences between the two such as the contribution between both between-groups and within-groups differences in intelligence (Suzuki & Valencia, 1997). As Rose (2000) fervently states, this debate "has been the preoccupation of what we may call 'Anglo-Saxon' psychometrics" (p.123).⁷⁰ There is renewed concern for how the two views accommodate the nature of phenotypical change as opposed to how much each view contributes to change (Ceci, Rosenblum, De Bruyn & Lee, 1997).⁷¹ The enveloping framework which grounds this debate is contested though as misleading in terms of presenting the two as opposing views on a continuum. The new framework as informed from insights obtained from molecular genetics and developmental biology illustrate the epigenetic view of these two complementary viewpoints (Bidell & Fischer, 1997). As Piaget originally wrote in 1971 and in agreement with biologist C.H. Waddington, predetermined genetic unfolding is simply not the manner in which development proceeds and he calls for envisioning development as emanating from a set of axioms which when fully developed cannot be derived from the axioms originally underlying it (Piaget, 1994).

At this point in time it can be stated that the author's leanings are towards a more-or-less equal contribution made by both nature and nurture but the following statement is firmly supported: at the very start there is biology⁷² ("psychology is inevitably based in biology"; McManus & Bryden, 1994 just as "mind had to be first about the body, or it could not have been; Damasio, 2006, p.xxvi) and it is common knowledge that single-gene disorders are identifiable as particular areas on chromosomes even though it is conceded that linking intelligence to genes is not as straightforward an issue (Plomin & Petrill, 1997). However, findings within mental-retardation research and Alzheimer's have evidenced single-gene linkages. Other single-gene linkages include phenylketonuria which results from the carry over of a single recessive gene from both parents (Plomin, 1997) yet the effects of this mutation can be nullified by the correct dosage of phenylalanine (hence the environmental effects on phenotype expression) (Sternberg, Grigorenko & Kidd, 2005). Genetic markers are of late starting to emerge in the molecular genetics field related specifically to intelligence and there is good reason to believe, at least for now, that this area of nascent research can only but add value to the field of intelligence and dynamic assessments of ability and potential (Petrill, 2005). It is conceded that pleiotropic (and even probabilistic) interactions make up the varied synchrony of manifest behaviour and intelligence functioning. The accepted form of genotypical influence on phenotype arises as a result of group genetic influences (Wahlsten & Gottlieb, 1997)

⁶⁸ Or phylogeny and ontogeny - this plays forth in a later section dealing with the heritability of intelligence (*h*).

⁶⁹ Considered the brain's computational unit which can grow or deplete in neurogenesis-inducing tasks (Bolhuis, 1997; Innocenti; 1997).

⁷⁰ Of course the fact that this is even considered an issue here is perhaps more telling of the author than the subject at hand!

⁷¹ Although Hunt (1997b) does make a point when he states that this debate represents one of those acrimonious areas in psychological research where "if you missed it the first time, you can always watch the reruns" (p.531).

⁷² Critics might well add that biology itself is a product of continual environmental input and that our very DNA is a result of parasitism (Zimmer, 2003) and that viewing the genome as a code or programme is limiting in terms of its not accounting for developmental branchings that govern gene activity (Oyama, 1994). How far back this argument can go is without a doubt thought-provoking. But the "environmental" influences referred to within this study are those that are least variable in terms of genotype alteration within a single human being within a single life-time. Of course there is evidence that phenotypical changes can in fact alter genotype within an individual in a life-time (Meng, Smith, Hager, Held, Liu, Olson, Penington, DeFries, Gelernter, O'Reilly-Pol, Somlo, Skudlarski, Shaywitz, Shaywitz, Marchione, Wang, Paramasivam, LoTurco, Page & Gruen, 2005; Phillips, 2005; Scott & Louis, 2005; Weaver, Cervoni, Champagne, D'Alessio, Sharma, Seckl, Dymov, Szyf, & Meaney, 2004). It is cautioned that intelligence itself is a phenotype, with at least half the variance of IQ being explained by non-genetic factors (Plomin & Petrill, 1997). Namely environmental impingements working via both experience-expectant (akin to the critical periods of development, say for language acquisition) and experience-dependent processes (Black & Greenough, 1997).

and the relation to high and low psychometric IQ has already been linked to variant allelic associations or frequency for two genetic markers (Ceci, Rosenblum, De Bruyn & Lee, 1997). It is becoming increasingly obvious for the need to view intelligence measurement from biological-environmental models as opposed to solely envisioning intelligence from psychometric perspectives only (recall that psychometric results are biological derivatives obtained from different manners of assessing for essentially the same purported construct; Sternberg, 1991). Depending on how studies are conducted and the measures utilised in the process of deducing what in fact contributes more towards intelligent behaviour, will often affect the final result reflecting a situation in which studies convinced of either the role of nature or nurture as superseding the other will often come to an expected conclusion (Scarr, 1997). Central to this nature/nurture debate concerning intelligence is the degree of brain plasticity and how environmental impingements on the brain often result in neuronal changes (both arborisation and pruning or the constraining of neural plasticity) (Johnson, 1994b; Kolb, 1994; Kozorovitskiy, Gross, Kopil, Battaglia, McBreen, Stranahan, & Gould, 2005; Nowakowski, 1994; Scheibel, 1997). The extent to which these changes remain permanent and are transferable (typically the goal within any dynamic assessment intervention) is contentious (Das, Naglieri & Kirby, 1994) especially as it pertains to development and learning⁷³ as well as task familiarity and task similarity (DeFrance, Hymel, Degioanni, Calkin, Estes, Schweitzer & Hymel, 1994; Greenough, Black & Wallace, 1994; Huttenlocher, 1994; Klauer & Pbye, 1994).⁷⁴ Transfer of skill and modified cognition is dependent on the nature of intervention which can range from standardised prompting to intensive mediation (Delclos, Vye, Burns, Bransford & Hasselbring, 1992) but this is of course where the choice of intervention becomes problematic in terms of feasibility. Intervention work within intelligence often assumes plasticity of intelligence in contradistinction to mainstream thoughts (Gulgoz & Kagitcibasi, 2004).

The current literature on domain-specificity and domain-generalty within assessment has brought to the fore the question of what exactly is being tested for and changed as a result of intervention. It has been argued that specificities of cognitive skills can be construed as less *g*-loaded than generalities (overall skill) which emerge when the totality of specificities is factored into a primary *g*-loaded factor for instance (te Nijenhuis, Van Vianen & Van der Flier, 2005). Are neuronal changes within the plastic brain brought about when cognitive interventions seek to induce behavioural change? These changes are often verified on a behavioural level (cognitive mental construct level) but not necessarily within a deeper neuronal level (cognitive neuroscience level) within dynamic assessment research. Nevertheless, should change occur at many levels of description, the remaining issue is whether what was changed will remain stable and enduring and how this change impinges back on to *g*-loaded constructs. This entire argument of course hinges on the assumption that *g* as an entity (statistical artefact or actual “thing” in the brain⁷⁵) exists (Lohman, 1997b). Psychometrics pleads for statistical artefacts to aid in explaining away findings on assessment instruments. Dynamic assessment pleads for behavioural change to explain away findings on assessment instruments. Are these varied approaches talking at cross-purposes? Nature vs. nurture, static vs. dynamic assessment, reification of statistical artefacts vs. identifiable mapping of such a construct within the brain are all issues which factor into the area of dynamic assessment within intelligence. These issues are not near to being resolved, but perhaps if these issues are considered as central some progress might well be achieved.

2.4.5.3.8 Static vs. dynamic assessment of intelligence and potential

Although this issue is mentioned as a separate heading it is placed here for the express purpose of allowing for the immediate seepage of thoughts and alignments which have been expressed above to filter through and thus permeate the issues highlighted below. The stance on dynamic assessment and static assessment within intelligence research, which this author takes, can now be better understood in lieu of the varying views thus far expressed (mind-brain, consciousness etc). This study seeks to explore a meta-theoretical framework for dynamic assessment and intelligence and in order to accommodate the many types of assessments to be had, practitioners would do well to take cognisance of the advantages that each has to offer the field of intelligence research. Dynamic assessment has never sought to overthrow nor do away with conventional testing (for which at times it is a firm supporter). Dynamic assessment follows a progressive and accommodating mandate which aligns itself with more static type intelligence assessments often serving in a complementary fashion. It redirects the attention of practitioner and researcher to the potentially rich source of variation to be found in many types of intelligence assessment protocols. The fact that this study exists is testament to the firm belief the author has in the potential that dynamic assessment has to offer but is simultaneously fully aware of the need to maintain and develop more static types of assessments. Both are equally necessary. Dynamic assessment is entangled in messy ontological, epistemological and foundational philosophical issues:

⁷³ The actual relation between developmental changes and their concomitant neural changes in terms of volume of axons and dendrites has not been adequately addressed (Huttenlocher, 1994) but that there is a decidedly direct link is clear (Horn, 1994).

⁷⁴ Neurogenesis has, in the past, been supposed to have resulted from enriched and stimulating environments. But even this “fact” is being reconsidered in light of new evidence against this hypothesis (Meshi, Drew, Saxe, Ansorge, David, Santarelli, Malapani, Moore & Hen, 2006).

⁷⁵ Spearman’s *g* factor as statistical artefact as opposed to the later idea of the speed of neuronal firing being equivalent to *g* as expressed by Eysenck and his colleagues (Eysenck, 1987).



- firstly it is located within psychology, which itself as a discipline hovers continually between a natural science styled endeavour to understand human beings as opposed to the more relativist and grounded understanding emanating from ethnographic discourses;
- secondly, it has developed alongside a seemingly opposed manner of assessment, namely, psychometric assessment which often concerns itself with the assessment of intelligence. Yet it can even be stated that dynamic assessment has done a service for mainstream assessment in broadening the view encompassed by psychometrics instead of ousting it and has not yet evidenced a dislodgment from psychometrics to date (Kozulin, 2005)
- thirdly, the field of intelligence research is so overwhelmingly well padded with empirical research data that the very construct of intelligence is still not definable
- fourthly, its history is one of amalgamation of European and American thought, often represented by very different views on human nature (rationalist philosophical bents versus more pragmatic empirical bents; Smith & Katz, 1996)
- fifthly, in trying to balance views and attend to various opposing philosophical stances on human behaviour and assessment it would seem that dynamic assessment is stretching itself rather too thinly, for in trying to please as many views as possible it has become hybridised in order to merely survive as a method of assessment. At once firmly advocating its primary tasks as method of assisting change in human thought through the process of assessment and advocating its staunch strictly science-like approach towards this same endeavour. Can dynamic assessment have its cake and eat it too? Perhaps, but in order for this to be realisable, its ensconced foundations need to be studied

At times, problems seem insoluble and intractable so much so that any path evidencing light at the end of a very un-illuminated tunnel seems almost too good to be true and any viable solution on the horizon is immediately investigated for possible avenues of pursuit. That this study is one such viable option can only be later assessed. Dynamic assessment offers a powerful humanising approach towards assessment;

- It is a method which seeks to humanise the research subject, a subject with cognitive as well as affective characteristics;
- It is an ecologically valid approach towards assessment which aims to be an unbiased, culturally relevant, clinically orientated, time-consuming enterprise in helping individuals achieve change in all its enlightened connotations,
- It was an early flourish and confluence of French, Russian and Jewish psychology later stirred up with American and German standardised empirical research and practice but always having as a goal, change, within assessment, acting as catalyst of growth within a context of examiner-examinee interaction

The estrangement of such obviously opposing views on how science should be practiced cannot be more firmly illustrated than by viewing this history of dynamic assessment. Dynamic assessment appears almost ethnographic in mode utilising a critical psychology approach towards the treatment of individuals, modes that are in complete dissentience to those more attuned to objective explications of individuals. Theirs makes for an uneasy alliance. Yet, such opposing and dichotomous views permeate and at times define what it means to engage in dynamic assessment within psychology. No wonder this field is at a crossroads, for it can be stated that psychology finds itself stranded at a similar junction unable to move. Insoluble and intractable? Or are we in fact asking the wrong questions?

2.5 Summary

The reasons why the above-mentioned aspects of human investigation were chosen for brief discussion (over-and-above those already stated) is the intertwined nature of each aspect's basic epistemology and how this filters out in how research is conducted (Wagner, 2003). Each aspect and its requisite beliefs held by the author influence and impinge on the other. To advocate that mind is brain or that consciousness is bundled nerve fibres and to concurrently maintain that intelligence is not similarly reducible makes for an uneasy and unworkable framework within which to work. Is it possible that, at once, a belief in such contradistinct issues can be held? That times change and along with it technology and understanding of how things work is a given, and although the author has stated her basic epistemological and ontological leanings, it is not to say that such leanings will forever be held static. For now, as the case rests and for what is to follow (however much or however little these leanings weave their way into the study) these affiliations are considered stable and enduring. As stated above in 2.4 the sorting out of philosophical issues at the outset is just as necessary a task for the author as it is for the field of dynamic assessment. The author is merely reflecting a trend necessitated by the field.

It may seem from what has transpired above, that the author is a firm believer in what is considered in some circles to be quite unseemly and old-fashioned. The irony of the situation is just this: although in support of reductionism, leanings towards hereditarian views and beliefs in a more general intelligence and all that goes with these beliefs as well as realist interpretations of reality as opposed to the relativist counterparts, there lies a firm conviction and intuition that change occurs throughout life, that emergent properties of systems loom below the surface and if tweaked will blossom and that no-one is a victim of their biographies. It is for this very reason that the ensconement of dynamic assessment within intelligence assessment in psychology is at its very axiomatised core a belief system which (an *a priori* system, struggling towards *a posteriori* distinctiveness) if given the right opportunities and correct advancements, will surely date conventional approaches towards



human intelligence assessment and replace the current edifice or core affiliations with a more informed, progressive, utilitarian and enlightened foundation.

No pangs of guilt are evident when acknowledging that in order to better understand the workings of humans in all their varied contexts, reducing each area of study is the better method to follow. There are countless arguments and debates concerning the validity of the scientific method itself, but the belief in this method's superiority in comparison with other advocated methods is up-held.⁷⁶ is it *the* best method per se? Probably not if one was to study all possible methods through all possible times and all possible places in time but the point is that we are not allowed such luxuries and hence need to conform to one tried and tested method which does have an enviable record in getting things right (despite all the misgivings of those who deny this and also despite the method's unfortunate histories within varying contexts - no method is perfect after all). In similar vein, relativism is an admirable undertaking when considering all possibilities of investigation and a critical approach to how things are studied and for what reasons aids in keeping the enterprise on its toes but that all approaches need to be accommodated, through all time and all locations threatens to unravel a method which has thus far proven helpful.⁷⁷ It is felt that it is fairer and safer to lay the cards on the table even though to have made choices between supposed dichotomies might in some circles be considered ill-conceived.

Choices do need to be made. One cannot forever dance in the middle accepting all views all the time. Some views are less extreme than others, granted, but it seems unlikely that the very human enterprise of seeking new knowledge will get anywhere without choices being made on fundamental issues.⁷⁸ Attention is now turned to dynamic assessment proper with a brief introduction to this manner of assessment which encompasses a glance at its historical growth as well as a brief look at who is considered by some to be the father of dynamic assessment; Lev Vygotsky and his role within the Soviet psychological enterprise. This section will conclude with a look at the current awareness of this method in the United States and the United Kingdom.

2.6 Dynamic assessment within intelligence

The aim of this theoretical study is to devise for dynamic assessment a framework which will provide a means for comparison of various models and theories within dynamic assessment as housed within the broader field of intelligence assessment. The framework developed for this purpose is the culmination of concerns spanning basic philosophy of social science research, the methodology of science and social science as product of the times, meta-concerns in terms of the general direction of advancing science practice and specifically issues pertaining to psychometrics; mathematical, statistical and measurement foundations of psychological assessment. Advances within psychology in general but more specifically within assessment have more often than not been steered in various directions according to agendas, whether implicit or explicit and have resulted in the position in which current intelligence research and dynamic assessment find themselves. Likewise, it is to be expected that the framework for this study has been similarly affected by historical contingency and it too will most likely be accepted as a product emanating from a time in which the concerns looked at were to be expected. The application and exploration of a meta-theoretical framework will be discussed in chapter 5. For the purposes of clarity, the general introduction to dynamic assessment and intelligence will be placed here due to the many comments which will be made in the following chapters which in turn will necessitate familiarity with both dynamic assessment and intelligence.

2.7 Dynamic assessment: fundamentals

Dynamic assessment is a manner of assessing individuals' at times hidden potential (reserve capacity) in a fluid, process-orientated, diagnostic, engaged and flexible manner in which aiding or guidance (via instruction and feedback) of cognitive skill acquisition is of prime importance (Campbell & Carlson, 1995; Elliott, 2003; Gillam & McFadden, 1994; Grigorenko & Sternberg, 1998; Kirkwood, Weiler, Bernstein, Forbes & Waber, 2001; Kirschenbaum, 1998; Kliegl, Smith & Baltes, 1989; Lidz, 1991, Lidz, 1997; Meyers, 1987; Minick, 1987; Sternberg & Grigorenko, 2002). It stands in stark contrast to the more product-bound approaches of mainstream psychometric and edumetric assessment (Craig, 1991; Gupta & Coxhead, 1988b; Resing, 1993; Slenders & Resing, 1997) by emphasising the change in performance (rate) and remedial strategies necessary to progress (Bejar, 1984; Brown & French, 1979; Campione, 1989; Wiedl, 2003). Rate of learning, amount of improvement (typical of the

⁷⁶ Although the author is in agreement with Faust and Meehl who state that "Knowing that science is the best game going and has produced remarkable achievements establishes little about its approximation to the optimal" (1992, p.201) one is left with the question of what is optimal? There is no benchmark really for this concept.

⁷⁷ In more instances than not it has proved helpful - although this will undoubtedly be debated.

⁷⁸ Who is to say that the enterprise of science itself is not just one more method of viewing life? Yes, this is accepted. However, to carry on in this manner will inevitably lead nowhere. One has to know when to say "stop" and proceed from a foundation as solid as can be made. Intelligence is a relative concept, to measure it is relative, to even consider it as a concept is relative. One feels uneasy with this type of argument as it becomes evident that to even try to explain or measure or assess it seems to be a futile endeavour. One might as well give up. That intelligence has been understood in terms not amenable to all views is acknowledged. This, however, does not necessitate giving up on trying other ways of explaining it and related concepts.



Feuersteinian and neo-Vygotskian views) as well as amount of aid necessitated (more modern views of gauging potential) are all methods of assessing for growth of learning or potential (Ferrara, Brown & Campione, 1986). The relationship between tester and testee as characteristic of strict neutrality is the hallmark of conventional testing which, if violated, would render the objectivity null and void (Greenfield, 1997) but not so with dynamic assessment (Lidz, 1992b). It represents greater all round diversity in assessment and the method's results extrapolate to a far wider field of application than mainstream assessment (Gupta & Coxhead, 1988a) leading to, at times, fairer and greater predictive diagnostic validity (Ferrara, Brown & Campione, 1986; Gredler, 1988; Resing, 1997) for below-average performers (both majority and minority groupings) on conventional IQ tests (Babad & Budoff, 1974; Budoff & Hamilton, 1976; Hessels, 1996). Movements are afoot within static-based modes of testing which seek to make such tests more functional, at least for special education populations in terms of prescribing treatment in respect of test results yielding another type of validity, that of treatment validity (Flanagan, Andrews & Genshaft, 1997).

The nature of instruction and feedback is not one of repetitious aid but of engaged understanding of the unique attributes of the individual being assessed (although this can vary dramatically depending on reigning circumstances) (Goikoetxea & Gondra, 1995) where the learner is a participant and not a subject in the active learning process (Svinicki, 1998). The relation can even be said to be one of personal understanding (Feuerstein, 1972). Synonymous terms with very similar philosophical backgrounds resonate with this method of assessing learning potential and include among others, learning tests (or the German *lerntest*), interactive assessment, trainability testing, testing-the-limits (via graduated prompting) and mediated learning (Brown, Campione, Webber & McGilly, 1993; Fernandez-Ballesteros & Calero, 2000; Swanson, 1995; Von Hirschfeld & Downs, 1992). It can be viewed as a strategy within the cognitive education approach where educability is understood to be synonymous with the modification of intelligence (Chartier, 1996). The basic rationale behind the method of assessment is that if a student can improve upon initial performance when aided (hence interactive; Lidz, 1991), resident potential exists with which much can be achieved⁷⁹ (Ukrainetz, Harpell, Walsh & Coyle, 2000). Testing-the-limits allows for the pre-establishment of an age-appropriate level of achievement according to which children are assessed via a step by step process of approximating the maximum levels of performance (Kliegl, Smith & Baltes, 1989). Depending on the level achieved for any one group, individual differences within the group can be assessed thus allowing for more accurate assessments of development level (De Ribaupierre, 1993).

Inherent contradictions pervade the mainstream assessment field especially within educational contexts (where dynamic assessment sits at ease; Daniel, 1997). In such circumstances, current functioning is utilised as indicator of future success (Kozulin & Garb, 2002) where in fact the learning process itself should be utilised; it is after all the goal of education to monitor the *learning process*. By observing change and by directly interfering with development one is more able to understand this process (Paour & Soavi, 1992). Learning and instruction as well as assessment and teaching are examples of dyadic approaches utilised in models encompassing Feuerstein's original notion of instrumental enrichment where there is a dual process carried out by both teacher and student as well as the processes of information gathering. One need only think, for instance, of Ashman's Process-Based Instruction (PBI) model (Ashman, 1992). Feuerstein's product is an applied instance of dynamic assessment more so than a separate theory of dynamic assessment (Feuerstein, 1994) where intellectual potential is "activated" via instrumental enrichment (an instance or interrelation of his theory of cognitive modifiability) (Lidz, 1992b; Messerer, Hunt, Meyers & Lemer, 1984). The acceleration of maturation through learning is an old Vygotskian notion (Das, Parilla & Papadopoulos, 2000) which seems to be rediscovered every so often in one guise or another. The emphasis within core dynamic assessment interventions is placed simultaneously on learning potential as evidenced through cognitive skill which is in turn founded on psychological-cognitive theory (Guthke & Beckmann, 2000a). Here the link between dynamic assessment, cognition and intelligence comes to the fore. Dynamic assessment is so widely applicable in so many contexts that relegating it within the intelligence realm only is short-sighted but this is the focus of this particular thesis. Common characteristics include (Carlson, 1994):

- The notion of inherent modifiability
- Competence and performance which are two divergent concepts
- Test performance as enhanced via interactive intervention
- Process which is paramount to product and
- Development of abilities which are better or at least complementary to developed abilities

Some researchers who do rely heavily on static based conceptions of intelligence yet render this dynamic, prefer a reference to intelligence as "cognitive activities" in order to reflect the process-based nature of intelligence (Carlson & Wiedl, 2000). The teacher or tester is a reflective person who not only monitors the individual being tested but engages in a process of self reflection, judgment and control (Reichenberg & Rand, 2000). Four general aspects which can be considered as characteristic of this approach include its tester-testee relationship which is newly defined as one of collaboration and sharing (S. Feuerstein,

⁷⁹ Of course there exist populations for which even extensive interventions have little effect.



2000); the process vs. product manner of assessment; the nature of the test and the various tests utilised and the manner in which the results are interpreted (Tzurriel & Klein, 1987). Teacher perceptions have been shown to change after dynamic assessment interventions are conducted, resulting in changed expectations of performer and performance (usually for the better) (Benjamin & Lymofsky, 2002; Bransford, Delclos, Vye, Burns & Hasselbring, 1987; Delclos, Burns & Kulewicz, 1987). The link between assessment and intervention as well as the link between assessment and environment and the nature of the process and product of behaviour are noteworthy characteristics of this approach (Meyers, 1987). The learning process is at the core of this manner of assessment and can be studied and observed in the actual learning situation through the analysis of learning curves as well as via the presentation of a learning or training phase during the pre-testing phase of a pre-posttest set-up (Hamers & Sijtsma, 1995). Both pre-testing as well as the monitoring of the learning process itself are necessitated if adequate intervention strategies are to be planned as advocated over seventy years ago by Vygotsky (Day, Engelhardt, Maxwell & Bolig, 1997). This is consistent with his notion of ZPD - how else is one to know of future growth if present rates are not known? Dynamic assessment overflows the boundaries between intelligence (typically the ability to solve unfamiliar problems in the environment) and achievement (the ability to extract and assimilate information within the cultural environment including the schooling system) (Cahan & Noyman 2001). Losardo and Notari-Syversen (2001) offer a generic "theoretical framework" the use of which some may question, and state that dynamic assessment encompasses six theoretical assumptions:

- A Piagetian constructivist perspective
- Vygotsky's social-interactionist view
- Vygotsky's ZPD notion (for long a paradigm more so than methodology; Kozulin, 2005)
- Scaffolding
- Feuersteinian mediated learning experience
- Self-regulation which encompasses some form of non-linear growth trajectory

Such a framework of course is characteristic of general practice from several theoretical positions but perhaps is yet to be seen in practical use as one model! Chapter 5 illustrates various models' use of the above-mentioned assumptions as theoretical underpinnings.

Dynamic assessment straddles curriculum-based assessment along with general adaptability within life making it expressly useful over a larger domain; i.e. low achieving individuals may well score higher on intelligence estimates (Elliott, 2003). Cognitive training and dynamic assessment have been equated as methods of inducing change thus further elaborating the methodological spectrum of possible alternatives for learning potential assessment (Scharnhorst & Büchel, 1995; Schneider & Ganschow, 2000). It places the individual, usually but not exclusively young children (Kahn, 2000; Klein, 1992a; Lidz, 2000b; Mearig, 1987; Samuels, Lamb & Oberholtzer, 1992; Tzurriel, 2000c, 2000d; Tzurriel & Haywood, 1992), at the centre of the assessment process and regards change within the child as criterion (Burns, Delclos, Vye & Sloan, 1996). Although dynamic assessment is also amenable to group assessment (Luther & Wyatt, 1996; Rand & Kaniel, 1987; Tzurriel & Feuerstein, 1992; Ukrainetz, Harpell, Walsh & Coyle, 2000) practical problems inhibit its wider scale usage especially in South Africa (Miller & Bradbury, 1999). The individual becomes a reciprocal respondent responsible for the self-monitoring of strategies whilst the assessor engages the situation as facilitator (Schneider & Ganschow, 2000). The notion of a critical age for development of certain skills akin to Piaget's stage-like model of human development informs the more clinical approach towards dynamic assessment as younger children are perceived to be more malleable to interventions of various sorts. Moreover, children under the age of five evidence inconsistent results on traditional measures of learning ability (Vye, Burns, Delclos & Bransford, 1987) due largely to the fact that evidence attests to many critical periods of development where neuronal overproduction is followed by selective tailoring and so on through progressive development (Dehaene-Lambertz & Dehaene, 1997).⁸⁰ Depending on the literature, Piaget's notion of cognitive structures being independent of environmental influences at this early age is either refuted or upheld (Campbell & Ramey, 1990). Socio-ecological variables within the child's environment during critical stages of intellectual development impinge on development trajectories (neurobiological as well as social) and results in rapid development of these intellectual skills (or not, depending on the context) (Blair, 2006; Haywood & Switzky, 1992). Piaget acknowledged the role of parents in providing opportunities for children to develop but Feuerstein and others have placed even greater emphasis on the roles of parents as active and influential modifiers of children's development (Tzurriel, 2000b). Recall that broad-based Piagetian theories are just that; broad and based on average developmental paths where individual development trajectories vary among children (Fischer, 1993). By assessing the individual and individual differences within structural models and doing so within context, dynamic assessment is able to escape averages, generalities and monolithic developmental ladders (Lautrey, 1993). However, a recurring theme in this thesis is the concern over statistical and clinical decision-making and how this influences individual lives for better or worse. External influences are more influential than

⁸⁰ Yet another reason why biological/neuronal studies need to be factored into cognitive/psychological studies especially in the terrain of intelligence assessment and learning potential change-based assessment (Estes & Bartsch, 1997; Feldman, 1997; Haith, 1997) .

previously thought by Piaget and development occurs within domain specific tasks as well as context-specific areas and is not bound by general development across domains (Case, Okamoto, Henderson & McKeough, 1993).

The mechanisms in place during the learning process co-occur in a contextualised situation where development is directed by both the individual (the brain constructs experience as well as regulates behaviour by the manipulation of symbols) as well as social forces such as peers, teachers and parents (who render the symbols meaningful to the child or peer via culturally evolved cognitive tools and do so in a co-constructed process) (Arievitch & Stetsenko, 2000; Jensen, 1992; Portes & Vadeboncoeur, 2003; Schwebel, 1992). This makes the entire programme one of socially embedded learning as opposed to mainstream isolated learning (Brown, Campione, Webber & McGilly, 1993). Collaborative mediation may result in emergent functions which have yet to express themselves internally; i.e. external manifestations of cognitive functions may over time become internalised but only after active collaboration (Kozulin & Garb, 2002)⁸¹ as well as the internalisation of activity via language and thought (Das & Conway, 1992). There is a particularly strong blend of constructivism inherent in understanding the developing child as well as teacher-student interaction (where learning theory and teaching practice needs to be bridged) (Meyer, Cliff & Dunne, 1994; Schur, Skuy, Zietsman & Fridjhon, 2002), which if taken in tow with the author's affiliation for reductionist and scientific manners of explication of development can be said to be at odds. This is deemed not to be the case. Here, this intuitive appeal of dynamic assessment's fundamental aims and allure is readily apparent and forms part of current test models (Guthke & Beckmann, 2000a; Jensen, 2000). Construction of experience is both an individual and environmental phenomenon that can be studied scientifically and is consistent with the views expressed by Vygotsky and Piaget in which the progression towards higher cognitive functioning is set about via mechanisms of construction of information and not merely progressive transferral (Lloyd, 1995; Mynhardt, 1995). This is of course the hypothetical best case scenario which we know is often not the case in reality⁸² especially, it has been argued, in school settings, where collaboration between teachers and students is anything but mediatory and any such efforts purported to manifest such mediatory characteristics are disguised (Beveridge, 2000).

The niche for dynamically assessed computer mediated approaches is manifest in situations where human contact with each individual is not feasible (Gerber, 2000) and offers mediated experiences which do not necessarily detract from the philosophy behind Vygotskian and Feuersteinian mediation techniques. The discipline of psychology has focused more on the individual within society than on society within the individual and as such the influence of the "social" is often vague in its meanings (intrapersonal, interpersonal and intergroup) (Finn, 2000). In fact, depending on one's philosophical affiliation or at the very least, the emphasis placed on various aspects within the broader contextualised approaches of understanding development such as mediated learning experience, socio-cultural approaches and/or skill transfer, these approaches can rest comfortably within a larger nested approach known as cultural theory, of which there are numerous models with varying degrees of emphasis on different aspects (Mastergeorge, 2001). As with many concepts within dynamic assessment, mediated learning experience (the pivotal aspect within Feuersteinian theory) as a definition is often conceptually blurred (Miller & Yager, 2001). Mediation is thus a fluid concept with changing emphases regarding its meaning and utilisation within differing contexts (Miller, 2003a). Mediation within the Feuersteinian model depicts the notion of cultural transmission of knowledge and the development of individual cognition within the broader culture (Deutsch, 2003), going beyond Piaget's decontextualised⁸³ theory of growth and development by placing the developing individual within a context. Vygotskian mediation follows the ideal of making available requisite tools (language and thought via signs and symbols) in assisting the developing child with progress into a zone of near development, without which this zone may not be crossed. The emphasis within this type of mediation is placed on the development of higher mental processes firstly via the aid of an adult (or peer) and thereafter the mediation of this effort via the child him or herself. Once the transfer from adult guidance to child-initiated guidance has occurred one can state that mediation has been successfully implemented (Karpov, 2003). The Feuersteinian notion of transcendence (transfer of mediated skill) and the transfer of mediation from adult or child within the Vygotskian system is evident. Feuersteinian mediation emphasises cognition-in-culture and Vygotskian mediation emphasises tools-towards-higher mental processes but both take cognisance of the fact that development does not and cannot take place within a decontextualised environment and both account for the

⁸¹ Recall (if readers can think this far back!) to the days in school where your more able peer beside you (unless you happened to be the lucky able peer) was able to explain something to you in a briefer and far more understandable format than the teacher. It is precisely because the information was channelled from an age related peer in the context of school friend that made the transfer of information easier (Mynhardt, 1995). One was also less scared to ask your peer than the teacher.

⁸² In academic treatises such as this, the hypothetical situation is often described, and so, statements of this sort are a frequent occurrence. The reality, however, is often severely different. Theory needs to anchor itself at some point in some manner and so sentiments often take the form of formulaic statements. There is nothing wrong with this and is necessary, but to stylise the prose by moving away from the reality will only confound theory at a later stage.

⁸³ Very much the main criticism levelled at psychometric and laboratory-based approaches towards the study of intelligence as tasks are assessed without due concern for context and content. Content and context invariably influence cognition (Wellman & Gelman, 1992). This is quite obviously true but perhaps one can argue that to study the degree to which a synapse carries current is rather contextless but that the overall effect of growth and development is very much context-bound. Here one can quite vividly view the disparate research agendas for both static and dynamic assessment (Gardner, Kornhaber & Wake, 1996).

internalisation of information as originally taught or copied from another. The bioecological⁸⁴ model of development and its close ties with epigenetic models of growth allow for a link to be made between dynamic assessment and intelligence, the latter emphasising innate structures more or less in keeping with Piagetian development within an abstract system where the individual's development is almost played out by the unravelling of its pre-determined code (Karpov, 2003). The adult mediator in this instance is not given as much regard for their role in aiding developing to newer levels (Haywood, 2003). Neo-Piagetian considerations of mediation re-look the role of external mediators as "tools" from which the developing child needs to remove themselves in order to progress to new levels of growth, in other words, revisiting self-mediation (which is similar to the sentiments echoed by Vygotskian notions of the mediatory cycle) (Haywood, 2003). Conceptual, procedural and metacognitive knowledge is mediated within the neo-Piagetian set-up via a process of questions and not via a process of hinting as is done in other set-ups favouring other dynamic assessment origins (Haywood, 2003). Children discover the rules for themselves and provide their rendition of a rule as opposed to applying a rule that is taught to them (Haywood, 2003) which brings one to the question of how rule identification is understood during both types of processes, for there is both an opportunity and a necessity for providing a rule to be applied and allowing for a rule to be discovered. Either way, a more experienced person is placed within the assessment situation; a commonality running throughout the broader working of dynamic assessment.

Dynamic assessment's definition is one built from a myriad of other definitions and as is the case with a definition for intelligence, a definition for learning potential or dynamic assessment is similarly vague and diverse or loosely structured (Ghesquière, 2002; Hamers, Hessels & Pennings, 1996; Lidz, 2001; Reschly, 1997). It often refers to a host of approaches (Feuerstein, Rand, Jensen, Kaniel & Tzuriel, 1987) evidencing a healthy diversity of approaches⁸⁵ (Bransford, Delclos, Vye, Burns & Hasselbring, 1987) but it does emphasise trainability of thinking⁸⁶ and reasoning abilities which mainstream assessment does not in general do (Resing & Roth-Van der Werf, 2002). Dynamic assessment, like its intelligence counterpart also lacks a unified theoretical framework (Campbell & Carlson, 1995). Linking diagnoses with treatment is the essence of this approach towards assessment which seeks to modify through active intervention and remediation (Campioni & Brown, 1987; Lidz, 1987a). Modifying the content as well as the structure of thinking, dynamic assessment aims to bridge gaps in cognitive skills and seeks to manoeuvre away from placement of individuals in categories towards inclusive understanding of malleable change where "true ability" as a concept is meaningless (Gamlin, 1996). Moving away from product-based approaches towards the assessment of individuals which in the past has been dealt with by simply ignoring the responsiveness of the testee to the test situation as well as ignoring strategies for interventions, dynamic assessment aligns itself to an attitude of holism (Bransford et al., 1987). It has much to offer the field of intellectual assessment due to its unique stance on measurement and modification of skills within a remedial framework (Budoff, 1987b; Das & Conway, 1992; Feuerstein, Rand, Jensen, Kaniel & Tzuriel, 1987). It seeks to mediate and remediate errors in thinking as opposed to simply ignoring them, as is usually the case (Laughon, 1990; Luther, Wylie & Rosenthal, 1996; Van der Aalsvoort, Resing & Ruijsenaars, 2002). Transfer of task related information is the goal of remediation but the difference between transfer and actual learning is often blurred and indistinct (Butterfield, Slocum & Nelson, 1992) with transfer estimates being notoriously difficult to achieve (Campioni, 1989; Crawford & Das, 1992). Indeed, lack of transfer was the major reason why early intelligence research into the relation between IQ and learning proficiency was not supported and eventually abandoned (Brown, Campione, Webber & McGilly, 1993). Transfer, maintenance⁸⁷ or even adaptability (for that is in effect what is being concluded and what Vygotsky envisaged within his ZPD; the eliciting of change through tasks that "provoke cognitive adaptation"; Craig, 2000, p.7) has been linked to intelligence measures evidencing higher intelligence levels concomitantly with greater ease of transferral (Brown & French, 1979; Day, Engelhardt, Maxwell & Bolig; 1997; Ferrara, Brown & Campione, 1986). Better use of metacognitive functioning eases transfer of cognitive skill and increases the duration of transfer (Hamers, Hessels & Pennings, 1996) between domains which is why metacognition is emphasised later on under the section discussing intelligence. This is one of the many examples or instances where the interchangeable nature of both static and dynamic constructs come to the fore. The notions of each are difficult to define, and the situation is compounded by continuous tautologous referral to definitions within and between the approaches. Hence, dynamic assessment's entanglement. It is wedged in a gulf seemingly unbridgeable (see figure 5 below) as intelligence and learning are related yet manifest as separate constructs (Kanevsky & Geake, 2004). A possible retort to the argument encapsulated in the figure may be that learning potential and IQ are not problematic for some studies.

⁸⁴ One should not confuse the ecological criticism against constructivism which states that what is perceived in the environment is not changed in any way by cognitive factors (constructivist) in perception with the contextually-based understanding underpinning the bioecological approach to human behavior (Grieve & Van Staden, 1988).

⁸⁵ Which is precisely where the problem for some researchers lies - the field's lack of coherent definition.

⁸⁶ And in some instances, trainability in skills and crafts, which are areas of motor ability (Ackerman, 1988). Robertson and Mindel's (1980) study is a refreshing look at the utilisation of trainability within more technical and physical set-ups. The author may well look into this at a later stage as South African is currently (2006) experiencing a shortage of skilled technicians in a number of areas such as plumbing, welding and the like. One could perhaps identify trainability of skill dynamically and here one would most assuredly not be using or testing for "static" intelligence constructs as commonly understood.

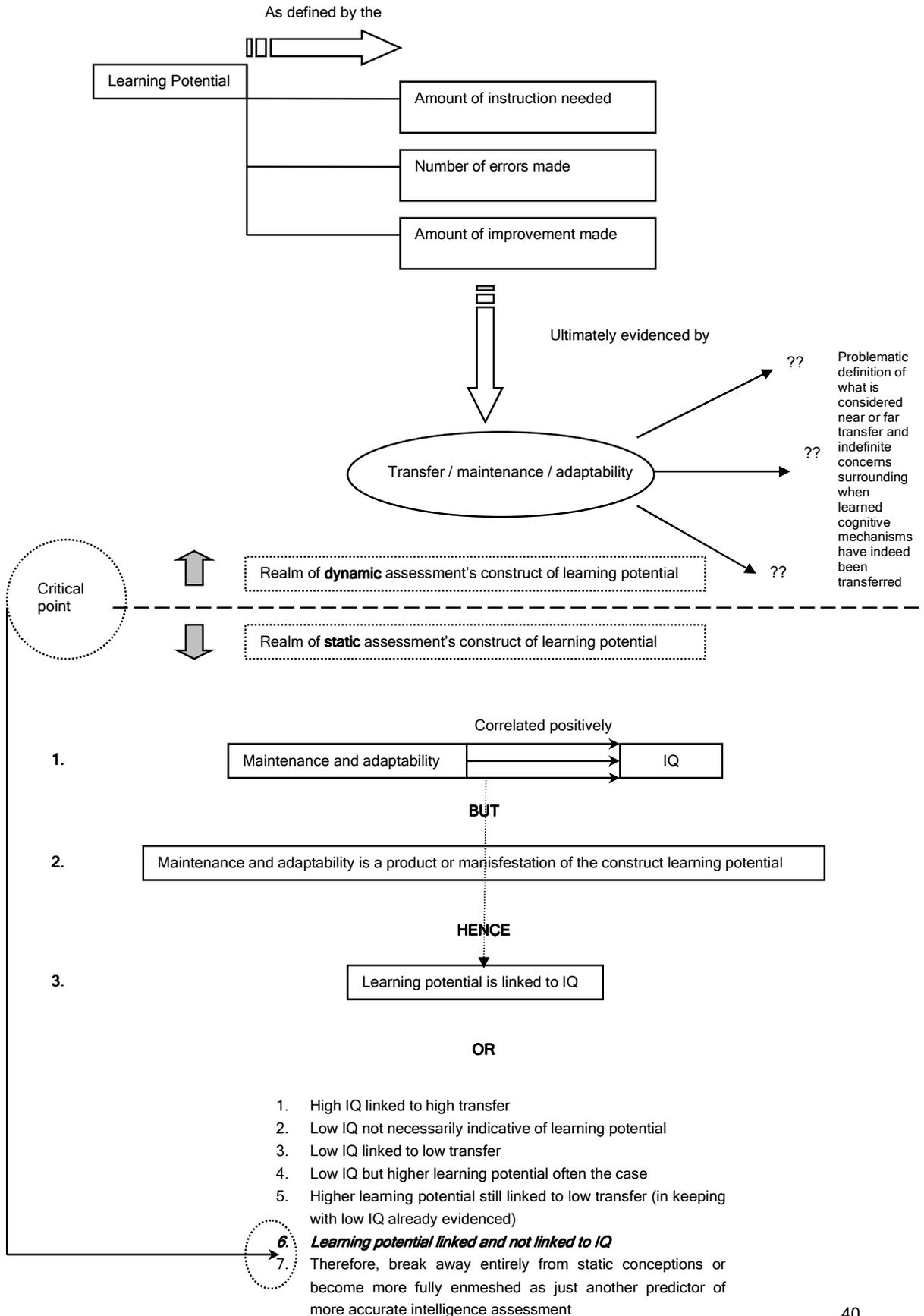
⁸⁷ Transfer and maintenance cannot be equated as conceptual constructs as it has been shown that dynamic assessment has predicted transfer but not maintenance within certain tasks, highlighting once again, the tentative nature of our semantics within this and the intelligence domain (Day & Córdón, 1993).



Just such a scenario as provided below may be of no consequence in any event as most studies assessing for learning potential validation are focused on below average performers where learning potential is a better indicator of future success as opposed to IQ or school marks (Miller, 1998) (not to mention the added group of low learning potential scorers for whom nothing is usually recommended).⁸⁸ But of course dynamic assessment is not only utilised within these populations which is why one is saddled with the issue. The literature is replete with efforts to link learning potential to intelligence whether manifestly or in a veiled fashion. Nevertheless, ultimately, the goal of any intervention is to have as a result the far reaching transfer of skills that can be utilised in disparate contexts with similar underlying principles.

⁸⁸ Perhaps this is unfair to highlight and use in this argument. Low IQ performers with paralleled high LP (learning potential) measures are the ideal group with which to work and advocate the necessary in terms of promoting dynamic assessment. However low IQ and also low LP individuals are considered out of the range of even dynamic assessment's grasp (the author is also not referring to mentally retarded individuals for whom early dynamic assessment initiatives allow them access out of predefined and discriminatory categories). Dynamic assessment could, then, be said to treat borderline cases more effectively than either of the extremes. Initial high-level performers can be disadvantaged by dynamic assessment interventions for whom working strategies are already functioning optimally and should thus be left to attend to their own cognitive devices (Snow & Lohman, 1984).

Figure 5 The inseparability of IQ and learning potential and the iron-clad grip of IQ upon dynamic constructs





Dynamic assessment assumes constant change throughout life and this is reflected in similar assumptions concerning changes within the assessment process (Sternberg & Grigorenko, 2002). Contextual as well as inherent characteristics are deemed equally important allowing for this method of assessment to be utilised in various cultural contexts (Cole, 1996a; Das, 1987; Guthke, 1993a; Haywood, Tzuriel & Vaught, 1992; Hessels & Hamers, 1993; Jensen, 1992; Schardt, Whitten & Gamlin, 1996; Van de Vijver, 1993). Other contexts include the learning disabled, mentally handicapped, socially and economically disadvantaged contexts (Khani & Gamlin, 1996), which is usually related to children (Missiuna, 1996; Tzuriel, 1996). Incidentally, through the intervention strategies of dynamic assessment, misdiagnosed individuals can at times be more correctly diagnosed (or at least diagnosed more accurately) (Popoff-Walker, 1982) which parallels the phenomenon of decreased diagnoses in the number of mentally retarded individuals with a concomitant increase in the number of learning disabled individuals seen in the United States for instance (Budoff & Friedman, 1964; Folman & Budoff, 1971; Shepherd, 2001).⁸⁹ It must be noted that such labels can themselves be considered as socially constructed conveniences for those doing the labeling and this in turn helps to perpetuate a situation already suffused with contradiction and myth (Valencia & Suzuki, 2001).

Mediation is consciously directed at children with the aim of assisting them in making sense of their environments which if not provided can lead to severe backlogs in emotional as well as intellectual development (Klein, 1992b). Here it is immediately apparent that the “social” plays a very important part in dynamic assessment ranging from clinical interventions to the more standardized approaches.⁹⁰ Mediation need not only be social but can also take the form of symbolic agent which according to Vygotskian thinking is revealed to the child through symbols proceeding from cultural transmission (easier to learn) to higher cognitive functioning symbols (more abstract) (Kozulin, 2002b). Clinical mediation (versus the more standardized approaches) is particularly pertinent in societies where disadvantages in education, lack of educational services and parental mediation are rife such as in South Africa (Bedell, Van Eeden & Van Staden, 1999; Engelbrecht, 1996; Shochet, 1992; Skuy & Mentis, 1992) which in the past approached poor performance from a child-deficit point of view largely inspired from the medical diagnostic model based on identifying weaknesses (Archer & Green, 1996; Bejar, 1984; Kriegler & Skuy, 1996; Robinson-Zanartu & Aganza, 2000) (“deficit” in mainstream assessment vs “potential” in dynamic assessment). It has been stated that the disreputable notion of deficit model thinking or genetic pathology model has resurged in the recent past with texts reminiscent of eugenicist thoughts (Valencia, 1997a; Foley, 1997) although one has to be exceedingly careful with statements “for” or “against” these types of texts. Not all statistical delineations of various intelligence assessment results are necessarily tinged with biased arguments. Nevertheless, there is a progressive trend towards the understanding of the processes involved in developmental assessment which vie for the attention of the intelligence researcher which emphasizes the reversal of the effects of sub-standard environmental influences as opposed to unthinking acceptance of irreparable “damage” (Pearl, 1997). There has in the past appeared a trend in which prescriptions for deficit model thinking could be said to have been applied (Valencia, 1997b). During various times, a counter-resurgence of alternative assessments operating in stark contradiction to these trends surfaced, among the most visible being dynamic assessment. Feuerstein’s mediated learning has proven exceptionally applicable in cross-cultural societies, the very population for whom it was intended. It has thus evidenced success within South Africa where mediated learning concepts are directly translatable into cognitive, emotional and cross-cultural dimensions (Skuy, 1996) notwithstanding its obvious time and cost related disadvantages (Skuy & Mentis, 1990). Familiarisation with test content has been a known factor in cross-cultural psychological testing for many years indicating its fruitful area of implementation quite early on (Biesheuvel, 1972; Ortner, 1972) although test familiarity as instituted within a practice session does not constitute a dynamic intervention (Lidz, 1992b). Interestingly, Tzuriel (2001) cites the work of South African researcher Shochet (1992) who looked at dynamic assessment’s applicability in tertiary education in South Africa but is cited as the only exponent of dynamic assessment within this context. Despite dynamic assessment’s applicability across a broad arena of implementation, it has found a particularly rich niche in South African tertiary education, a trend, it seems, which is relatively unique to this country.

The typical mode of a dynamic assessment intervention closely follows a test-retest design which is punctuated by varying levels of mediation (Budoff, 1987a; Campione, 1996; Elkonin, Foxcroft, Roodt & Astbury, 2001; Hamers & Resing, 1993; Klein, 1992a; Lidz, 1987a; Lidz & Pena, 1996; Taylor, 1987; Tissink, Hamers & Van Luit, 1993). Mediation via a process of observation as well as participation results in lower common variance accounted for as opposed to the utilisation of two tests for instance (Haywood & Tzuriel, 2002). Along with decreased explainable variance newer change-based IRT can more successfully accommodate

⁸⁹ What was considered a “moron” by Goddard is no longer considered as such (1912) nor is a “backward” child backward (Burt, 1937; Fletcher, 1991). Is it possible that this type of labeling is still prevalent? Another niche which dynamic assessment can quite readily fill is the re-evaluation of cognitive functioning and the resultant move away from the medical model of categorization although as it is currently understood, learning disabilities are in fact due to “basic disorders in specific psychological processes that are a reflection of neurological, constitutional, and/or biological factors” and not due to emotional or physical disorders nor are they related to general intelligence (Swanson, 2005, p.409). This is particularly pertinent here as learning disability research has often evidenced deficits in working memory, which as a construct is often related directly to intelligence.

⁹⁰ At times the vague definitions of what in fact dynamic assessment is can be seen in encompassing measures of potential where by merely readjusting statistical givens on a traditional IQ test (Taylor & Richards, 1990) sufficient change has said to have been made. That this would suffice as “dynamic” is questionable.

change as a construct (see chapter 4 for more on this). Depending on the nature and underlying philosophy⁹¹ of measurement, the intervention can take place via clinical and intensive mediation through to a structured progressive hinting and somewhat more standardised approach (Burns, Delclos, Vye & Sloan, 1996). This format of assessment allows for better prediction of school results, yields more information than mainstream testing on both the strong and weak cognitive points during performance and the information gathered is relevant to the instruction being given (Meijer, Oostdam & Van der Sluis, 2002; Tissink, Hamers & Van Luit, 1993). Depending on the amount of control exercised within this structure, it can be loosely classified as a classical quasi-experimental design (Klauer, 1993). It is, however, considered as undesirable within standardised assessment (Klauer, 2002) but this is counter-intuitive for at least one reason: that of construct validity (Hamers & Sijtsma, 1995; Lidz, 2003) for how much closer can one come to assessing learning than in the actual process of learning (Lidz & Gindis, 2003)? It is envisaged that the future of intelligence assessment will focus on issues of content and construct validity thus enabling better extrapolations of test results to other aspects of learning and intelligence (Kamphaus, Petoskey & Morgan, 1997). Feuerstein's Learning Propensity Assessment Device (LPAD) is in fact both a product and informer of the theory that underlies it, namely, structural cognitive modifiability (SCM) (Feuerstein, Feuerstein & Gross, 1997) and has been in use since the 1950's (R.S. Feuerstein, 2000). This process manner of considering human functioning within the broader environment is reminiscent of the thinking of Piaget and Vygotsky, although the latter two do not cohere entirely on various points of view within educational theory (DeVries, 2000). In this manner, substantive theory is upheld in opposition to the more traditional hypothetical theory (more on this in chapters 3 and 4).

Aspects considered as error within static assessment such as fatigue, stress and lack of attention for instance are what make for diagnostic decision-making within dynamic assessment (Lidz, 1997). The rationale underlying static and dynamic types of assessment are radically different, philosophically and fundamentally but as Guthke and Beckmann (2000a) so aptly point out, static components are included within dynamic assessments but not the other way round thus giving dynamic assessment more of an edge so to speak. Moreover, most individuals do not utilise all their resident potential and this will inherently not be picked up within static assessment in any event (Resing, 2000). Due to the unique nature of the interaction (Missiuna, 1996) between the testee and the tester in dynamic assessment however, validating variable constructs can become problematic especially when the interaction between the testee and assessor also varies between test situations (Burns, 1996) where specific target group interventions are necessitated and thus change from one context to the next (Guthke, Beckmann & Dobat, 1997). Reliability and validity issues from psychometric theory start asserting their presence in this regard. Over and above the preoccupation with prediction, dynamic assessment aims to explain in addition to predicting scores in school tests thus making it a more instruction-based approach (Ruijsenaars, Castelijns & Hamers, 1993). It creates a profile of the learner's behaviour by adapting and attenuating intervention programmes to more readily suit the individual and in this way compiles an individualistic account of the person within the assessment situation (Greenberg & Williams, 2002). The process-based approach allows for explanation of cognitive deficits usually far in excess of anything offered by most mainstream assessments. Three main attempts at modifying traditional psychometrics include a re-interpretation of test score results to reflect a more culturally attuned approach; the modification of test items which evidence contextualised individual functioning (cultural and developmental) and lastly an attempt to modify functioning beyond that which is recognizable as manifest (Feuerstein, Rand, Jensen, Kaniel & Tzurriel, 1987).

Mediation effectively spans purely qualitative interventions in which intensity and duration (minutes to years) is emphasised in a clinical set-up such as offered by Feuerstein and Jensen for instance (Elliott, 2003) where domain-specific tasks are avoided due to their similarity with school related tasks (Greenberg, 2000; Lidz, 1997). It also spans standardised quantitative testing within a decidedly more psychometric approach, as offered by Budoff, Campione and Brown with the adaptation of certain psychometric tests to reflect a more dynamic approach such as Swanson's cognitive processing test which is based on information processing theory (Budoff, 1987a; Campione & Brown, 1987; Feuerstein, Feuerstein, Falik & Rand, 2002; Jensen, 2000; Lauchlan & Elliott, 2001; Minick, 1987; Sternberg, 2000a; Swanson, 2000; Wiedl, 2002; Wiedl, Guthke & Wingenfeld, 1995). The emphasis in dynamic assessment education and clinical set-ups is on people as opposed to scores (Wiedl, 2002) and normative functioning (Das & Naglieri, 1992) which is divergent from the emphasis placed in mainstream preoccupation with only scores in a number conscious society (Tzurriel & Haywood, 1992) (although there is increasing awareness and acceptance of developmentally and criterion or curriculum referenced tests which facilitate further the integration of instruction and assessment; Kahn, 2000; Lidz, 2000b). Dynamic assessments are concerned with understanding and aiding individuals in their development and many models do not emphasise normative profiling or classification to categories of achievement (Jensen, 2000), but such stratified norming does indicate the degree of loss of performance in certain cognitive areas which will then need to be dealt with in terms of understanding why performance is so low (Jepsen, 2000). Assessing for higher education potential is a particularly thorny issue, especially in South Africa where assessment of general thinking skills may allow for greater numbers to access higher education, but without the requisite domain-specific skills many may well struggle to cope with the system.

⁹¹ Recall that once a method or measure or statistical technique is employed to discover data or knowledge, consciously or unconsciously a decision has been made regarding the underlying assumptions of what it is one is studying (Ridgway, 2000). Many may simply be unaware of this.



Moreover, it is well evidenced in the literature that domain general dynamic assessment interventions do not transfer well to other domains, hence the need to dynamically assess domain-specific skills (Samuels, Killip, MacKenzie & Fagan, 1992) which was attempted as early as 1987 in South Africa utilizing Feuerstein's LPAD (Skuy, Archer & Roth, 1987). There is a need for both types depending on the circumstances. A balance between general thinking and domain specificity is needed. It is well known that knowledge structures already in place aid in the generation and accumulation of yet more knowledge in a snow-ball type effect (Minnaert & Janssen, 1996).

Static measures can be utilized and extended dynamically (Lidz & Thomas, 1987) and is a process often turned to in South African research (Murphy, 2002, Murphy & Maree, 2006). The role of assessors is an emphasised concern within dynamic assessment as they do not play a passive role but rather seek to engage the individual in such a way as to encourage development and change within the transactional process (Ghesquière, 2002; Grigorenko & Sternberg, 1998; Lidz, 1997; Minick, 1987). They can be considered as filters for children aiding in the expansion of their cognitive repertoire as well as guiding them in their cognitive choice of strategies (Jensen & Feuerstein, 1987). The assessment procedure can quite rightly be referred to as a consultation (Lidz, 1981). Cultural minorities often experience language difficulties and this is typically the situation in the West where there is limited English language proficiency in some instances;⁹² but this is not necessarily indicative of language deficiency (Lidz, 1997; Losardo & Notari-Syverson, 2001). There is a monumental difference between the two and this difference is often ignored. Hence dynamic assessment interventions and alternative assessments are more often than not non-verbal in mode (Lidz, 2001; McCallum & Bracken, 1997) and rely heavily on pictorial test matter (Schur, Skuy, Zietsman & Fridjhon, 2002) very similar to cross-cultural modes of assessment (Cortada De Kohan, 1972). The main goals of dynamic assessment within a clinical set-up are characterised by concern with the:

- initial level of performance, used within conventional testing but here it informs the mediator as how best to handle the particular individual
- nature and amount of mediation
- nature of the deficient cognitive functions
- non-intellective factors, currently perhaps the most under emphasised aspect within assessment in general
- maintenance and transfer of learning
- type of modality used to channel the test and the
- effects of various mediatory strategies

(Tzuriel, 2001, pp.47-48)

At this juncture in the discussion on dynamic assessment fundamentals it has become evident that the approach is one of assessment intermingled with teaching most often within the framework of Vygotsky's ZPD or some such similar model (Snow, 1990). Table 1 bears testimony to this notion and illustrates the commonalities and differences between assessment and teaching within dynamic assessment focusing specifically on two particular research efforts. Table 2 illustrates the main differences between dynamic and static assessment. Table 3 presents a comparison between the two approaches but from an alternative view point depicting the reality of the situation regarding the utility of both methods of assessment. This thesis, although fundamentally in support of dynamic assessment cannot ignore or put aside the very many positive aspects encountered within traditional forms of intelligence assessment and in so doing the author seeks to objectify arguments within the discipline of psychology which prides itself on its scientific status after all. Figure 6 illustrates the author's understanding of dynamic assessment's take on development as emanating from specific research agendas.

⁹² This is another major point of contention within South African universities at the moment. Determining language proficiency and general aptitude for academic achievement are two completely separate issues in terms of inherent intelligence. Failure to progress due to language problems and failure to progress due to lack of sound schooling or a lower level of intellectual performance are three areas which are often difficult to tell apart. But these aspects will need to be pried apart if one is to successfully choose from a population those who will likely succeed even though language is a barrier as well as prior low quality education as determinant. It is the contention of this thesis that dynamic assessment indeed has a large role to play (even within tertiary education as dynamic assessment is not only applicable to individuals of a certain age; De Beer, 1991) but if it is to do so in any measurable way, it will need to re-look its stance and placement within the larger intelligence assessment field.



Table 1 Assessment and teaching within dynamic assessment: commonalities and differences (Brown, Campione, Webber & McGilly, 1993, p.175)

Assessment and instruction compared	
Zone of proximal development	
Collaborative assessment environments (the work of Campione and colleagues)	Collaborative (reciprocal) teaching environments (the work of Brown and Palinscar and colleagues)
<u>Main similarities</u>	
<p>Based (loosely) on Vygotsky's learning theory Involves guided cooperative learning with expert feedback Strategy modelling by experts (apprenticeship model) Externalizing mental events via discussion formats On-line assessment of novice status Help given, responsive to student needs Aimed at problem solving at the level of control structure (metacognition, intentional learning, learning to learn) Understanding measured by transfer, flexible use of knowledge</p>	
<u>Main differences</u>	
<p>Goal - individual assessment Test - knowledge and strategies Aid - standardized hints Hints - hard to easy to measure student needs</p>	<p>Goal - cooperative learning Teach - knowledge and strategies Aid - opportunistic Hints - easy to hard to scaffold student progress</p>

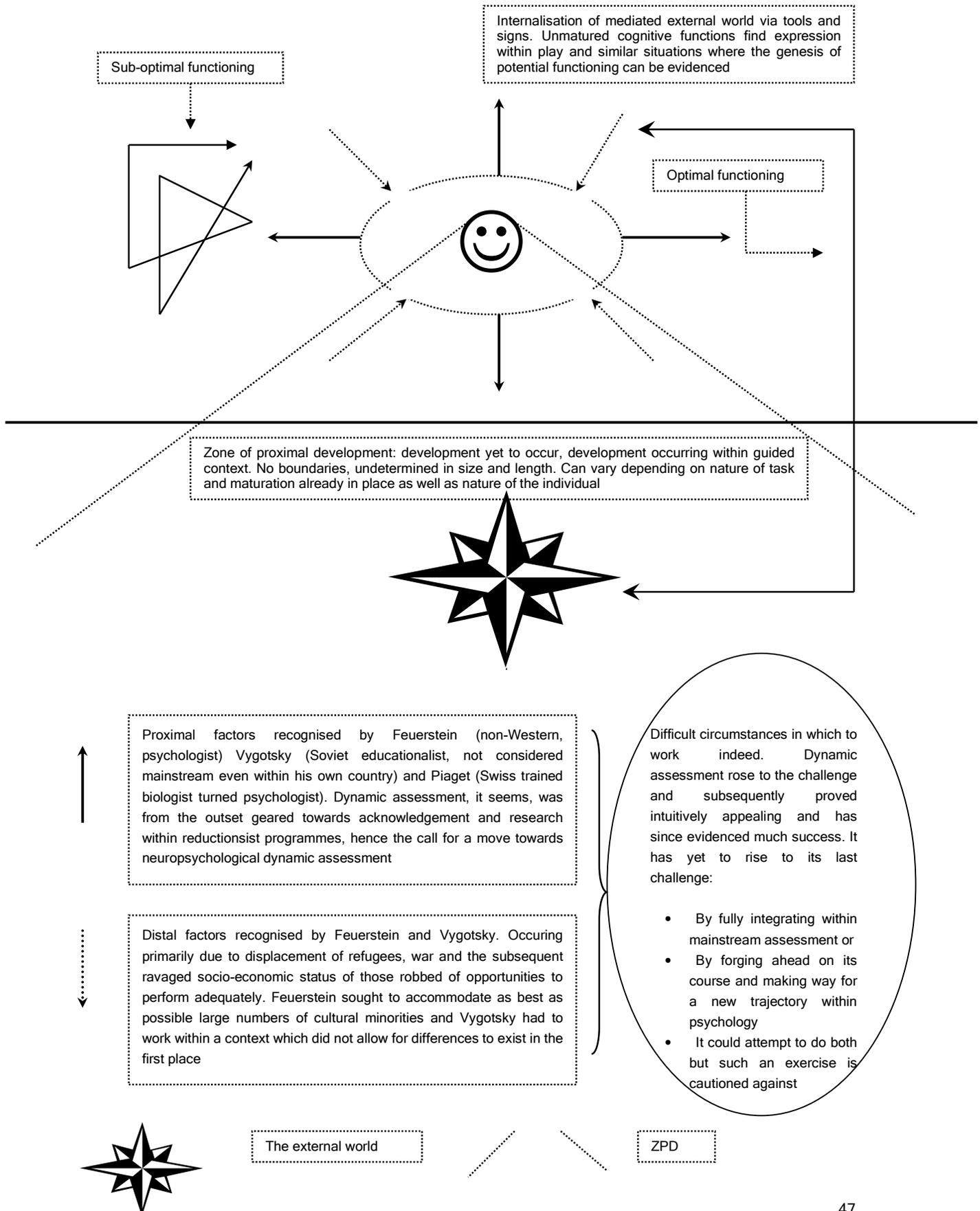
Table 2 Main features differentiating dynamic assessment from static assessment (Morrison, 2001 p.308; Tzuriel, 2001, p.7)

<i>Dimensions of comparison</i>	<i>Dynamic assessment</i>	<i>Standardised testing</i>
Goals of testing	<ul style="list-style-type: none"> • Assessment of change • Assessment of mediation • Assessment of deficient cognitive functions • Assessment of nonintellective factors 	<ul style="list-style-type: none"> • Evaluation of static performance • Comparison with peers • Prediction of future success
View of the learning process	<ul style="list-style-type: none"> • Dynamic, unpredictable 	<ul style="list-style-type: none"> • Predetermined and linear
Orientation	<ul style="list-style-type: none"> • Processes of learning • Metacognitive processes • Understanding of mistakes 	<ul style="list-style-type: none"> • End products (static) • Objective scores • Profile of scores
Assumptions about the learner	<ul style="list-style-type: none"> • Active - invents new knowledge 	<ul style="list-style-type: none"> • Passive - acquires existing knowledge
Context of testing	<ul style="list-style-type: none"> • Dynamic, open, interactive • Guidance, help, and feedback • Feelings of competence • Parents and teachers can observe 	<ul style="list-style-type: none"> • Standardised • Structured • Formal • Parents and teachers are not allowed to observe
Origin of standards	<ul style="list-style-type: none"> • Evolves with the knowledge invented in the process 	<ul style="list-style-type: none"> • Externally determined and applied to the learner
Role of evaluation	<ul style="list-style-type: none"> • Ongoing - part of the learning process itself 	<ul style="list-style-type: none"> • End point of learning
Interpretation	<ul style="list-style-type: none"> • What people learn, what they say they learn and how they learn it 	<ul style="list-style-type: none"> • What learners can produce
Purpose	<ul style="list-style-type: none"> • Recording, interpreting and reinterpreting process 	<ul style="list-style-type: none"> • Measuring and judging according to existing norms
Interpretation of results	<ul style="list-style-type: none"> • Subjective (mainly) • Peak performance • Cognitive modifiability • Deficient cognitive functions • Response to mediation 	<ul style="list-style-type: none"> • Objective (mainly) • Average performance
Assumption about the evaluator	<ul style="list-style-type: none"> • Partial and limited - needs learner's perspective 	<ul style="list-style-type: none"> • Neutral and objective - unchallenged authority
Nature of tasks	<ul style="list-style-type: none"> • Constructed for learning • Graduated for teaching • Guarantee for success 	<ul style="list-style-type: none"> • Based on psychometric properties • Termination after failures
Focus	<ul style="list-style-type: none"> • Whole learner situated in a particular context 	<ul style="list-style-type: none"> • Isolated competencies

Table 3 Playing devil's advocate: some not so sinister aspects of static assessment and some problematic issues within dynamic assessment

<i>Problematic issues within dynamic assessment</i>	<i>Positive aspects within static assessment</i>
<ul style="list-style-type: none"> • As currently practised and envisaged by older models of change assessment, the gain score issue remains problematic although recommendations as to its renewed manipulation within IRT change-based models are considered in chapter 4. • The time taken to administer clinical versions of dynamic assessment interventions are prohibitive within school settings as well as <i>en masse</i> testing of potential tertiary education students • The costs involved are greater due mostly to the length of time needed to administer the assessment • The need to extensively train moderators and mediators within a dynamic assessment set-up is cost and time intensive • Can at this stage only be utilised for underachieving populations which have for any number of reasons not been allowed to explore their fuller potential and has not really entered into assessment of average to above average performing students from normal backgrounds (this is a contentious statement and worthy of more debate) • Perpetuates the misguided notion that dynamic assessment is useful only for sub-performing populations and does not successfully market itself as a packaged product for an all-inclusive manner of assessment • Is not yet effectively linked with biological and neurological psychology thus falling behind in a potentially and manifestly rich area of research where funding would most likely be easier to procure • Is currently “all over the place” in terms of a methodology. Although in fairness, this can be considered a good thing • Original thoughts behind this model were usually just this: thoughts and models. The necessary theoretical backing is sometimes vague, although in some instances this is not the case and as dynamic assessment has moved into the future various models are relying more heavily on sound theory and empirical research 	<ul style="list-style-type: none"> • Has a long history of venerable research attesting to its credibility on a number of issues • Although obviously tainted with nefarious motives throughout its historical development, the field of intelligence was punctuated by steady maturation within modelling of its characteristic traits that cannot be swept aside • Thoughtful and humane approaches towards the assessment of masses of individuals did not always leave much room for intelligence assessment to veer off in directions warranting greater concern for the individual • Dissatisfaction with mainstream assessment has at times appeared fashionable which can be unduly harsh and simply unscientific in rhetoric • There is much to be said for the physiological basis of intelligence and decrying or dismissing reductionist approaches of biological views on intelligence is also not scientific • Contextual, bio-ecological, systemic, environmental and many such development-framed inclusive models of intelligence are intuitively appealing and most likely the more correct explanation of what and how intelligence is and functions. But such models are inherently complicated and are further complicated by the sparse methods currently available to fit them into models evidencing intelligence and potential • Nature evidences a bell curve distribution concerning a number of issues and ignoring this regarding intelligence is not only short-sighted but scientifically unfair and doing so does a disservice to the enterprise of knowledge-gathering. Life in its varied forms is often “unfair” and finding intelligence scores distributed across a spectrum is not something against which we can fight. The situation is that there is such a distribution and the best we can do is to regard each segment with dignity and fairness realising concurrently that no-one is a victim of their biography. Low functioning performers exist, have existed in the past and for the foreseeable future will exist and no amount of interventions will cause this distribution to change. Dynamic assessors must understand this well even though this type of statement is one which is very rarely encountered in dynamic assessment texts. Having said this, the support for dynamic assessment from this author is nevertheless upheld.

Figure 6 Dynamic assessment and the understanding of development



2.7.1 History

Dynamic assessment has a history with disparate origins depending on the manner in which you choose to view it but can be said to have a long past but short history (Haywood & Tzuriel, 2002; Lidz & Elliott, 2000b; Wiedl, 2002). Some may laud Lev Vygotsky as the founding father due to his unique concept of the zone of proximal development within a socio-cultural theory (Elliott, 2003; Hamers, Hessels & Pennings, 1996; Hegarty, 1988), which states that cognitive performance, when aid is provided, will result in the best measure of ability (Hamers & Sijsma, 1993; Meijer, 1993; Shamir & Tzuriel, 2002) thus birthing the learning test (Guthke, 1982). Reality as such is never met with face-to-face but engaged with via tools of mediation (Netchine-Grynberg, 1995). Depending on the level of support received by a child on a task, varying outcomes can be expected; performance at the functional level is expected with minimal or no support as opposed to optimal performance when supported in a task (Suizzo, 2000) and clearly distinguishes between “performance” and “competence” (Gelman, 2000).⁹³ His approach was later built upon by other Soviet educationalists and formulated more extensively regarding its educational implications (Haenen, 2000). Others prefer to view Alfred Binet as the progenitor due to his notion of investigating ability during the process of a test and his idea of a continuously developing latent trait. Binet looked at correct responses as indicative of ability whereas Piaget preferred to look at errors; a novel notion at the time but routinely accepted within current dynamic assessment today (Anastasi, 1998; Chalmers & McGonigle, 2000; Lidz & Thomas, 1987; Sternberg, 1997a; Styles, 1999; Ukrainetz, Harpell, Walsh & Coyle, 2000) although he was not really concerned with remediation (Carlson, 2002).⁹⁴ He also offered interesting ideas on remediation but due to its generality the idea receded into the background (Brown, Campione, Webber & McGilly, 1993).

A brief interruption ...

Galton and Binet are often discussed within the intelligence literature in a dichotomised fashion, the somewhat usual mode or tendency when contrasts wish to be drawn and summated. There is surely usually more to this type of exercise than polarising views, nevertheless, Eysenck (1986) characterises the Galton-Binet differences towards intelligence as follows:

- Cognitive ability and its study was a science to Galton as has already been noted above but to Binet intelligence was a statistical artefact (here it is apparent that substantive construct and abstract construct are already evident - a theme of incongruence that was to play out till the late twentieth century)
- Genetic heritage was a predominant determinant of intelligence within Galtonian intelligence research whilst Binet was concerned with the educational environs (one has to recall that both men were working within two different contexts; the former a scientist and the latter a psychologist interested in education but most academic psychologists in the early twentieth century were more concerned with theoretical issues as opposed to practical ones; Reese, 1993)
- Galtonian measurement was predicated upon physiological bases and Binet's predicates lay within more naturalistic settings
- At times, one feels that some of our forbears have been unkindly treated by history and their reputations besmirched by incidental information vastly disproportional to much exceedingly eminent work contributed to the field of science by these scholars. Subsequently, and in keeping with the human tendency to label and categorise, some scholars become boxed in an ethereal realm from which they have no escape. Galton's work was prodigious by any standard and incredibly wide ranging in thought and scope. Binet was not immune to the tried and tested methods of his time although he did much for the future course of events within intelligence assessment by side-stepping conventionalities. Vygotsky himself did not entirely eschew the intelligence test either. Modern-day researchers should revel in the leaps made by such individuals but be likewise weary of inflating or downplaying various issues which concern us today merely because it is politically correct to do so. History is not biased but our interpretations are

Others would stake a claim for the founding figure to be Reuven Feuerstein who likewise emphasised what individuals did incorrectly in the hope of coming to understand their functioning better and did so within a mediated learning experiential environment (Büchel & Scharnhorst, 1993; Shamir & Tzuriel, 2002). At times both Vygotsky and Feuerstein are credited as equal co-contributors to the field (Haywood & Tzuriel, 2002). Feuerstein may be cited as the founding figure due to the development of his test battery within this domain (Skuy, 1989) and his published and widely recognised work on the subject matter (Lidz, 1992b). It is due largely to Feuerstein and his followers that the renaissance of dynamic assessment has flourished

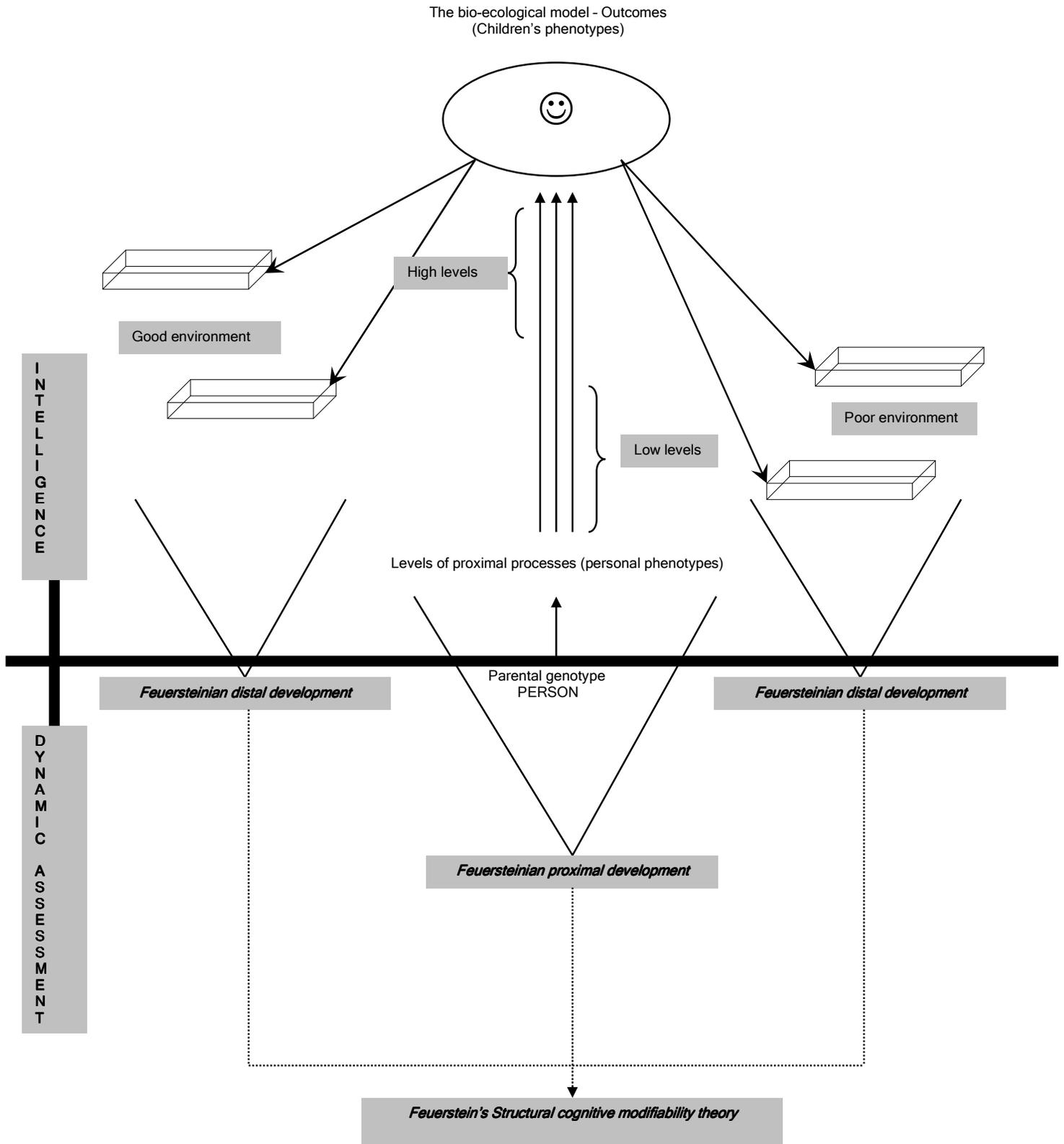
⁹³ It has been shown that students who do well in school due to their parrot fashion learning of material at times experience a decline in marks at tertiary education institutions. Here, their performance is good but competence less so. This may work the other way round for students who have not yet evidenced maturation in certain cognitive areas and hence perform poorly but may well evidence superior competence. This competence-performance distinction is central to educability and assessment of at-risk and mentally retarded individuals (Paour, 1992).

⁹⁴ Incidentally Binet had also measured skull circumferences in his day in keeping with the times, so to paint a non-biased view of this figure as progenitor of dynamic assessment one should be in possession all the facts (Feuerstein & Feuerstein, 2001; Styles, 1999).



as it has especially its gradual development in the West (Sternberg & Grigorenko, 2002). His emphasis on a bio-ecological model of human functioning emphasises distal and proximal factors and thus takes cognisance of indirectly impinging variables such as genetic heritage, early childhood experiences as well as cultural and socio-economic factors playing in on the developing child. It is in keeping with developmental theories of intelligence such as evidenced by Ceci's bio-ecological theory (Ceci & Bruck, 1994; Ceci, Rosenblum, De Bruyn & Lee, 1997; Lohman, 2005; Miller, 1997). The similarity of the bio-ecological and Feuersteinian theories/models can be viewed in figure 7 below. Some of the early literature dating from the 1920's and onward regarding educability and pedagogy in general is quite prescient really in its treatment of measuring the learning process (Brown, Campione, Webber & McGilly, 1993). Although not making up the majority of the work a fair section of the material is pleasingly modern in outlook which leads one to wonder why certain ideas have indeed taken so long to gain a stronghold within mainstream pedagogy, education and assessment. De Weerd (1927) for instance, begins her article with a statement echoing sentiments eerily similar to the ones heard today "the whole scheme of formal education is based upon the fundamental concept of *improvability*" (own emphasis) (p.547). She does go on to state that "the educator has always been interested in this *capacity* of the individual and has measured it in a more or less direct way through class achievement ... we have relatively little material on the *learning* or *improvability* of children under classroom conditions" (own emphasis) (p.547). This was written seventy nine years ago.

Figure 7 Intelligence-dynamic assessment link: the bio-ecological intelligence model and the Feuersteinian dynamic assessment model (bio-ecological model from Ceci, Rosenblum, De Bruyn & Lee, 1997, p.314)



SCM (structural cognitive modifiability) assumes that modification ensues regardless of the etiology and severity of a condition as well as regardless of age. This is due to the brain's plasticity and lifelong engagement in learning (Birnbaum & Deutsch, 1996).



Socio-economic factors as well as cultural mediatory factors and their influences are felt as early as the third grade in school (Portes & Vadeboncoeur, 2003). It would make sense to intervene as soon as possible even if only from a neuronal development point of view where synapses in the frontal and occipital cortex areas for instance undergo growth until the age of twenty (if not longer) (Neville, 1994). This is in addition to the fact that remediation in general becomes successively more difficult as age increases (Humphreys, 1988) (but not necessarily impossible!). Intellectual factors are only a part of the greater spectrum of aspects influencing development and mediated learning experience places cognitive, emotional and cultural dimensions on a par with intellectual factors in the model (Skuy, 2002). Bio-ecological theory supports this notion as it has been shown that environmental influence has differential effectual outcomes on heritability (h) where h is a notion of populations and not individuals (Ceci & Bruck, 1994; Conway, 2005; Gordon & Lemons, 1997). It represents the degree of phenotypical (observed) differences within a population; that is results from genetic differences within the population (Grigorenko, 2004a). "Heritability is the ratio of genetic variation to total variation in an attribute *within* a population" (original emphasis) (Sternberg, Grigorenko & Kidd, 2005, p.53) and explains percentage variation accounted for, not the construct attested to. Feuerstein's emphasis on the modifiability of cognitive processes can be traced back to the time he studied under Piaget whose models of child cognition are synonymous with the developmental tradition but he later parted ways with Piaget to study under André Rey (Burgess, 2000).

Together with mediated learning experience in which the environmental challenges are filtered and attenuated for the child and which also outweigh the distal factors in the magnitude of its influence, the governing model of human development is all-encompassing and indeed very modern for the time in which it was being proffered as viable alternative to mainstream testing (Kozulin, 2002a). Vygotsky's socio-cultural approach and Feuerstein's mediated learning approach both reject the dichotomous appraisal of cognition as naturally occurring phenomenon and learning or instruction as a cultural tool. Vygotsky viewed learning, culture and development as inseparable (Mastergeorge, 2001) as any instructional interaction reflects a social interaction (Perret-Clermont & Bell, 1987). Human beings, unlike other species,⁹⁵ interact with intentional agents and engage in cultural learning from as early as nine months and acquire linguistic and other symbols necessary for communication within their cultural grouping (Tomasello, 2001). Culture and development integrate seamlessly within the learning and processing situation (Kozulin, 2002b) but Miller (1997) warns that although cultural aspects of psychometric intelligence often results in the revisiting of specific theories of intelligence, it is not always the case that the nature of intelligence itself is revisited. Cognition is framed within culture as "cultural ontogeny affects biology in a very direct way" (Feuerstein, Feuerstein, Falik & Rand, 2002, p.73), although cogent arguments are made for other sides to the debate of where and how intelligence arises. Culture, states Hunt (1997) cannot account for measurable variables and so there is no way of tracing their effects within a causal model. There are many other individual researchers as well as research groupings which can attest to having developed their own unique manner of dynamic assessment (Guthke, 1993b; Hamers & Resing, 1993; Laughton, 1990; Lidz, 1981, 2003; Resing, 1993) but most of these schools find their points of origin in one or more of those mentioned whilst some groups prefer to blend in contributions from various originators to fit their needs (Shamir & Tzuriel, 2002). The literature is often divided on the issue of the history of dynamic assessment with statements revealing that it is relatively new to statements revealing its venerable and aged heritage (Hamers, Hessels & Tissink, 1995). What is most likely being hinted at is that philosophically, the method is old and can be traced back to the early nineteenth century (despite the early Greek's writings on potential). The method is likewise young in terms of its being generally known about and widely practised or conversely not well-known even today (Freeman & Miller, 2001; Haywood & Tzuriel, 2002). Recent research evidences that dynamic assessment still has not left much of a noticeable footprint outside small societies advocating its message. One retort to this is that this is very much the story of science where numerous ideas have had to wait years, decades and in some instances centuries before the ideas latched on in the mainstream imagination. The alternative retort is that this is very much an indication of its potential demise as equally many "good" ideas in the history of science and social science have been relegated to anecdotal status. Who can tell? This is an "unknown"; the one "known" is that the idea underlying the philosophy is unflinching in its tenacity to continually surface through the ages. One must ask oneself: why? The intuitive appeal of the idea continues unabated.

Many schools of thought, have, since the days of Binet, Vygotsky and Feuerstein developed within a framework which can be considered "neo"; as in neo-Piagetian and neo-Vygotskian and so on. However, a number of isolated researchers had, in the early decades of the twentieth century, already experimented with the concept of a malleable intelligence (Lidz, 1992a) including, among others, Otto Selz (1935) (in Klauer, 2002); G.R Ortat (1959), E. Haeussermann (1958) considered the mother of dynamic assessment (Lidz & Elliott, 2000b), H. Schucman (1960) (in Lidz, 1991, 2001) with South Africa reporting the results of a dynamic-like type of assessment as early as 1961 (Lloyd & Pidgeon, 1961). A pupil of Wundt, Meumann (1922), the psychometrician Kern (1930) and DeWeerd (1927) are other names in the early decades of the twentieth century to whom the idea of a learning test concept can be traced (Guthke, 1992; Guthke, Beckmann & Dobat, 1997). European pioneers within this tradition trace origins to the works of, among others, Quetelet (1835) the Belgian researcher (see above) and his contemporary

⁹⁵ This is contentious. There is evidence that higher primates do engage with one another as intentional agents and transmit information about symbols to each other (Rumbaugh, 2002). More research is however needed to vindicate such a claim.



Esquirol (1838), a French psychiatrist, who argued for a differentiation to be made between performance and prognoses (unfavourable conditions can lead to unfavourable outcomes⁹⁶). Binet and Simon, (1908), Stern (1928), a German psychologist who studied the relation between environmental influences and intellectual abilities; Vygotsky (along with his students Luria and Leontiev who modified his theory of ZPD; Hamers, Hessels & Pennings, 1996), Kern (1930), who was concerned with improving intellectual performance via training and using trainability as diagnostic indicator of aptitude; Selz (1935), who concentrated on modifiability of intelligence; Piaget, Rey (Feuerstein's mentor), who researched mental plasticity; Volle (1957), who pioneered the testing-the-limits approach outside the domain of Rorschach testing where it had previously been utilised and concluded that the manner in which testing was conducted influenced the outcome, especially in low performing individuals. Boesch (1952, 1964), was influenced by Piaget and concentrated on patterns and variations of results in place of viewing only correct responses on a test. Hurtig (1962), the French psychologist who made strides in differentiating children who had suffered environmental backlogs and those who were truly retarded and Schmidt, (1969, 1971) who drew heavily on the work of Boesch and utilised his theory but further refined the testing-the-limits approach by integration of empirical and theoretical data (in Wiedl, Guthke & Wingenfeld, 1995).

European contributions towards the historical development of dynamic assessment, on the surface, does seem to have a richer history as opposed to the Western development which may partially explain why dynamic assessment has been more eagerly followed and accepted within countries outside the United States and the United Kingdom. Possibly this has resulted in more resistance to this method from mainstream assessors currently more at ease with typical intelligence tests. This conclusion is substantiated only from what has been gleaned from the literature as well as from results from studies dealing with dynamic assessment awareness surveys. Early 1920-1930's research emphasised the ability to learn as paramount and was considered, at times, more informative than the static measures hitherto used (Lidz, 1987a). It was during the 1930-1950's that the neuropsychologist Andre Rey formulated tests which were utilised later on by his student Feuerstein in the 1950's but these two decades were a period of relative quiet on the dynamic assessment front, mostly due to a variety of social, political and scientific reasons (Kozulin & Garb, 2002; Pascual-Leone, Johnson, Baskind, Dworsky & Severtson, 2000). The 1940's evidenced the insight garnered to detail the relation between intellectual achievement and intellectual potential noting the lack of covariation between the two measures (Lidz, 1987a). Practice effects had been studied in the 1920's but coaching studies came into their own in the 1950's. Original attempts at documenting the effects of coaching on performance were mainly conducted in the United Kingdom, although the emphasis was not on the improvement of capacity as such (Lidz, 1987a). Process orientated means of assessing intelligence as well as emphasis on testing the limits to which individuals could aspire were also characteristic of the 1950's and incidentally this process nature of intelligence assessment is sometimes considered quite modern in approach where there is a turn towards assessing what individuals can and cannot do within their own limits (Baltes, 1998). Sporadic attempts at assessing educability took centre stage during the 1960's and interestingly, it was A.R. Jensen who very early on stated his views concerning the biased nature of mainstream intellectual assessment batteries, deferring to more dynamic alternatives for various minority groups (Lidz, 1987a) and it should be noted that ecocultural factors also need to be factored into the readiness to limit bias even within similar cultural settings (Church & Katigbak, 1987). Learning potential measures in the United States were also becoming increasingly evident during this decade (Budoff, 1987a) due predominantly to increasing dissatisfaction with psychometric testing (Kozulin, 2005).

The 1970's witnessed an explosion of sorts in comparison to the previous decades and saw the deployment of dynamic assessment methods of mental abilities (Carlson, 1994). This decade was witness to the filtering in of cognitive science principles into the arena of intelligence; hence the hybridised field one sees today (Das, Naglieri & Kirby, 1994). During this phase, various research groups came to the fore with various points of emphases under the umbrella term "dynamic assessment" (which is partly the reason why this manner of assessment is so varied in its definitions and practical implementations). Included among others are the works of Feuerstein (who had, as mentioned, laid the ground work decades prior to this) as well as Budoff, Campione, Brown and Haywood (instrumental in bringing to the United States the work of Feuerstein) (Lidz, 1987a). Carlson and Wiedl had started their work on testing-the-limits approach in the 1970's and on into the 1980's and provided evidence for the validity of dynamic assessment as alternative approach, working specifically within the information processing paradigm (Lidz, 1987a). The 1990's and early decades of the twenty first century can be characterised by a more sedate and sceptical attitude towards this approach but ironically was simultaneously only starting to take off in South Africa (Murphy, 2002; Murphy & Maree, 2006). It was also expanding its field of research application to populations other than the traditional sphere of application, and includes among others mature students, the elderly,⁹⁷ gifted (culturally diverse and

⁹⁶ However obvious this may seem to us today in 2006, recall the period in which these esteemed researchers worked. The early 1800's was a period chiefly characterised by industrial growth, scientific development (enlightenment), the need for labour, urban migration, slow but steady creation of a rising middle class, interest in culture, greater apportionment of time to leisure activities, internecine war and an altogether renewed attempt to understand human beings and their place within the world, cosmos and universe. The founding and consolidation of new empires, countries and territories resulted in independence of many forms. It is not surprising then, that studies into mental faculties started to blossom when they did (Brinkley, 1993; Delius, Gatzemeier, Sertcan & Wünsch, 2000; Gribbin, 2003; Roberts, 1995).

⁹⁷ Lest it be forgotten, neurons are created in the adult brain - some hope for dynamic assessment neuronal correlates approach that is advocated in this treatise (Quartz & Seknowski, 1997b).

usually undetected), at-risk foreign language learners, psychiatric, prison inmates, early brain damaged individuals, blind, deaf and speech impaired populations and is applicable to both domain general and specific contexts (language, arithmetic, writing competence, science and biology for instance) (Alfassi, 2002; Bolig & Day, 1993; Chan, Ashman & Van Kraayenoord, 2000; Glaspey & Stoel-Gammon, 2005; Guthke, 1992; Kaniel & Tzuriel, 1992; Keane, 1987; Keane, Tannenbaum & Krapf, 1992; Kester, Pena & Gillam, 2001; Kirschenbaum, 1998; Kozulin, 2000; Kozulin & Garb, 2002; Lidz, 2004; Lidz & Elliott, 2000b; Samuels, 2000; Schur, Skuy, Zietsman & Fridjhon, 2002; Silverman & Waksman, 1992; Stanley, Siegel, Cooper & Marshall, 1995; Schneider & Ganschow, 2000; Tzuriel, 2000a, 2001; Ukrainetz, Harpell, Walsh & Coyle, 2000; Wiedl, 2002; Wiedl, Guthke & Wingenfeld, 1995; Wiedl & Schlittke, 1995). These are populations which have until recently been side-lined from mainstream assessment and interventions and for whom targeted mediatory intervention programmes are deemed more suitable due their unique life contexts. Dynamic assessment is also moving into areas concerned with neurological disorders where both individual and group administration of assessment is being researched (Haywood & Miller, 2003). There is some conceptual overlap, namely, the approach towards understanding what is maximally possible after sustained injury as opposed to what is typical functioning and the effect of actual intervention within the injured patient as opposed to merely assessing it (Haywood & Miller, 2003). Perhaps the most valuable asset within the dynamic assessment arsenal kit!

Issues such as psychometric aspects, costs and time are major factors continuously lamented as constraints within this method of assessment. However, a number of prospects have since the 1990's become available in order to allow this manner of assessment a more palatable existence. New models within item response theory are currently being developed as this thesis is being written (see chapter 4 for change-based IRT models) which, simply stated, provide enhanced technology for the generation of items and assembly of tests but is as yet unable to provide deeper insight into the process of intelligent functioning (all the more reason to look towards biological models of intelligence) (Schmiedek, 2005). The ideas of measuring change are not new, but it has only been within the last two decades that dynamic assessment has come to the fore in the literature (Guthke, Beckmann & Dohat, 1997; Lauchlan & Elliott, 2001). Although not by any means exclusively so, a sizeable research output within dynamic assessment is conducted in the following countries: the Netherlands, Israel, Germany, United States of America, Canada, Belgium, Europe in general, the United Kingdom and South Africa. Research output from South America, Australia and elsewhere is also evident but the mainstay emanates from the first four countries. Dynamic assessment is often greeted with enthusiasm by many practitioners and school psychologists in terms of its theoretical underpinnings and has received attention from many eminent scholars within the intelligence assessment field. Despite the very often positive mentions of this approach it is nevertheless plagued by a number of disadvantages which are similarly highlighted by these scholars; this study being merely one such attempt at providing a framework to aid in guiding the sub-discipline. The most often cited aspect which it is assumed will aid in the theoretical and methodological growth of dynamic assessment is the advent and development of various item response models (Pennings & Verhelst, 1993; Resing, 1993; Schöttke, Bartram & Wiedl, 1993; Sijtsma, 1993a, 1993b) which will go some way in making dynamic assessment more psychometrically sound - perhaps the most criticised aspect of this approach (Minnaert, 2002). These models are based on newer conceptualisations of what it means to validate a construct as well as a move away from the more traditional concept of construct validity⁹⁸ as is currently endorsed in some dynamic assessment literature (Carlson & Wiedl, 2000) (more on this topic is discussed in chapter 4). Nevertheless, the current trend within psychological assessment is to view dynamic assessment methods as complementary to mainstream assessment (Büchel & Scharnhorst, 1993; Minnaert, 2002; Resing & Roth-Van der Werf, 2002). In some instances though, dynamic assessment is utilised in an incidental fashion such as serving informal functions within a more traditional criterion and norm referenced manner (Freeman & Miller, 2001). The need to more readily integrate the two methods is called for.

The one common thread specific to dynamic assessment is its role within cultural assessment scenarios (Van de Vijver, 1993), and although not in any way limited to assessments of culturally diverse groups, finds its niche in this context specifically due to the method's manner of assessing for a largely "untouched" construct of ability; i.e. pure ability devoid of cultural influences (Van de Vijver & Jongmans, 2002). Vygotsky's task was one of assessing intellect in a socio-political sphere which denied that intellectual differences existed (Gindis, 1995a; Guthke, 1993b; Guthke, Beckmann & Dohat, 1997) and where ignoring the issue of social class and the reduction of the psychological to the physiological was frowned upon (Gilgen, 2000; Van der Veer, 2000). Yet ironically, this was a time in Soviet history where support for educational psychology, both moral and financial was considerable (Haenen, 2000). This prompted the furtherance of his approach in treating handicapped children by advocating the need to redress the influence of the socio-cultural environment (particularly the adult environment in which the child grew up) in which the disabled person functioned (Gindis, 1995a; Kerr, 1997). His socio-cultural research programme however began to disintegrate during the 1930's and was replaced by the burgeoning of activity theory (considered a branch of Vygotsky's cultural-historical approach) which bespoke of the activities humans were involved in as crucial towards the development and emergence of mind (brain) (Toomela, 2000). Likewise, Binet's task, to assess for and define different levels of ability within different economic groupings in France prompted his manner of assessment. Socio-economic factors play a negative role when

⁹⁸ Which has been referred to as a continuum as opposed to an all-or-none concept (Gross, 2002). This reflects the myriad definitions for concepts such as intelligence for instance.



assessing for intelligence, especially in societies which base such assessments on the prevailing schooling (Tellegen & Laros, 1993). Poorer individuals are often less likely to be able to afford better education and so the cycle ensues but the need to identify gifted students within economically and culturally varied populations manifests and the utilisation of dynamic assessment as technique of assessment-intervention is uniquely placed to assist in this assessment of such populations as well (Lidz, 2001).

Feuerstein's need to assess immigrants seeking asylum in Israel (largely an immigrant country) from as far a field as Ethiopia and Morocco also resulted in a the manner of testing now known as learning potential assessment (Deutsch, 2003; Feuerstein & Feuerstein, 2001; Goldberg, 1991; Gutiérrez-Clellen & Pena, 2001; Kozulin, 2002b; Tzuriel & Haywood, 1992; Zeidner, Matthews & Roberts, 2004). Wide-scale disruption within Europe during and after the second world war resulted in the displacement of countless children and adults seeking residence elsewhere where forms of assessment were woefully inadequate in terms of assessing different cultural groups as well with dealing with the added effects of war (Morphet, 1986; Tzuriel, 2001). Learning within traditional and rural Ethiopian culture was propagated largely via means of imitation and observation and was therefore not verbally based (Katz, Kizony & Parush, 2002). Moroccan, Turkish and Surinam immigrants living in the Netherlands have also benefited from dynamic assessments (Hessels & Hamers, 1993) and the pressing need to continue with dynamic assessment research in the Netherlands can be supported by the ever-increasing number of minority groups in this country (Hessels, 2000). Kozulin and Presseisen (1995) differentiate between types of individuals whose higher order cognitive development and mediated learning experiences can effectively be categorised into one of four areas. The typical categorisation for displaced persons' lack of mediated learning experience is due to displacement of some sort and the concomitant lack of developed higher order thinking processes results in the lack of adequate education. It is not difficult then, to understand why dynamic assessment has gained wider acceptance in the countries mentioned above as opposed to countries which still propound mainstream assessment. There is a need to assess minority groupings (Ruijsenaars, Castelijns & Hamers, 1993) who often present with substantially lower levels of cognitive functioning when assessed within mainstream testing situations (Hessels & Hessels-Schlatter, 2002). These attempts can be categorised as culturally sensitive assessments or ecologically more inclusive (Guthke, 1993a; Lidz, 1981), sensitive to not only the culture but also sensitive to deviations within the specific culture. In other words one must pry apart cultural difference (those not in need of intervention but who may be in need of a culturally-sensitive tool) and those suffering from cultural deprivation (in need of cognitive deficit intervention) (Jensen, Feuerstein, Rand, Kaniel & Tzuriel, 1988; Rosas, 2004; Tzuriel, 2000b). Cultural deprivation occurs in any culture and conformance to one culture as evidenced in typical behaviour can be vastly at odds with another culture's normative behaviour (Helms, 1997).⁹⁹ Cultural diversity is seen in many contexts as an asset within any intervention (Robinson-Zanartu & Aganza, 2000) and should be capitalised upon, not dismantled and set aside. Emanating from a different culture and having been the recipient of poor mediational learning experience are two very different etiologies of poor performance (Feuerstein, Rand, Hoffman & Miller, 1980). The child who has received poor mediation will evidence blurred perception of reality, be apathetic and withdrawn and their subsequent isolation results in their fragmented experience of the world (Klein, 2000). There is of course also another sub-group of low performers who evidence neurological problems (Swanson, 2000) and so the landscape of poor performance does not represent an even or smooth territory. Cultural deficit typically occurs within a culture; in other words, delayed development as specific to the culture in question using cultural norms against which to judge deficits. This definition of cultural deficit is then not at all similar to the notion of "other culture" deficit according to which one culture is subverted by another. Culture as utilised within dynamic assessment is thus not tinged with biologically inherited differences between cultures but deficits within cultures (Helms, 1992). This is of particular importance within the dynamic assessment approach towards understanding intellectual functioning and veers away from mainstream intelligence work in this regard.

Regarding the more traditional notion of therapy within psychological science, Haywood (2000) has made important and insightful strides in combining cognitive education into psychotherapeutic interventions where many misdiagnosable aspects of behaviour can be seen to emanate from errors in cognitive thinking as opposed to traditionally accepted pathologies of behaviour. Such an interdisciplinary tool can and has proven thoughtful in approaching sub-optimal performance in culturally different populations. This is clearly not the picture presented by those who merely originate from another culture. The latter is deemed very important as it assumes a fully functioning cultural society without the subsequent need to test these cultures via mainstream and largely Western intelligence batteries which more often than not results in bias to some degree (Hessels & Hamers, 1993). The notion that in order to assess non-westerners one needs to treat people as educable mentally retarded (Sternberg, Grigorenko, Ngorosho, Tantufuye, Mbise, Nokes, Jukes & Bundy, 2002) says more about the test scenario and the reigning concept of intelligence than it does about the various test populations! The need to, for instance, differentiate between children with language differences vs. those with language impairment is of concern and is an area specifically attuned to dynamic assessment approaches (Gillam, Pena & Miller, 1999; Gutiérrez-Clellen & Pena, 2001; Pena & Gillam, 2000). Developing expertise is a notion not favoured by an exclusive reliance on Western concepts of what it means to be "intelligent" or well equipped. Developing expertise is culture-wide phenomena which is difficult to assess with Western testing methods only

⁹⁹ In this regard the reader is urged to take a look at table 26.1 p.521 in D.P.Flanagan, J.L. Genshaft & P.L. Harrison (1997). It is interesting that such notions took such a long time to seep through to mainstream understandings of how best to go about assessing for "intelligence" between different cultures.

(Sternberg & Grigorenko, 2002; Sternberg & Grigorenko, 2001c). Language proficiency bias has also posed serious problems especially for immigrant populations (the majority of whom are children) resulting in confusing learning disabled populations with culturally diverse ones (Cole, 1996b). This critical aspect reflects dynamic assessment's concern with distinguishing between cultural difference and cultural deprivation (Tzuriel, 2002) and can be closely interwoven with both Feuerstein's and Vygotsky's conceptualisation of mediated learning experience and higher order thinking processes (Kozulin & Presseisen, 1995). Immediately, the predicates of dynamic assessment can be seen. Research has also shown the efficacy of utilising dynamic assessment within environments catering for mental retardation as opposed to the more classically aligned mainstream assessments (Hessles-Schlatter, 2002a, 2002b). Metaphorical conceptions of intelligence are the preferred analogies within dynamic assessment where the sociological and anthropological contexts are of deep and consistent concern (Campbell & Carlson, 1995). Change directedness occurs in a second-order manner where change is effected from within the individual and is not imposed from without (pouring information into a system versus the deeper understanding of material from within the system).

2.7.1.1 Soviet ideology and Vygotsky's manoeuvrability within the system

As discussed above, depending on who is viewed as the "founder" of dynamic assessment different emphases can be placed on various aspects within the larger movement. The author has decided to elaborate on the times and work of Lev Semenovich Vygotsky¹⁰⁰ (1896-1934) for the following reasons:

- Vygotsky's work is very well known within the field yet one finds in the mainstream literature that his thoughts are merely descriptive of his opinions regarding the zone of proximal development with little elaboration into the reasons behind this thinking
- Understanding Vygotsky's importance for dynamic assessment is better understood when his thoughts are contextualised perhaps more so due to the nature of Soviet psychology's history (Sutton, 1988)
- Vygotsky was the first researcher within this tradition to develop and offer a more systematised approach towards a newly restructured "defectology" in Russia (which had already a long history in the Soviet Union)
- Vygotsky was separated both in place and time from the continental European and American trends within psychology yet his ideas have proved so fertile in these geographic disparate locations (Resing, 1997)
- Soviet psychology¹⁰¹ often followed its own trajectory on a number of fronts within psychology and education and more often than not had to dance to the tune of the ruling political party (not always the case but often so). Academic and research work was often politically infused with ideology¹⁰² and reflects a substantial difference between the West in terms of intellectual tradition and approach to the study of psychology
- In keeping with the thesis's main emphasis on meta-theory, Vygotsky's work is perfectly placed to better understand the times during which he worked even though some of this ideas are quite modern (Gouws, 1997). The meta-theoretical framework that has been developed in chapter 3 and 4 will not be implemented here but the spirit of the framework will be felt (see chapter 3 for Madsen's meta-theoretical framework)
- The collectivist background from which Soviet psychological science sprung has much philosophically and ideologically in common with a similar notion of collectivism in South Africa even though this has and is undergoing remarkable changes in both countries; it is the historical origins with which this brief section concerns itself
- Vygotsky's ruminations over a "psychological crisis" is similar to one discussed within this study (see chapter 3). Where was and is the discipline headed? Recall that he was not a psychologist by training (he studied law and philology; Cole & Scribner, 1978; Kerr, 1997 and professionally practised psychology for only ten years; Grigorenko & Kornilova, 1997) and so perhaps his views were less influenced by specific context. He was primarily a thinker for whom the area of psychology suited his vision for application (Kozulin, 1990)
- It has been remarked that the research work of Vygotsky and Piaget (and many others besides) was, to put it bluntly, quite mundane when considered today. But this is precisely the point; when one considers that the former was born at the end of the nineteenth century and worked under, at times, rather bizarre conditions and made almost galactic¹⁰³ strides in terms of the then current thinking can one and should one hold such individuals to current standards? It seems that much of the literature regards them with due respect but the author has on occasion reasoned away the necessity of having to once again deliberate on the works and times of Vygotsky when there is now well over half a

¹⁰⁰ Depending on the text available, one will find his name as Semionovich Vygotskii as well as various other renditions. For the sake of consistency the above will be utilised. The same applies to other Russian names.

¹⁰¹ It has been the author's experience that much of the literature on Soviet psychology is tinged with ill-defined concepts which may be due to the direct translations of works, phrases and words from the original Russian. Sometimes, it is difficult to appreciate the essence of an argument when given in vague conceptions.

¹⁰² The reader must be cognisant of the fact that much of Soviet psychological work remains to be translated and when it is translated is up to the discretion of the particular researcher's interests and perhaps tacit biases. As Wertsch (1981) remarks, Soviet psychology should be understood in terms of its underlying philosophy which can of course also be said of Western psychology.

¹⁰³ Some even consider it futuristic, so much so that time has yet to unravel the fullest grandeur of his theory (Robbins, 2001).



decade of research and development in the area of dynamic assessment. The following digression seeks to imbibe the reader with a sense or feel of what it may have been like to inhabit Vygotsky's shoes, at least from what can be ascertained from the literature and to paint a picture of why the man thought as he did given his circumstances. It is always necessary to revisit the past if one is to make sense of the present and this is as true for dynamic assessment as it is for any discipline for in order to more fully understand where we are today, understanding the past is a prerequisite

- Gillen (2000) cogently argues for a closer understanding of what has made Vygotsky as great as he is currently. The article is insightful and has made the author re-consider her position on Vygotsky or at least on her position concerning others' views on him. Notwithstanding the article the brief discussion is still, it is thought, warranted
- The author has been told on occasion that it is never a good reason to conduct research because it is interesting, but the retort to this is that Soviet psychology is unquestioningly interesting and this area of work hardly ever features in course work within psychology, at least not in South Africa. This brief digression will focus more so on the times during which Vygotsky was active more so than on his theory of ZPD per se and will not focus on the many fine issues (phylogeny and ontogeny as well as thought and language issues) within the works of Vygotsky due to space limitations

By way of a very brief summary to Soviet psychology, table 4 below encapsulates the nature of the discipline's path through the period spanning two hundred and fifty years. The trajectory followed throughout the twentieth century will become clearer during the discussion below which focuses on historical impingements on Soviet psychology.¹⁰⁴

¹⁰⁴ Sahakian (1975) uses "Russian" and the author has used "Soviet". After the dissolution of the Soviet Union, "Russian" is usually used. However for sake of clarity, the author uses Sahakian's (1975) choice.

Table 4 Placement of dynamic assessment precursors within the broader context of Russian psychological history (Sahakian, 1975)

A cursory glance: Pedology as placed within the broader context of Russian psychological history	
<i>Precursors of Russian psychology</i>	<ul style="list-style-type: none"> • The Russian enlightenment - chiefly characterised by M.V. Lomonosov (1711-1765) who founded the Moscow University in 1755 and is considered the founder of Russian philosophical materialism • Russian associational psychology - P.M Lyubovsky, who initiated experimental psychology and put forward the notion of associationism as explanation of mental processes (the mind is organised via a process of organisation emanating from the ideas of Locke and James; Harnish, 2002). Binet, in France chose to break away from mainstream associationism thus propelling the study of intelligence towards the direct measurement of processes as opposed to the then current manner of associationist research (Lautrey & De Ribaupierre, 2004) • The revolutionary democrats - materialistic psychology, active in and around the 1860's concerned with issues such as psychology's placement within the broader field of science, its link to other areas of science and the questions surrounding physiological and mental processes • Early experimental psychology - founded by N.N. Lange (1858-1921) who founded one of the earliest psychological laboratories. This period also witnessed progress in personality, general and comparative psychology
<i>The reflex period</i>	<ul style="list-style-type: none"> • Reflex theory of mental activity - I.M. Sechenov (1829-1905); considered the founder of Russian physiology • Conditioned reflex or classical conditioning - I.P.Pavlov (1890-1936) • Typology - the study of human individual differences and psychological types - B.M. Teplov (1896-1965) • Orienting reflex - E.N. Sokolov who investigated the neural mechanisms of the orienting reflex • Reflexology - V.M. Bekhterev (1867-1927) who influenced Watson but was historically eclipsed by Pavlov • Reactology - K.N. Kornilov (1879-1957) - who challenged and criticised the work of Bekhterev, stressing the link between the physical and mental, hence the reaction of organisms to the environment (biosociological) • Theory of the dominant - A.A. Ukhtomsky (1884-1942) whose work centred chiefly around neurology thus further inspiring experimental psychology
<i>The pedological period</i>	<ul style="list-style-type: none"> • Pedological theory - P.P.Blonsky (1884-1941) who emphasised the importance of the relation between heredity and environment within child psychology
<i>The dialectical era</i>	<ul style="list-style-type: none"> • Dialectical materialism (1930-1950) was the main authoritative representation of Russian psychology during the period of test bans. Prior to this movement human beings were considered products of genetic heritage and environmental influences. Education as notion now intervened heavily. This movement was a fusion of German Hegelian and Feuerbachian philosophies mixed with Marxist thought culminating in the primacy of "the social" in many aspects of theories developed during this period • Sociohistorical or cultural historical development - L.S. Vygotsky, A.R. Luria and A. Leontiev (more on these researchers in the discussion below) • Formation of psyche as activity - S.L. Rubinstein (1889-1960) who sought and advocated the unity between the mental and the physical; that mental is fully integrated into the physical development of the organism, that human beings change as society changes and that theory and practice should function as a unit • Involuntary memory - A.A. Smimov and P.I. Zinchenko who viewed involuntary memory as the result of goal-directed behaviour (i.e. learning and retention is aided by the nature of the task and its context)
<i>Psychology in Armenia and Georgia</i>	<ul style="list-style-type: none"> • Relevant research

Learning disability research ("defectology") features heavily in the history of Soviet psychology (Bauer, 1959; Coles, 1982; Goldberg, 1982) and represents a substantial segment of this psychology. Special education started as early as 1806 where the state supported schools for deaf and blind children and for a period Russia was on par with like-minded Europe (Malofeev,

1998). Biological Soviet psychology for instance kept abreast of the latest written work from Europe in order that they not start to lag behind but this was at odds, at least ideologically, with the reigning Tsarist regime whose philosophical approach to life did not include reductionist models based on Western science (Wundt's work was criticised by both leftist and rightist groups!) Yet, in what can only be considered a fantastic turn around, Russian progress in materialist monism originated and continued unabated during the latter half of the nineteenth century with many decrying Wundt's status as originator of the laboratory tradition which was concurrently being instituted in Russia (McLeish, 1975; Todes, 1984). Russian psychology was in no small measure unaffected by research in Europe¹⁰⁵ both before and after the revolution (Kozulin, 1986) and special education in Russia developed rapidly during the 1920's after the loss of resources during the revolution (Grigorenko, 1998). The war resulted in many displaced people having lost out to education, a scenario similar to that experienced by many refugees the world over. The Institute of Defectology opened its doors to the study and treatment of children with developmental delays in 1929. The theoretical work of Vygotsky (who was appointed as scientific leader; Van der Veer & Valsiner, 1991) emphasised the interplay of socio-cultural variables within development (Malofeev, 1998) during which social semiotic mediation bridged the gulf between words and the world (Shotter, 1993; Wozniak, 1996), the linguistic version of tool as mediator being particularly important to Vygotsky (Leiman, 1992; Vygotsky, 1981b). Human consciousness was, for Vygotsky, situated within the social, cultural and historical environment (Wertsch, 1998), for him knowledge is social in origin even if that knowledge remains trapped within one person; it is because of "the social" that much knowledge is created (Brown & French, 1979; Miller, 1989). Sign within semiotics can take on slightly different meanings; for some it is a tool which represents that which is in reality is otherwise unrepresentable (a map of the Kruger Park) and that which is more readily representable (an arial photo of the Kruger Park) (Leiman, 1992; A. Leontiev, 1981). Vygotsky (1978) differentiates between tool and sign (symbolic mediation; Elkonin, 2001), with the former relating to external changes in objects and the latter relating to internal psychological changes effected by mediation. Play was also of particular importance to Vygotsky who stressed the symbolic aspect of it as mediatory tool in which culture is actualised and embodied within language and gestures (Kozulin, 1986; Nicolopoulou, 1993) and during which children could create their own zone of proximal development (Tzuriel, 2000b). The nature of the tie between language and cognition is a contested one with nativists positing a language centre in the brain progressively deployed throughout development and language as a product of a co-constructed process of learning. The relation between language and learning then becomes one of learning language, learning about language and learning through language (Grieshaber & Ashby, 1997). The results of play provide future impetus for the reorganisation of psychological functions in later life (Nicolopoulou, 1993). Vygotsky's method was christened "experimental-developmental" and was a process-based dynamic analysis of child development (Vygotsky, 1978). He fully appreciated the melding of two manners of locating psychology within the broader field of knowledge discovery and acquisition, thus later prompting his concern with the crisis in psychology. The syndrome of disability is defined as a social problem encompassing the individual but not leaving it there (Grigorenko, 1998). Russian defectology is mainly characterised by the following (Grigorenko, 1998):

- Social remediation and rehabilitation in contrast with the "listing" function within mainstream assessment where individuals are screened and systematised
- The replacement of the loss of certain functions by higher mental functions. However, those individuals evidencing less severe impairment are often not accommodated in this approach
- The subsequent individualisation of specific learner problems but an equal disregard for those mentioned above
- Regardless of the specific disability, mediation within the ZPD is fundamental to rehabilitation
- Mental development and disability is conditioned more by the social sphere than genetic heritability

The Soviet 1920's was a decade characterised by intense interest and strides within the domain of child development, an area considered important due to the future socialist stability of the country and is noted by the contributions of Vygotsky, Kornilov, Krakov and Smirnov among others (Kostyuk, 1972). Child education moved from a concern with "learning through doing" to a concern with "creating a person fulfilling the explicitly denied demands of the social and political order" (Bauer, 1959, p.43). The notion of person-in-society akin to Vygotsky's "mind in society" or "mind in context"¹⁰⁶ or "context and cognition" (Lidz, 1991) and "union of the person in situation" (Snow, 1998a) is very evident here and the stress that this notion plays within dynamic assessment is highlighted (Beals, 2000). Two main types of mediation are evident within Vygotsky's definition: metacognitive and cognitive mediation (Karpov & Haywood, 1998). The former deals with the self regulation, planning, checking and evaluation of behaviour and is encompassed within executive processing of the self (a feature heavily regarded within the Feuersteinian

¹⁰⁵ Developmental psychology for instance is an example of a research area in which parallel developments were taking place in geographically disparate areas of the globe with findings in child development being echoed in Piaget's research, Russian developmental research as well as mainstream North American child developmental psychology. These schools emphasised certain aspects in contrast to the other schools, but by and large, major characteristics of the typical child developmental paths were delineated in similar theories and models of development (Super, 2005).

¹⁰⁶ Gillen (2000) points out that Vygotsky never wrote a book entitled "mind in society" and probably did not consider his ZPD notion as terribly important. The fact that much literature contradicts this notion has resulted in the author leaving in this one sentence. One wonders what Gillen (2000) might make of sweeping statements such as Vygotsky being "the most ingenious, prolific and encyclopaedic humanists of our times" (Gindis, 1995b, p.99).

model of instrumental enrichment¹⁰⁷). The internalisation of socially mediated experiences was brought and cemented into Soviet psychology by Vygotsky and is considered one of the most consistent attributes of dynamic assessment models¹⁰⁸ (Karpov & Gindis, 2000). Vygotsky melded the notion of a plastic brain which could be shaped by the social environment (Van der Veer & Valsiner, 1991) so it would be incorrect to over-simplify his idea of a social aspect only in his theorising. Polarising concepts which are not in fact opposed is a key feature within the reconstruction of science and history and is something to be guarded against (more of this in chapter 3). Secondly, cognitive mediation aids in specific skill acquisition that is needed to perform school tasks for instance. Galperin paid particular attention to cognitive processes within learning and was later to develop a research agenda concerning the teaching and learning process and initiated his research in the 1950's which is parallel in time to the original research conducted by Feuerstein (Haenen, 2001). Haywood (2003) goes one step further in his conceptualisation of neo-Piagetian mediation and includes self-mediation as a further goal to the independent development of the individual where the mediator effectively fades away to be replaced by the adequate and sole functioning of the individual (Miller, 2003b).

Grigorenko (2004b) further stipulates four sub-divisions within Russian psychological research, namely,

- *1917-1936*: the period preceding the 1936 State decree of Pedology
- *1936-1950*: the period framed by the cessation of Pavlov's work. It was during this period that many psychologists were accused of cosmopolitanism (working too closely within Western research traditions) and moving away from Marxism
- *1950-1980's*: during which the reign of terror having subsided allowed for cross-collaborations within Western psychologists yet still maintained a semblance of Marxism
- *the current period*: which emphasises tolerance and is in no way impeded by any ideological barriers
- note that all these contextual influences resulted in intelligence research traditions substantially different from mainstream Western models

One can see from what has transpired above how powerful learning and learning disability research was in the Russian psychology programme and how it came to be that later Western thoughts on early Soviet psychology was influential in the development of various strands of dynamic assessment within both the socio-cultural and psycho-educational realm (Lidz & Gindis, 2003). Most cognitive education programmes have as their genesis classroom experiences (Haywood, 2001b) and it is clear where the field of application lies for this manner of assessment. Dynamic assessment can rightly be referred to as a psychoeducational model of process-based assessment (Haywood, Tzuril & Vaught, 1992). This is in keeping with Feuerstein's three-pronged approach towards child assessment: the scientific theory of child development, the educational aspect of child pedagogy and the ethical understanding of meaning and action within education (Hadji, 2000). Acknowledgement of theoretical educational psychology and school psychology (theory vs. practice) is evidenced in Feuerstein's bridging of the two disparate fields (some may view them as disparate) (Burden, 2000) and his thinking in terms of modifying general intellectual skills as opposed to domain-specific skills which can be traced in this line of thought. Anecdotal commentary has often illustrated the lack of fit between academic studies and the reality in which such skills are supposedly necessitated (Sternberg, 2000b),¹⁰⁹ so in a way, the enhancement of general functioning can be viewed with more confidence. Applying theoretically inspired models in

¹⁰⁷ It has been noted by scholars that Feuerstein does not adequately cite Vygotsky as intellectual predecessor within mediation research as some ideas are clearly very similar to those of Vygotsky (Miller, 2003b) especially with statements such "what are the chances that this child can go beyond himself" (Goldberg, 1991, p.37) harkening back to the level of next development (ZPD). There are however departures from Feuerstein's enterprise and Vygotsky's conceptualisation; acquisition of mediated concepts and cognitive modifiability are different aspects within the respective theories for instance (Moll, 1986). However, one must be mindful of the fact that early Feuersteinian work was conducted within this mode during the 1950's, a period in which most of Vygotsky's work was only available in Russian. It was only to emerge much later (in the 1970's and onward) in English translations (Wertsch & Tulviste, 1992). Nevertheless, there seems to be a general consensus indicating Feuerstein's alignment with Piagetian developmental learning and thinking and Vygotskian appreciation of the importance of social interaction (Falik, 2000). Many research efforts during the 1960's and 1970's across Europe, Israel and the United States had invoked similar alternative concepts towards the study of intelligence without reference to Vygotsky in any event (Guthke, 1992). Perhaps this is an instance of an idea whose time has come.

¹⁰⁸ Vygotsky's work can be considered as work in progress, thoughts on a model but not a fully developed theory. Hopefully though it is not a world-view masquerading as theory (Hunt, 1997).

¹⁰⁹ At present the author tends to agree with a Feuersteinian predilection for enhancing general cognitive skills as it is expected that a domino effect will commence during which skills learned and modified during intervention programmes could initiate a cascading effect or spill-over into other areas. However, for many practising school and educational psychologists reality necessitates a more subject-specific intervention programme. The author has at times wondered about her application of subject-specific skills in the real world and is pleased to announce that although there are numerous aspects of the curriculum that have never been utilised (and probably will never be) the underlying cognitive changes undergone have proved invaluable. Related to the aspect of practising school psychologists and the at times far from perfect surroundings in which they work, models such as mediated learning experience can and have been systematised into a form which is more practical in circumstances demanding more assessments within shorter time span, notable the Mediated Learning Experience Rating Scale (Lidz, 2000a).

practical day-to-day running of intervention programmes within schools is very much in keeping with the spirit of Feuersteinian approaches.

Metacognitive mediation involves parental guidance of spontaneous (natural, everyday, empirical or unsystematic) concepts whereas “scientific” conceptual knowledge (theoretical) is gathered at school via peer and teacher mediation, scaffolding or assisted performance (Karpov & Bransford, 1995; Lidz & Gindis, 2003; Portes & Vadeboncoeur, 2003; Van Geert, 2000) where verbal definitions start to take charge of thought (Brockmeier, 1996). Such scientific concepts need to be taught directly as many instances of what humankind now takes for granted (understanding gravity for instance) is not immediately apparent. Although the authors add that learning scientific concepts (Vygotsky’s terminology) should be a constructed experience (Karpov & Haywood, 1998). In the 1920-1940’s unrealistic demands were often made by the state to push delayed children into mainstream education, often resulting in innovative methods of attacking the problem. A major aspect to come from this state of affairs was the need for remedial education and the accompanying need to differentiate those children who merely required teaching in another mode versus those who truly required specialised interventions. Unlike Western psychology which is a testimony to fragmentation (which is not necessarily a negative feature as much good research has come of it), Soviet psychology progressed along smoother lines in terms of trajectories and traditions although there is also no such notion as “Soviet psychology” just as there is no such thing as “American psychology” (Hydén, 1988; Valsiner, 1996). The era during which Vygotsky lived witnessed relatively few traditions within Russia which was of course to change later on (Van der Veer, 2000). This trend was aided by the effective ban for 25 years on the practise of psychology as a discipline as well as a ban on sociological surveys (Malofeev, 1998). School psychologists were effectually eliminated in 1936 and were only reinstated more than 40 years later (Grigorenko, 1998) due to the prevailing spirit of equality of all and the subsequent placement of all national groups under one Soviet Union (Tzuriel & Haywood, 1992). Concerning the banning of tests in 1936, Jarovsky (1989) highlights the differing interpretations of intelligence as viewed in Soviet society and American society during the first half of the twentieth century. Americans, he states, were concerned with the nature-nurture debate due to their “dream of purely achieved status” and being anxious to improve upon techniques of further justifying inequality whereas Soviets aligned themselves with the idea of upward mobility in terms of intellectual recognition. People in the lower strata of societal functioning were not placed there intentionally and “great talents were imprisoned within the uncivilised” (p.346). Nevertheless, three main Soviet traditions can be identified:

- Pavlovian psychology (akin to Skinnerian behaviourism) - in which the beginning of the subject domain of higher nervous activity was presented in 1903 at the International Medical Congress in Madrid (Brožek, 1972b; Ushakova, 1997). This research tradition is of particular importance regarding the cross-pollination of ideas *from* the Soviet Union *to* the West and Pavlov’s influence is evident in the works of both Russian and Western researchers such as Bekhterev, Anokhin, Asratyan, Thorndike, Yerkes and Watson (Kozulin, 1986). Although the influence of Pavlov is well known, it must be noted that the development of behaviourism and “objective psychology” did occur more or less simultaneously, Pavlov’s work was pre-revolutionary (Bauer, 1959) and was influenced by (some would claim was based upon) the work of others including Sechenov who is considered the father of Russian physiology (Gilgen & Gilgen, 1996; Sahakian, 1975). Soviet psychological science’s past is an historical inventory of physiology as well as mental studies and the two have often become so melded that the one (mental studies) seems to have been subsumed within the physiological counterpart (Zinchenko & Gordon, 1981). In keeping with general philosophy of science issues then nascent within Western psychology, reductionist understandings of human functioning must have been a welcome addition to the repertoire within research, although this research tradition has, within the Soviet Union, been considered the domain of physiology as opposed to psychology (Brožek, 1966, 1972a; O’Connor, 1961). It is difficult at times to draw a distinction between various research domains under the blanket term of physiology or psychology (Bauer, 1959). Pavlov acknowledged the inner life of subjective consciousness but understood that objective research within this domain would prove problematic (Ushakova, 1997). Eminent Soviet psychologists both endorsed and applied Pavlov’s research tradition. American behaviourism generated much output during the 1930-1950’s and was heavily influenced by Russian reflexology (Brennan, 1982) more so by Bekhterev and Sechenov than Pavlov in fact (McLeish, 1975; Valsiner, 1988), the latter’s work and influence often being overemphasised. Vygotsky deliberates upon the consideration of thought as pure reflex and discusses the role of mediation as an observable means to reach unobservable thoughts (Vygotsky, 1994b).¹¹⁰ Biologically conditioned responses as well as historically derived (cultural) responses are part of the developing child’s repertoire of skill and skill acquisition (Vygotsky, 1994e) but the main focus within his pedagogical work was not with the laws of heredity as such as much as the role that such

¹¹⁰ Although the author does not read Russian and has had to rely on translations of Vygotsky’s works from various translators, it is evident that his style is quite lyrical and flowing which, from an aesthetic point of view, is refreshing but can be a bit tedious when trying to get to the crux of what he is trying to say. One finds that when reading a passage on subject x that he seems to take quite a while in getting to the point. Cole and Scribner (1978) point out that his style is general and almost no raw data is ever presented. But in all fairness, Vygotsky was a pioneer on a number of fronts and it would be unwise to be overly critical of his work from the stand point of a century in the future. Nevertheless it is a unique style and at times so lucid in thought in terms of the fusion of biological and environmental concerns that the author feels almost as if Vygotsky would have been quite comfortable bridging the psychological with the physiological as has been alluded to already.



heredity plays within development (Vygotsky, 1994f). This rich area of research seems to have dominated branches of related research into perception, memory and brain plasticity. However, just as there are gaps evident in the theory and models of Piaget so too are there gaps in Pavlov's work (but there would be, it is over a century since his Madrid presentation)

- Abstract mental operations analysis research best exemplified by Rubinstein (akin to structuralism and Gestalt psychology) who sought to turn psychological experimentation in the Vygotskian sense into a more educational one in keeping with Marxist philosophies of active change (through change one is able to come to greater understanding of objects in the world and the way to obtain knowledge about objects in the world is to set about attempting to change them) (Guthke & Wingenfeld, 1992; Lidz, 1991). Ironically, Rubinstein's work was considered too bourgeois in the 1940's which resulted in his revising his work and turning towards the Pavlovian tradition of ideas (Hydén, 1988). There is continuous flux between genetic and environmental heritage, a process to which the developing child has to grow accustomed. There is also the social which services the internal via mediatory agents such as adults which guide the growth of children. A departure from Western thought on the matter of age-dependent maturation occurred with the likes of Rubinstein's questioning of the nature of this growth which is dependent not only on age but on the information accumulated by children as well as the nature of their varied activities (Kostyuk, 1972). Vygotsky himself differed with Piaget on the matter of pure age-related growth stating that development and instruction were commensurate and integrated in such a way that instruction and development co-occur. Hence his theoretical offering by way of the ZPD to this question (Allal & Ducrey, 2000; Karpov & Bransford, 1995; McLeish, 1975; Van der Veer & Valsiner, 1991) added to the notion that development does not necessarily lead to learning (Feldman & Fowler, 1997) in a straightforward information processing fashion (Gindis, 1995b). For instance, no matter how mathematically gifted a child is, no manner of improvement or regard for mathematical concepts will be forthcoming unless there is adequate mediation on the part of some social force (Vygotsky, 1994f). It is necessary to differentiate between organismic developmental stability across cultures due to common genetic ancestry and variant changes brought about by learning in different cultures (or set-ups and schooling environments). Learning and development co-occur and this means that variant changes in development are due to learning environments (Niaz & Caraucan, 1998) an aspect underappreciated by Piaget but closely seen to by Vygotsky and also by dynamic assessment's concern with the learning-to-learn approach. In this regard, due deliberation is given to both developmental and learning issues within modern theories of cognition (Niaz, 1998). The training of skills via instruction has become prominent in the literature since the early 1980's (Borkowski & Konarski, 1981). Vygotsky (1978) referred to this method of development observation as instrumental and noted how erroneous it was to insist on the separation of the two. For instance, to study the child regardless of the environment (reminiscent of Binet) is characteristic of a natural endowment approach versus the study of achievement due to the environment only, regardless of the natural processes at work is to miss the vital link of the interaction between natural ability and environmental concern (Vygotsky, 1981a). "Child development is least of all like a smooth process sheltered from external influences" (Vygotsky, 1981b, p.151) and the case is vividly illustrated in his testimony of the deaf child's development which may be delayed only due to the lack of an appropriate environment and that this delay is far from being one of genetic causality (Vygotsky, 1981c) and should be dealt with as a social problem. He did not of course deny the problem's biological nature¹¹¹ (Vygotsky, 1994a). He was keenly appreciative of Darwinian theory¹¹² (as was the whole enterprise of the study of child development as geared towards a materialist-evolutionary concern for the field; Rahmani, 1966). He sought two lines of human development as explanatory modes of human social existence; the first detailing biological evolutionary adaptation and environmental press bringing about inclusive fitness and the social or historical development, upon which he seemed to place more emphasis and its regulation of the further development of "socialist man" (Vygotsky, 1978, 1994d). The importance allotted to instruction and development as co-occurring aspects within his theory is echoed in current dynamic assessment efforts as well as Feuersteinian theory which is principally instruction coupled with assessment (Messerer, Hunt, Meyers & Lerner, 1984)
- Vygotskian cognitive developmental psychology (akin to the later Piagetian and Feuersteinian developmental and contextual approaches). It is noteworthy to consider Van der Veer and Valsiner's (1994) three reminders about Vygotsky's work which they consider to be blind spots;
 - although the Soviet Union did experience isolation, both academic and social particularly during the reign of Stalin, Vygotsky's thoughts can nevertheless be considered interdependent with European and American academic thoughts and this can be seen in some of his work (Porges, 1998; Vygotsky, 1978)
 - Vygotsky, although emphasising the socio-cultural context, did in fact place a great deal of emphasis on the individual developing person

¹¹¹ This makes for the interesting slotting in of dynamic assessment models and theories within intelligence assessment where intelligence is itself defined from distinct vantage points; from neural mechanisms to social processes (Lohman, 2005; Pretz & Sternberg, 2005).

¹¹² No longer considered a theory but a fact yet at the time of Vygotsky and his contemporaries, public opinion was divided (unfortunately a similar divide seems to forever seethe under the surface when it comes to public opinion).

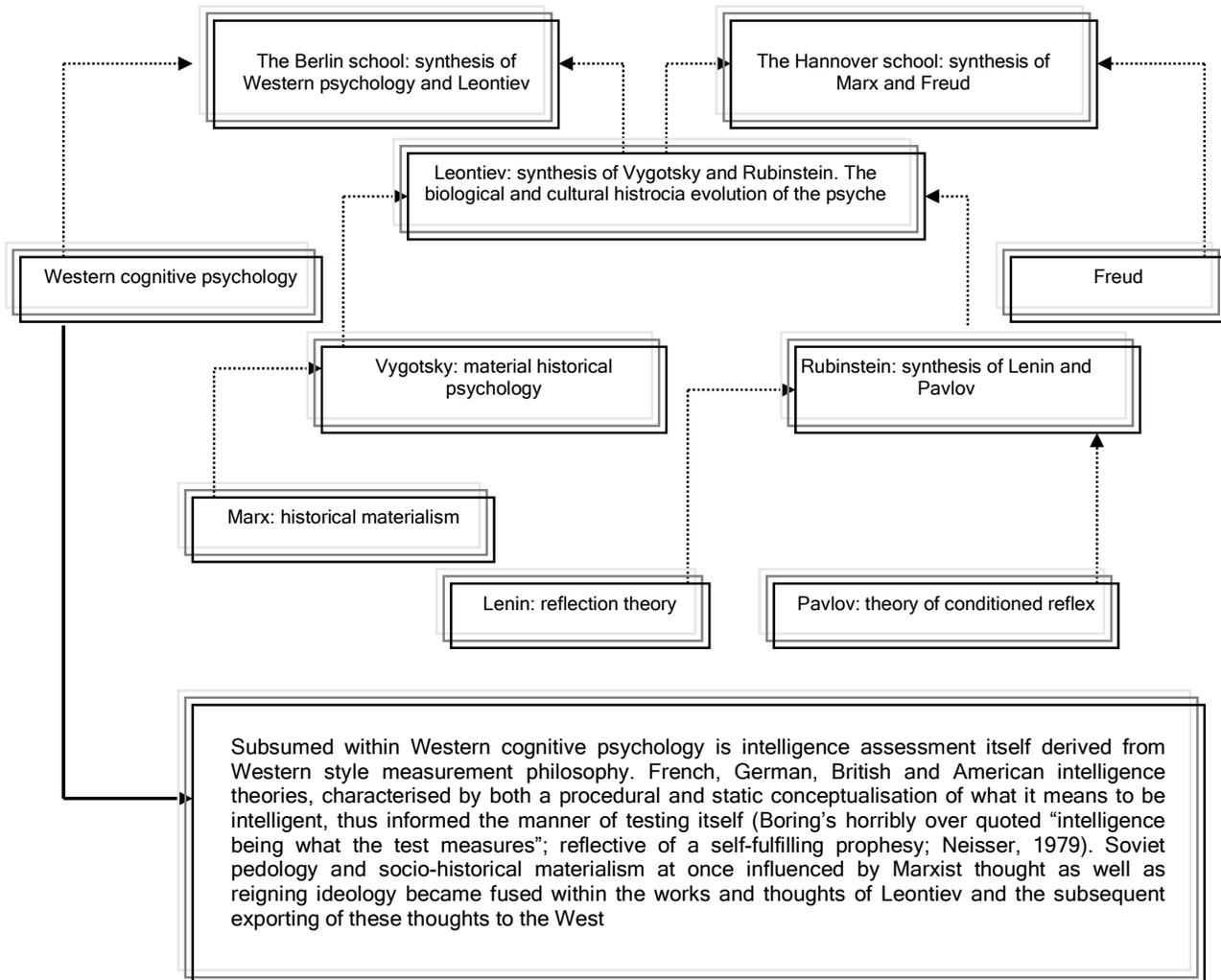


- o Vygotsky's emphasis lay more heavily on the cultural aspect of social mediation (cultural experience, cultural behaviour and cultural methods of reasoning; Vygotsky, 1994c), more so than on individual peer and teacher mediation. "Culture, generally speaking, does not produce anything new apart from that which is given by nature. But it transforms nature to suit the ends of man" (p.59). Also, peer, parent and teacher mediation has often been cited in positive terms where functioning within the zone of proximal development is considered advantageous, but is this the case in all instances?

Marxist philosophy on the theory of society is an underlying and prevalent feature of these three approaches but widely disparate in their efforts as separate traditions (Coles, 1982; Shotter, 1989). The case of ideology has at times been critiqued for being overstated and oversimplified in some literature (Bauer, 1959; Doehring, 1982; Valsiner, 1988). Such oversimplified renderings of Marxist thought can, at times, be misconstrued as a mechanistic treatise for socialist life which it in fact is very much not (Joravsky, 1989). Yet reading the literature generally leaves one with the impression that many aspects of Soviet academic thought including Vygotsky's ideas on consciousness and cognition were traced along lines of a division of labour (Emihovich & Lima, 1995; Moll & Slonimsky, 1989).¹¹³ As O'Connor (1966) points out, much of Soviet psychology has been oriented by dialectical materialism; a period lasting from 1936-1950 (Sahakian, 1975). One rarely reviews Russian/Soviet psychological literature without noting Marx's contributions in some form or another (Avtonomova, 1995). Figure 8 illustrates the linkages between various thoughts maintained by philosophers and psychologists within Soviet psychology.

¹¹³ As Emihovich and Lima (1995) point out, Vygotsky differed from traditional Marxist renderings of labour where the focus was on the mastering of nature via tools as opposed to the mastering of cognitive processes via psychological tools.

Figure 8 Linkages between schools of thought with emphasis on Soviet systems (Madsen, 1988, p.444)



That this background somewhat intrudes on pure and objective research is often viewed with scepticism by Western intellectual counterparts (Gray, 1966). However, such political dominance and resulting influences cannot be understated such as after the 1917 revolution where anything which smacked of bourgeoisie tendencies was literally obliterated before it could raise its head. There is of course the understanding that among many Soviet psychologists, theories and implementation of ideas were voluntarily moulded by Marxist ideology with many advocating its necessity, so it would be incorrect to state that such ideology was pervasively forced upon academics (Cole & Cole, 1979). Joravsky's (1989) definition of ideology as pertains to Soviet psychology is decisive in its depiction of the state of affairs "[an] unacknowledged dogma that serves a social function of unverified belief assumed to be proven truth because they serve the interests of the group that shares them" (p.ix). During the latter half of the nineteenth century, some psychological articles were infused with a mixture of science, ideology and politics (Grigorenko, 2004b; Grigorenko & Kornilova, 1997; Todes, 1984) and the intertwined nature of science and society was never more marked than after the 1917 revolution in which psychological studies focused on the individual in socialist society (Valsiner, 1996).¹¹⁴ It has even been extolled that Soviet psychology originated within the realm of political action and not in the laboratory (McLeish, 1975) but the role played by reigning ideology cannot be discounted when considering Soviet psychology

¹¹⁴ A lengthy citation from Todes (1984, p.544) construes this situation perfectly: "The records of the St. Petersburg Censorship Committee testify to the basic antagonism between developments in biological psychology and Tsarist ideology. Perhaps more important they indicate that because the Tsarist state was unwilling to halt development of science and medicine it was compelled to acquiesce to the gradual erosion of a critical feature of official ideology. Several factors, particularly the rise of professionalism and positivism, limited the rate of that erosion, and biological psychology could not, of course, disprove the concept of an immaterial 'spiritual aspect of man'. Yet ideas can be defeated by simply being ignored, and the increasing legitimacy of biological psychology in Russia did provide alternatives to the traditional explanation of human psychology. In this sense, the fears of the Tsarist censor proved well-founded, and biological psychology displaced a key element of Tsarist ideology within an ever-widening sector of Russia's intelligentsia".



(Oleinik, 1996).¹¹⁵ The rallying call of “all like one, one like all” is evidence of the nature of the times in which retarded children’s education was ignored (Malofeev, 1998) and the task of creating a homogenous society without class distinction was very much part of the socialist nation ideal (Kozulin, 1987).

One cannot unequivocally state that Western psychological traditions are typically individualistic even though there is an identifiable trend witnessed in these various traditions. It has always struck the author as odd that although Vygotsky adhered to Marxist philosophies (he was a leading Marxist theoretician; Luria, 1979) and sought to reform psychology by integrating Marxist thought within his approach to child assessment, his novel approach was effectively banned in the Soviet Union (Cole & Scribner, 1978; Kozulin & Presseisen, 1995; Yaroshevsky, 1996). “To look at Vygotsky’s book *Pedagogical Psychology*, one had to have a special pass from the KGB that would admit one to the restricted reading room in the Lenin library where the book could be read” (Davydov, 1993 in Kerr, 1997, p.4). This should give one an idea about the period and circumstances during which Soviet psychology was being developed.¹¹⁶ During the 1930’s both Luria and Vygotsky’s work was a blended version of Soviet and Western psychological ideas, seeking to utilise what both traditions could viably offer (Joravsky, 1989). Luria’s pioneering ethnographic work in Soviet Central Asia preceded work conducted within anthropological studies in the United States (Hunt, 1994). McLeish (1975) offers two reasons as to his unpopularity by stating the lack of sufficient Marxist quotes in his texts and his affiliations with pedology (the general all-round study of child development; Sutton, 1988) and hence testing. There is no such thing as a Marxist psychology even though it is in some instances referred to as such (Madsen, 1988), for Marx never developed such a system but did remark on questions of psychology (Rubinštejn, 1987). There was a fervent attempt at creation of Marxist dialectical materialist psychology (Hydén, 1988) and science in general (Tobach, 1996). Ageyev (2003) in one of the more user-friendly articles on Vygotsky does an admirable job of highlighting various difficult issues that Western students face when reading Vygotsky. Marxism and its related philosophy is not the mainstay of Western education and the connection with Vygotsky is often strange to some. Why were Soviets so immersed in this philosophy? Ironically, states Ageyev (2003), Vygotsky is often decontextualised which goes against the grain of his cultural-historical theory! Much of Soviet history is indeed perplexing especially when seen in the light of favourable and unfavourable art, music, dance and theatre and cultural life in general where state approval was often the precursor to success. Trends and governing regulations were often fickle with many Soviet artists (musicians and dancers for instance whose works were considered atonal, naturalistic or cacophonous were discouraged; McLeish, 1975) and scientists becoming blacklisted almost overnight for some supposed transgression which supposedly reflected the State in a bad light. Goldberg (1982) casually refers to this state of affairs as not atypical and to be accepted as part of Soviet history. Brožek and Slobin (1972) claim that Vygotsky was the first Russian psychologist to depict, in entirely psychological terms, a Marxist account of the socio-historical nature of human consciousness and adhered to a dialectical materialist account of psychology which can briefly be described as:

- Subscribing to materialist monism (mind is brain, see above section on consciousness)
- Determinism (the interests in biogenetic studies of development was a characteristic of 1920-1930’s Soviet psychological science)
- Reflection (consciousness is a reflection of external reality)
- The unity of consciousness and activity (consciousness as formed by the activity in which one engages)
- Historicism (the development of consciousness out of human history)
- The unity of theory and practice (theory is only as good as its application in practical contexts usually in educational, child, work psychology as well as psychopathology; Teplov, 1961)

¹¹⁵ Of course very much the same can be said of Western psychology which is similarly a product of capitalism and individualism (Eskola & Weckroth, 1996). One need only think of the appeal of pop psychology where you are made aware of a problem you do not have only to be offered the solution for half the price of the competitor!

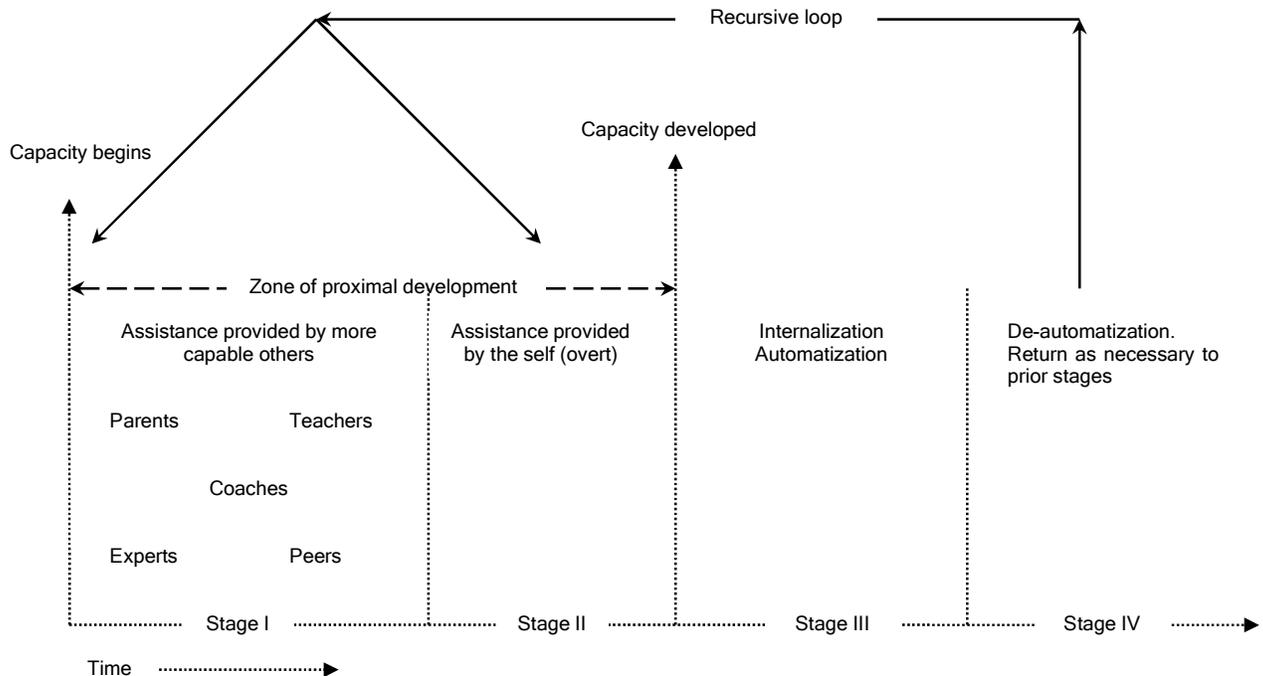
¹¹⁶ Vygotsky’s ideas and theories were developed during particularly hazardous times with major events during 1917 - 1928 being characterised by civil war, “desperately poor conditions, famine, the forced exile of some intellectuals, ...a struggle to eliminate all aspects of bourgeois society...[as well as the rise of dialectical Marxism] that accompanied the assumption of power by Stalin” (Gilgen & Gilgen, 1996, p.9). The period from 1928 till his death in 1934 coincided with the Stalinist Era during which psychology (like many other fields of scientific inquiry) were systematically “purged” of unwanted and undesired suspected bourgeois influences. Science in early twentieth century Russia witnessed “explosive institutional growth [which] was accompanied by the abolition of entire disciplines, and outstanding achievements routinely co-existed with backward doctrines” (Krementsov, 1997, p.777). Vygotsky’s ideas were not immune to these intellectual and brutal purges either and did not exist in a socio-political vacuum. Psychologists, who during the 1920’s, did not affiliate themselves with the ideals of Marxism as espoused by Stalin, were simply not allowed the freedom of continuing their research (Kozulin, 1990) and those who did not follow Marxist dictates were either exiled or subdued (Grigorenko, 2004b). Furthermore, Kozulin states that Vygotsky (among others) did not escape criticism and that his “cultural-historical theory was branded as leading to an anti-materialistic revision of psychology and [was] considered to be a reflection of bourgeois inclinations” (p.240). Without going into dire detail, Vygotsky’s plight is seldom the focus point in the many articles and books that discuss his work, but is thought of here as primary importance in the development of dynamic assessment as it pertains to his particular bent on semiotics and socio-cultural aspects of child development. Such socio-cultural and political influences cannot be ignored when considering a study such as this proposed one. Another point to consider is the effect of prevailing models of assessment ca. 1920 and to what extent Vygotsky and others were influenced by them (if at all, as Russia was quite isolationist at this time in terms of scientific progress and subsequent publication of scientific results in general).

One can clearly follow a thread in Vygotsky's thoughts regarding the utility of dialectical materialism. The focus on external impinging social reality in the formation of consciousness, the notable influence of historical processes in the development of mind (brain) and his application within pedagogical or pedagogical contexts. Interestingly, the above accords with similar notions as advocated within a positivist framework which has indeed been remarked on (Bickley, 1977). Such positivist interpretations were encouraged during Stalin's reign but rigid and unwavering adherence to strict positivism where objective facts were assimilated without being processed actively by the individual was anathema to Vygotsky (Van der Veer, 1996). One needs to tread carefully over the terrain of reigning popular conceptions of Soviet psychology and the movements within and around it by individual researchers. Bauer (1959) distinguishes two types of Marxism, which can be read with the above in mind, namely, a "vulgar Marxism" which affirms a popular rendering of the common understanding of Marxism which concerns itself with materialism, mechanist and deterministic understandings of behaviour and the second understanding of Marxism which is a more reflective one in the understanding of consciousness and the processes involved in active perception all the while acknowledging a separate yet extant reality. The conflict between the two conceptions played out in areas other than Soviet academia. The see-saw notion one gets from reading the history of Soviet psychology is likewise attested to in the works of various Soviet psychology scholars. Certain theorists and researchers work was considered acceptable only to be termed unacceptable later on and is indicative perhaps of the complicated role played by ideology and how academics and others had to manoeuvre within the system. One of the key features in common among the group of researchers led by Vygotsky, was the emphasis on practical applicability and the development of a psychology in practice and is exemplified in the works of, among others, Vygotsky's own research with handicapped children, Luria's studies of twins, Leontiev's study of school children's concept formation development, Zaporozhets' work into child mental development, Galperin's studies of the manner in which tools were mastered by children and Zinchenko's work in memory (Zinchenko, 1982).

Vygotsky's ZPD stands in stark contrast to the *g*-centred research in mainstream assessment and the almost natural urge among mainstream assessors is to try and quantify a concept which seems to lie beyond such an attempt (Ageyev, 2003). However this is precisely the problem within current dynamic assessment: the need to quantify and model change. The author often wonders whether Vygotsky would not turn in his grave if he were to witness such flagrant disregard for a process wholly entrenched in qualitative assessment.¹¹⁷ However, Vygotsky himself did not conduct any experimental validation of the ZPD in his own short life-time leaving it open in a sense to future adaptation and experimentation (Sternberg & Grigorenko, 2002). Perhaps the need to quantify the ZPD is another symptom of mainstream assessment. Vygotsky's ZPD was only "recently" researched for the first time in American literature in 1964 (Das & Conway, 1992) and was brought to the fore in this country in particular by Budoff and his team who were among the first to apply Luria's modified version of Vygotsky's ZPD (Hamers, Hessels & Pennings, 1996). The beginnings of *g*-related research had ignited following Spearman's seminal paper in 1904 so Vygotsky would most likely have known about it. ZPD is fundamentally a language-centred socio-cultural theory of development and the main proponent of such theory is A.N. Leontiev, a Soviet theoretician who argues that the individual changes qualitatively through life, being influenced by social and economic factors (Alfassi, 2002; Coles, 1982) and is not set by biological heredity (Kostyuk, 1872). Vygotsky's ZPD is illustrated in figure 9 below.

¹¹⁷ He could also possibly be overjoyed at the strides made within psychobiology. Who knows?

Figure 9 Tharp and Gallimore's (1988, p.35) depiction of Vygotsky's zone of proximal development



He speaks, too, of original learning as being chiefly characterised by responses to environmental stimuli much in keeping with the Pavlovian tradition (Leontiev, 1961). Leontiev criticised Western tendencies for describing psychological reality as a registry of functions without context, as the best experimentation occurs outside laboratory settings (notwithstanding the obvious severe set-backs) (Shotter, 1989). This is in keeping with current disability research which promotes the understanding of disability within the real-life situation (Grigorenko, 1998). Along with Vygotsky and Luria, Leontiev contributed to what was later to become cultural-historical activity theory (also founded upon principles of dialectical materialism; Sheng, 1996) which is self explanatory in terms of its power as an implementable tool (Hydén, 1988; Lee, 2003). Although his theory can be viewed as an extension of Vygotsky's, there are elements within Leontiev's theory which are not present in Vygotsky's (Hydén, 1988) and his deliberations on the concept of activity was the more generally accepted one (Wertsch, 1981).

Later, followers of Vygotsky replaced the idea of semiotic mediation with that of activity theory where practical actions took precedence in theory and research (Robbins, 2001). Vygotsky, although never acquiring theoretical status in his theory, used the term activity largely to denote a process of being engaged in something or simply being active (Hydén, 1988). Mental processes are the direct result of human activity. The notion of labour is particularly evident here and hence its link to Marxist philosophy. Soviet psychology took pains to mitigate the effects of capitalist notions within its own developed "labour psychology" (Guthke & Wingefeld, 1992; Kotelova, 1972; Noskova, 1996). It is through such activity that learning occurs and is very similar to the Piagetian notion of learning by engaging with reality (Das & Conway, 1992). Activity or labour was so fundamental a concept during the early decades in Soviet psychology that even the realm of emotions was considered most amenable to study when viewed from a labour point of view (man at work becomes equated with his emotions) (McLeish, 1975). Marxist concern with equality of distribution and the resulting perplexity by members of more capitalist societies is perhaps one reason why such early dynamic assessment models were anathema to Western countries (Rand & Tannebaum, 2000).¹¹⁸ Vygotsky has thus been referred to as a meta-psychologist and for very good reason (Kozulin, 1986; Robbins, 2001, 2003). Of course one must take cognisance of the parallel developments within the West at this time where individuals were studied so as to better understand them not always necessarily to help them. Vygotsky's idea that thinking was dialogue transferred is superbly rendered within a social system where the individual becomes such through others (Shotter, 1993) and it can clearly be seen how such mediated learning was to become such a prime concern for Feuerstein later on. Individual consciousness becomes secondary in the transmission of social consciousness and his ideas are similar to those of G.H. Mead and are

¹¹⁸ There is a slight clash of ideology but surely can we not say that dynamic assessment has moved on since the early conceptions of its being polarised due to its "collectivist" leanings? Westerners no longer have as their excuse the misunderstandings of the basic philosophy behind dynamic assessment and likewise, non-Western countries cannot forever lament the West's preoccupation with individualism, after all, the world is progressively globalising.



traceable to the works of French psychologist Pierre Janet who in turn was influenced by Emile Durkheim and the French sociologists (Cole & Scribner, 1978; Kozulin, 1990; Nicolopoulou, 1993; Van der Veer & Valsiner, 1991; Vygotsky, 1981b; Wertsch & Tulviste, 1992). Janet's concern for the "other" within development seems to have influenced Vygotsky's thoughts in this regard (Kozulin, 1986) but Vygotsky was the first psychologist to introduce the aspect of culture into the study of the nature of being human (Cole & Scribner, 1978). The Gestalt movement and its appeal to holism found a receptive audience generally in Russia at the time and with Vygotsky specifically (Van der Veer, 1996), which is hardly surprising when one considers the manner of the Gestalt approach to perception as opposed to behaviourist approaches. Among other Western psychologies, Vygotsky also critically analysed the works of Piaget, personalism and behaviourism (Brožek & Slobin, 1972; Kozulin & Presseisen, 1995). However programmes such as Gestalt psychology and psychoanalysis¹¹⁹ was deemed bourgeois (as well as departing quite radically from "objective psychology" characterised by Pavlovian and Bekhterevian psychology; Wozniak, 1996) and so not appropriate for study, resulting in the later unfashionable Vygotskian school which was all but abandoned for many years (Guthke, 1982) until it became fashionable once more around 1956 after the death of Stalin (Kozulin, 1986; Toomela, 2000; Van der Veer & Valsiner, 1994).

This abandonment can perhaps best be understood with Vygotsky's programme being labelled anti-Marxist and bourgeois. However the further one delves into the reasoning behind this successive banning and resurgence of his work the more difficult it is to understand. Leontiev and his research group, for instance, aligned their theory more in keeping with the 1936 communist party decree stating that pedology (educational psychology) should be banned.¹²⁰ This was most likely to do with the assumption of equal intellectual endowment in children; a means of moving away from a class society so despised by Bolsheviks (Gilgen & Gilgen, 1996). It was maintained that it was on the labour of the masses that society prospered and not on the intellect of the intelligentsia but there were also other reasons why Stalin has chosen pedology as part of his purge policy.¹²¹

The social determination of thinking had as a result educational thoughts on development assuming socially mediated learning representing complete development (Grigorenko, 2004b) which of course did not manifest. Nevertheless, Gindis (1995a) maintains that an objective history of pedology has yet to be written. There was thus a turn away from previous pre-soviet physiology to a more inclusive social psychology (Bauer, 1959;¹²² Koltsova, 1996; Vasilev, 1996). In effect the decree meant the following (Abul'khanova-Slavskaya & Brushlinsky, 1996; Bauer, 1959):

- The school psychologist's activity was to be drastically reduced in scope and activity with the subsequent increase in status of the educator in matters psychological
- No more testing was to be allowed
- The expurgation of the study of personality
- Industrial psychology was similarly disbanded
- The concept of man was now more purposive and conscious
- Training was to receive more emphasis as opposed to the role played by the environment (the latter which Vygotsky and those before him emphasised)
- The task of psychologists was to churn out socialist citizens as opposed to laying emphasis on the biological being's interaction with the environment (one can see how Vygotsky's programme came under severe limitations)

This played out in the subsequent banning of such tests in East Germany at the time as well as the subsequent disappearance of intelligence tests in China following the goings-on in the Soviet Union and the later cultural revolution in China¹²³ (Guthke &

¹¹⁹ The Russian psychoanalytic society was in formation during 1923 and Stalin announced certain abandonments of psychology in 1929 (Santiago-Delefosse & Delefosse, 2002).

¹²⁰ Incidentally, the Tsar banned the teaching of philosophy in 1850 (Joravsky, 1989). One cannot deny the influence of ideology on current academic and psychological thought. This played out in Vygotsky's thoughts on child development to a point and he was, even though ahead of his time, still a product of it.

¹²¹ Stalin's son, Vasily, could not pass pedology tests apparently! (Gindis, 1995). So perhaps Stalin's reason was partially motivated by this.

¹²² Life under Stalin was progressive in one sense (the nationalisation of various aspects of the economy - the five year plans - and the distribution of resources to far more people than during the reign of the Tsars; Solso, 1996) but equally wrought with paranoia and suspicion. America's "reds under the beds" syndrome seems to have echoed Russia's early "comrade today, threat tomorrow" syndrome. What becomes of science and scientists during such hectic times? Unfortunately, they are most often the ones to "disappear". Vygotsky died of tuberculosis but one cannot but help wonder about conspiracies! After all, Bekhterev was poisoned by Stalin's assassins in 1927 (Medvedev, 1996) and Bekhterev considered Stalin paranoid (Joravsky, 1989) seemingly with good reason. In fact Sabina Spielrein's father and brother were both arrested for deleterious pedagogical research and her brother was subsequently executed as a Trotskyist (Santiago-Delefosse & Delefosse, 2002). Stalin chased Trotsky all over Russia giving him no rest until he finally assassinated. Likewise, psychologists who were deemed as encouraging anti-regime notions could be sentenced for Trotskyism! (Bauer, 1959). To boot, Trotsky was considered an advocate of bourgeois psychoanalysis, as well as an opponent of materialism and Marxism (Tugaybayeva, 1996).

¹²³ Large-scale group administered tests were also banned in Israel in 1985 as passed by the Ministry of Education, following what has now become known as the "anti-test syndrome"; protests surrounding the use of assessment tests for placements in schools (Zeidner, Matthews & Roberts, 2004). During the Second World War, the use of educational tests was minimised in Japan following on from a viewpoint which stressed that individuality as a concept should be eschewed (Sato, Namiki, Ando & Hatano, 2004). Problems surrounding tests and their administration are thus not limited to only a few countries at certain time periods but have been an omnipresent issue throughout the ages.



Beckmann, 2000a; Shi, 2004). This is perhaps a contributory reason why dynamic assessment or the learning-test concept took hold in Germany as strongly as it did (Kormann & Sporer, 1983). There was a strong current and flow of ideas between the former East Germany and Soviet Russia; ideas which were not necessarily transported to the West (Robbins, 2001; Woodward & Clark, 1996). The works of Western psychologists were originally translated into Russian but were increasingly filtered so as to guard against ideological contamination (Joravsky, 1989). Soviet psychology journals were likewise banned for 23 years and were officially declared acceptable only in the 1950's after Stalin's rule. Even the major journals ceased publication from 1932-1934 and publications remained sparse until 1946 when serial publications of the Academy of Pedagogical Sciences was inaugurated (Bauer, 1959). Leontiev's emphasis on socio-historical Marxist influenced approaches towards human behaviour moved away from Vygotsky's original concern with sign within culture. Although it is possible to identify Marxist feelings within Vygotsky's work, he criticised others' works for abusing Marxist writings and inappropriately attempting to integrate Marxism into their work (Robbins, 2003). In fact so tumultuous were the times that Luria most likely changed his research direction from psychoanalysis to clinical neuropsychology due to its distastefulness.¹²⁴ Luria's extensive work in neuropsychology has had much influence in the discipline today and intelligence models have been based on his research into brain malfunction; one need only think of the Das and Naglieri model of the assessment of attention, simultaneous-successive coding, and planning (Angus, 1985; Das, 1998; Das & Naglieri, 1992; Das, Parilla & Papadopoulos, 2000; Naglieri, 1997). Vygotsky's work has indeed seeped through to a very deep level in the West at a time when the Soviet union had all but banned Western scientific literature and severed ties with Western academic counterparts (Valsiner, 1988).¹²⁵ Vygotsky's work was continued by the Kharkov group (of which Leontiev became leader; Woodward & Clark, 1996 and was populated by Vygotsky's students and co-workers who had "disserted" him; Van der Veer & Valsiner, 1991) during the 1930-1940's in which internal (mental) events mirrored and corresponded to external activities (Galperin & Talyzina, 1961; Kozulin, 1986). Years later during the 1960's various aspects of the Kharkov group's work was considered as representative of Soviet development psychology including "perception as action" as pronounced by Zaporozhets as well as the concept of "step-by-step formation of intellectual actions" of Galperin (Kozulin, 1986). Galperin was a follower of Vygotsky's socio-cultural programme but developed its educational implications further and was involved in the foundation of the Kharkov school (Haenen, 2001; Wertsch, 2000). He deviated somewhat from Vygotsky on his emphasis on more reductionist approaches to the study of development criticising Vygotsky's all-encompassing and overarching sub-areas of concern (Van der Veer & Valsiner, 1991). He preferred to scale down as opposed to enlarging the subject domain. Passing away in 1988 he was one of the last people to have personally known Vygotsky although Vygotsky never joined the Kharkovites many of his students and followers did. Hence, emphasis within Russian dynamic assessment has traditionally been placed on two aspects; diagnosis of learning aptitude and Galperin's learning-teaching experiments (Lidz & Gindis, 2003).

"Destalinisation" or the depoliticising of psychological science resulted in the resuscitation of Vygotsky and the subsequent translations of his works into English and the criticism of Leontiev's work by Russian psychologists as well as the general restructuring of Soviet psychology (Bishop & Solso, 1996; Hydén, 1988). These 1960's translations culminated from the West's invigorated concern with cognitive aspects in psychology (after the so-called grand collapse or downfall of behaviourism; Green, 2001; Turner, 2001) which had under the behaviourist tradition been a non-issue (Valsiner, 1988). It was also during the 1960-1970's that intelligence made its come-back in Russian psychology when tests were performed once again (Grigorenko & Kornilova, 1997). Vygotsky's work often reveals itself as thoughts and models in search of grounded theory, which he did not have time to formulate due to his untimely death. Translation of Vygotsky's work has undergone a parallel "Americanisation" which resulted in the loss of his poetic and philosophical style (Kozulin, 2002 in Ageyev, 2003) and has also been "tainted" by misunderstandings of his work primarily due to cultural differences between Soviet and American life (Ageyev, 2003). Ecosystemic models of human functioning closely align with non-Western dynamic assessment precursors (Valsiner, 1988) and as Robinson-Zanartu and Aganza (2000) point out, it is easy to understand why a systems thinking approach in dynamic assessment was only a relatively recent phenomenon in the West. It is interesting to note here that perhaps Vygotsky's unadulterated readings are more in keeping with South African collectivism as evidenced by the nature and culture of the reigning "Ubuntu" concept which essentially bespeaks of the greater good of the group versus the greater good of the individual (individualism as opposed to collectivism or American vs. Russian or European vs. African; Eskola & Weckroth, 1996 and where the psychology of person-in-society and the behaviour that accompanies it differs from that in many Western countries; Filatova, 1996). Over and above this rather simplistic rendering of another culture, is what Ageyev (2003) refers to as high- and low-context communication cultures which dictate the nature of communication. Westerners tend to be blunt and straightforward vs. high-context communication cultures which do not engage as directly. What is not said or done is equally as important as that which is said and done. This applies strongly to South African higher education circumstances. Perhaps this is why dynamic assessment has witnessed subsequent greater reception in countries presenting with non-individualised emphases. This is,

¹²⁴ One gets the feeling that something similar was on the go during Nazi reign during which scientists relocated *en masse* to the West. The West has won over countless protégés due to totalitarian dictatorships and lucky for them that they did not languish in the gulags and concentration camps (Applebaum, 2003; Bishop & Solso, 1996; Freeze, 2000; Solzhenitskyn, 1974).

¹²⁵ Recall the now over-used example of Russia's isolationist policies resulting in among others the biological "science" of Lysenko! (McLeish, 1975).



however, changing as the world progresses to a more global and amorphous state. All cultures express in some form or another intelligent functioning even if it is not directly referred to as such and even within one single European framework competing alliances and histories evidence the varied paths followed by intelligence researchers. French research following after Binet, German research after Wundt, British work after Spearman and United States' following in the wake of, among others, Thurstone and Thorndike (Sternberg, 2004d).

Regarding learning disabilities, which in the West is attributed to internal functionings of the child, socio-cultural understandings emphasise external factors and stress that learning abilities are in fact learned. Neurological underpinnings are not, however, swept aside for one need only look towards the work of Luria on brain functioning and injury and who cited Vygotsky as his mentor. Socio-cultural and neurological theory become intertwined in the fuller understanding of the individual. Dialectical materialism as basis for much Soviet work is itself descriptive of research which poses contradictions in the hope of finding "true" information (Coles, 1982). The continuous interplay between polar opposites (or at least which seem to be opposed) brings about development. One can express it as the individual within society, the neurological within the larger system and learning abilities within socio-historical contexts. Such "activity" throughout life is not the equivalent of stimulus-response behaviour as understood by mainstream psychology as the mental intervenes and mediates responses. Pavlovian psychology was likewise criticised for this approach to human behaviour. Vygotsky transcended S-R research by including mediation yet also did not veer off into introspectionist psychology where consciousness studies often sought explanations by referring back to that which they were attempting to explain; i.e. a tautologous system (Kozulin, 1986; Leiman, 1992). Activity is a loose term and has undergone changes in definition since its inception in the 1920's (Kozulin, 1986).

The appreciation for the greater context in which humans live their lives formed part of Soviet understandings of life and regarded Western experiments within psychology laboratories with scepticism (Coles, 1982). Echoing the dialectical relationship between contrasting aspects, Vygotsky followed on in the tradition of seeking to integrate both learning and the development process, denying that the two worked in isolation. His thoughts and views on human development were not clearly defined into ready-made categories such as strict biological psychology, or wholly hermeneutic and cultural but have been described as being "in between" these various approaches (Brennan, 1982; Shotter, 1989). He also utilised static conceptions such as mental age and the validity of standardised assessment as reliable measures of psychological performance but did offer trenchant critique (Lidz & Gindis, 2003). This thesis advocates dynamic assessment as method of assessment but cannot ignore much good research that has emanated from static conceptions of intelligence and perhaps one should consider why Vygotsky would have tacitly endorsed these views. It is unlikely in science and psychological science that any one approach is so all-encompassing as to describe fully the behaviour of individuals. Likewise his conception of mediation is also not a unitary description and ranges from mediation of activity to the acquisition of scientific concepts (Leiman, 1992). In essence, Vygotsky's theoretical contributions towards a more fully integrated dynamic assessment in 2006¹²⁶ can be listed as follows (Lidz & Gindis):

- Psychological tools are in need of mastering by the developing child and this takes place within varied contexts such as language, cognitive and social development. Such tools are not individual implements but reflect social symbolism and communication (Kozulin & Presseisen, 1995). Tools can of course also refer to physical objects, which are made and utilised by human beings in a way not used by higher primates and other animals (Luria, 1994). The former relying on 'native physics' and the latter on visual cues with no use of symbols in anyway (Vygotsky & Luria, 1994). At least one can say that the same cognitive processes used are dissimilar, where primates may use spatial cues for manipulation of the environment, humans make use of causality
- Learning and instruction are intimately intertwined yet not identical (Kostyuk, 1972) and seeking the separation of the two will not result in a contextualised and situated approach as cultural-historical theory dictates (Menchinskaya, 1972). Vygotsky's emphasis on the interdependence of instruction on development is evident in his discussion concerning the appropriation of academic concepts through schooling (Vygotsky, 1994g). Hence the need to continuously inform assessment from mediation and vice versa (one of the main reasons why dynamic assessment is so burdensome within today's time-hungry and cash-strapped society). It is still the case in more instances than not, that dynamic assessment although attractive for a variety of reasons is still very difficult to implement in practice (Elliott, 2000)
- ZPD's size can change throughout life and is not fixed. This is in keeping with modern-day understandings of the plastic brain. ZPD can shrink and grow depending on the nature of intervention as well as on the timing of such intervention. The length of intervention or training will also likely change depending on the nature of the task and the level of experience. In keeping with cognitive theory which states that the more one knows the easier it becomes to know, ZPD and the handling of it becomes compatible with the broader encompassing cognitive theory
- The larger the number of currently maturing functions within the developing child the more indicative of potential as opposed to those functions which have already matured

¹²⁶ Although not as fully integrated as it perhaps should be.



- Shared or joint activity (with more able peers and teachers) stimulates the growing functioning of the child
- In order to ascertain or pinpoint maturing functions, collaboration is necessary
- Functions which are not yet matured cannot easily be used in assessment so the next best aspect to look towards is the level of imitation which indicates the readiness of the child to engage socially with his function (which is most likely the reason why Vygotsky emphasised the importance of play as measure and developer of cognitive functioning towards higher mental functioning but has been criticised due his overemphasis on imitation). Such higher mental functioning was assessed across domains including memory, attention and decision making (John-Steiner & Soubberman, 1978; Kozulin, 1990)

Another giant in developmental psychology with whom Vygotsky is often compared and contrasted within the psychological development and educational literature is Piaget (DeVries, 2000; Duveen, 2000; Feldman & Fowler, 1997; Lloyd, 1995; Matusov & Hayes, 2000; Moll, 1989; Niaz, 2001; Santiago-Delefosse & Delefosse, 2002; Smith, Dockrell & Tomlinson, 2000; Tryphon & Vonèche, 1996). It can be postulated that Piaget approached the study of cognition from a structural position whereas Vygotsky did so from a functionalist point of view (Campbell, 1993; Moll, 1989). Piaget's work consists of many theories in contrast to Vygotsky's somewhat undeveloped scheme and much less experimentally validated theory (Smith, 1996). They are juxtaposed in terms of the respective theories' emphases on the individual within the social context and the social context as impinging on the individual as well as their respective views on accommodation and assimilation (Piaget) and ZPD (Vygotsky) (Van Geert, 2000). Vygotsky criticised certain aspects of Piagetian thinking (Santiago-Delefosse & Delefosse, 2002) yet read and commented on Piaget's work (Tryphon & Vonèche, 1996b). For instance, mediation does not feature as prominently in Piaget's works as they do in Vygotsky's (Haywood, 2003) even though it is appreciated that mediation of cognitive functioning does occur via a process within the developing child; the structural overlap between different stages require that certain functions become available (Case & Edelman, 1993). As Martí (1996) states "the individual, endogenous, operatory, universal constructivism, which accounts for the progress of the Piagetian subject, is opposed to the social, exogenous, semiotic, and contextual development inherent in the Vygotskian subject" (p.57). The two did converge on the aspect of interaction in which mind collaborates with environment (Wozniak, 1996) and their overall aspiration of humankind towards rationality, although they placed emphasis on differing aspects (Piaget's end-point was a universal human rationality whereas Vygotsky proffered rationality in terms of its functional utility within the state and economy; Wertsch, 1996). The degree of overlap between the two is testament to the utility of a combined summation of their approach and the application of both frameworks in work with developing children (Bidell, 1988; Smith, 1996). Although Piaget was never to personally meet with Vygotsky, he did express admiration for his work and lamented about not having read his work earlier (Guthke & Wingenfeld, 1992). Piagetian tasks were employed in Russia during childhood assessments and the Soviet interest in Piaget can be explained by the need for the "new Russia" to conceive of a similarly "new man" and secondly Piaget's approach was one of the few options viable to the new Soviet psychology in terms of not being expressed nor couched within a bourgeois philosophy (Tryphon & Vonèche, 1996b). Evolution has been offered as an amalgamating framework within which to unify and view psychology; as the rationale is that all behaviour is a result of evolutionary adaptation and change. Piaget, did after all, enter the field from having studied biology to better understand the evolutionary practices at work within psychological descriptions of behaviour (Bjorklund, 1997; Li, 1996) and developmental psychology has obvious roots in evolutionary theory (Weinert, 1987). Viewing mathematical skill within both approaches evidences the differences between the two. Piaget emphasised the logical progression of understanding as the individual matures (innate preparedness) versus Vygotsky's emphasis on mathematics as a cultural tool which is mediated to children (Bryant, 2000; Resnick & Nelson-Le Gall, 2000). The dual roles and processes involved in development within the individual functioning within a broader context often results in the comparison of these two researchers' works. Yet a call for the synthesis of what can only be referred to as two great traditions (as many scholars have since made enormous strides within neo-Vygotskian and neo-Piagetian theory) makes more sense than a complete divorce and is encouraged generally but more so from an educational viewpoint where the learning process and development co-occur (thinking develops from teaching and teaching develops from thinking) (Shayer, 2000; Sylva, 2000).

2.7.2 Current trends

Awareness

Although more surveys may have been conducted, three surveys, conducted within the United States and United Kingdom evidence similar results in terms of the recognition of dynamic assessment as manner of assessment as well as the utilisation of this method within practice (Deutsch & Reynolds, 2000; Haney & Evans, 1999; Lidz, 1992a). Lidz's 1992a study surveyed 120 school psychologist trainers in the United States who co-ordinated graduate cognitive assessment courses. Due to the length of time that had passed since formal implementation of dynamic assessment models in the 1960's and 1970's Lidz (1992a) was of the opinion that information pertaining to these models should have seeped through to practitioners as well as academics in some form or another. Diagnoses form the mainstay of a school psychologist's agenda making this population a suitable target for answering questions concerning dynamic assessment. In essence Lidz's (1992a) survey encompassed questions concerning the familiarity with the model; the extent to which it is utilised; the manner of first becoming aware of the model; the extent of incorporation or reference to dynamic assessment within cognitive assessment courses; possible reasons as to why models



were not included in courses if knowledge did exist about the method; views concerning the assets of the model and lastly views concerning its major limitations. Results of this survey indicated that a clear majority of school psychology trainers were at least barely familiar with dynamic assessment. Of those citing “some” to “quite” familiar only 24% utilised the method. Hence there was till 1992 high awareness but little implementation. Most of those who stated their awareness of dynamic assessment became aware of it through reading.¹²⁷ Within the courses themselves and of those respondents stating at least minimum familiarity with dynamic assessment, 55% spoke about it in class, 32% assigned reading material but only 13% actually taught skills.¹²⁸ The main reason for lack of inclusion of course material within courses was due to the already full programme, followed by a lack of skills to teach the programme with fewer responses citing scepticism about the model. Also the most often cited assets of dynamic assessment were its process-based nature as opposed to mainstream emphasis on product. This was followed by the model's relatedness to intervention and its decreased bias in cross-cultural assessment. The obvious disadvantages cited were the model's demand on technical expertise which the respondents felt they were not equipped to handle along with the time taken for such administration and being able to utilise such assessment within a school setting. Time constraints as well as the fact that neophyte practitioners have to re-learn basic assessment paradigms (and in so doing critically re-appraise their mainstream training) has often been highlighted as problematic within the training of dynamic assessment practitioners (Losardo & Notari-Syverson, 2001; Meyers, 1987).

Haney and Evans' (1999) survey was a follow-up on the Lidz (1992a) study. They questioned the extent of dynamic and non-traditional assessment familiarity and utilisation, once again in the United States.¹²⁹ This survey consisted of 10 multiple choice questions and received a total of 228 responses (46% of the total), thus a larger sample was obtained in comparison to the Lidz (1992a) survey. Of the respondents who answered 93% had over 8 years experience in the field. For the sake of expediency only main results will be looked at here. Eighty-nine percent of the respondents indicated their involvement with children in one or other setting but only 42% indicated minimum to fair familiarity with dynamic assessment. Sixty-four percent indicated that they did not conduct dynamic assessments themselves citing lack of knowledge as the main reason followed by time constraints. The majority indicated that their use of dynamic assessment was limited to children with learning problems as well as those from minority backgrounds. Dynamic assessment was utilised as it enabled a greater understanding of pupils' strengths and weaknesses. In keeping with Lidz's (1992a) results most had come to know of dynamic assessment via reading the literature. The WISC-III as well as the Stanford-Binet IV was the test most often used for assessing minority children. It would seem that the Lidz (1992a) and Haney and Evans (1999) surveys concluded very much the same thing and that in a seven year period nothing much had changed. The results of a United Kingdom study will now be briefly looked at.

Deutsch and Reynolds (2000) conducted the first survey into dynamic assessment as available via training and the perceptions regarding the model within a sample of educational psychologists in the United Kingdom. The main difference in this survey was that the sample chosen to answer the questions had already been exposed to dynamic assessment training between 1994 and 1999. The main aim was to determine how effective the psychologists found the training to be; the extent to which they used dynamic assessment in their practices as well as the perceptions regarding advantages and disadvantages of the model within educational psychology. The questionnaire was piloted and consisted of eight multiple choice questions. As with the former two surveys only main results will be highlighted. A 74% response rate was recorded (out of a total of 119 questionnaires). Eighty-five percent of the respondents were employed by a local education authority and over half had over 6 years experience in the field. Interestingly, it was the experienced psychologists who had the most interest in dynamic assessment (perhaps this is due to years of dissatisfaction within mainstream assessment). As with the two prior surveys, the majority of psychologists had come to know about the method through reading material or had heard about it from colleagues. Forty percent did not use dynamic assessment with over 52% only having used it for less than three years. Of those using dynamic assessment in practice, most spent less than 2 hours per week on the method. Time limits and difficulties of maintaining a dynamic assessment intervention programmes as well as lack of suitable support structures are mentioned as possible reasons as to the low level of utilisation. The linking back of information to the classroom as well as the cost of dynamic assessment materials were indicated as

¹²⁷ Incidentally, the author came to know of dynamic assessment in 2001 after having contemplated the notion of how one could study a person's intellectual potential especially if current product-based scores were not very good indicators. After having searched for terms such as “potential” and “learning to learn” or “teaching thinking” (Narrol, 1996; Niaz, 2001) and so on in various databases it very soon became evident that a research area entitled “dynamic assessment” existed. Apart from a dedicated search into this area, all knowledge has since been obtained from reading the literature, so Lidz's (1992) results are quite understandable. More research results within dynamic assessment in South Africa is available (Murphy, 2002; Murphy & Maree, 2006) although it is now dated. The author has already started a preliminary enquiry into conducting a similar survey of awareness of dynamic assessment among psychologists in South Africa which, it is hoped, will be conducted in 2006-2007.

¹²⁸ The author did have the pleasure of delivering four lectures to third year undergraduate students in the psychology department at the University of Pretoria in 2003. Students were allowed a choice of questions to answer in the final year exam and the dynamic assessment module received the lowest number of answered questions! This could be due to a lack of familiarity with psychometrics or the lack of knowledge regarding the roots of dynamic assessment (Russian psychology and so on). Either way, it resulted in fewer papers to mark.

¹²⁹ Appendix 2 details the content analysis of a questionnaire sent out to over 100 dynamic assessment researchers and/or practitioners across the globe in 2005. Due to the paucity of the results, as mentioned, the results are included as an appendix only. After having the studied previous surveys, the author understands more so than before that the questions were alarmingly detailed thus perhaps contributing to the small number of responses.



negatives. Thirty-four percent indicated that their dynamic assessment training was only partly adequate in obtaining their professional goals citing insufficient basic training as main reason for not meeting their training requirements. Dynamic assessment advantages include its flexibility, its positive experience-creating manner of assessment, its interactive nature, its practical advice-bearing manner offering an alternative to mainstream psychometrics, its culture-fairness and its rich source of information. These were cited as the perceived advantages of the method. Deutsch and Reynolds (2000) contend that the best known researcher within dynamic assessment in the United Kingdom is Feuerstein, who, they state, has not encouraged utilisation of his model outside Israel. In sum, there is a need to circulate more information about dynamic assessment within the United Kingdom.

Greenberg (2000) reflects on the four aspects necessitated by those thinking of adopting a dynamic assessment approach in practice and in keeping with the above-mentioned factors in terms of dynamic assessment's disadvantages, the following is highlighted:

- Various models and theories need to be understood before the embarkation into dynamic assessment
- The devices available need to be understood in terms of their workings
- Dynamic assessment is inherently a loose set of tailored approaches and depending on the context and person being assessed needs to reflect this individuality of assessment
- The need to link up dynamic assessment with classroom intervention poses a particular problem as there is usually a lack of human and financial resources

This is complemented by the following comments from Kaniel (2000) who states that according to his definition of dynamic assessment the following is endorsed:

- Dynamic assessment should include as an inherent feature of its approach assessment and intervention
- What makes the situation dynamic, is the dynamic adaptability of the assessment to the person
- The tasks involved in assessments should reflect the interests of the client and should take place in a relaxed atmosphere
- People themselves are dynamic beings and both process and product should be assessed
- Due to its malleable approach towards assessment, dynamic assessment presents with an eclectic array of techniques in its repertoire and hence should not and usually is not bound by certain strategies (as has been previously mentioned this is perhaps one of the characteristics with which some mainstream assessors have a problem)

In addition to this, the usefulness of professional societies and governing bodies as well as informal groupings may well aid in the further awareness, acceptance and influence of dynamic assessment. Now that a sketch has been illustrated regarding the beginnings of dynamic assessment an equally brief account will be given as to its place within the larger realm of intelligence with which it is most commonly yet not exclusively associated.

2.8 Intelligence

This term is either accurate in its varied semantic interpretations or it falls woefully short of any apt description to date. Some view it with awe and others with suspicion; frequent attempts are made to deny its importance and no less are attempts made to leverage its status as scientific. It appears at once as scientific and pseudoscientific, a misnomer waiting for reclassification into a system more worthy of study or an area already substantially endowed with copious literature attesting to its stature as recognised domain of necessary interest and research. Dynamic assessment needs to be placed within a domain and it is usually the domain of intelligence in which it is fatefully lodged, much to the dismay of some and elation of others. There is much to do with intelligence about which we are uncertain but there is something which none can deny: intelligence's history. Whether it be vilified or regaled, it is nevertheless rich in its historical deposits of research findings and ongoing debates. The innate need (it seems) to classify, judge, compartmentalise and determine a sense of hierarchy can perhaps be said to issue forth from our evolutionary heritage as mammals where the establishment of some sort of controlling system was necessitated in order to live in a more harmonious fashion than that which is frequently reflected in nature. The study of intelligence is a continuous activity carried out by most people everyday in life, our jobs as researchers is merely to systematise this knowledge and information into one system known as science.¹³⁰ With this rather unscientific introduction, a brief detour into the realm of intelligence is now taken.

¹³⁰ A relativist stance once again is proffered as acknowledgment towards other systems of knowledge, but as is now known, the author views "science" as the more relevant system and does so without qualms.

2.8.1 Desperately seeking a definition¹³¹

There is no definitive definition of intelligence and no-one knows what it is (Neisworth & Bagnato, 1992; Newell, 1990; Sternberg, 1997c; Undheim, 1987). Moreover it is a tiresome exercise to try and define it (Jensen, 1982b) as there are seemingly unlimited manifestations (psychometrically and neurologically) of what it supposedly is (Fuster, 2005) and also how it is methodologically represented (Henry, Sternberg & Grigorenko, 2005). It is at once a workhorse and a diva because “the construct is extremely useful, but we do not have a proper definition of what it is and what it is not” (Wilhelm & Engle, 2005, p.7). The word “intelligence” was first utilised in Spencer’s 1855 evolutionary-oriented text on psychological principles which is particularly significant given the author’s own definition of intelligence below (Jensen, 1998b). To understand that the concept was couched in evolutionary terms upon initial use and to witness the definition come full circle some one hundred and fifty years later is indeed simultaneously surprising and unsurprising depending on one’s point of view and frame of reference. Two major theorists within the field have this to say about intelligence and it is necessary to quote verbatim just for decisive and emphatic clarity:

- Butterfield, Siladi & Belmont (1980) - **intelligence develops** (p.96)
- Anderson (1994) - **intelligence does not develop** (p.1)

There are multitudinous views in between the above-mentioned. There is no agreed upon definition of intelligence, primarily because there is no agreed upon construct and secondly, when a construct is defined, it is presumed measurable (see chapter 4) when it is not necessarily the case. So within this contrived situation psychologists are left wanting a construct, a definition and a manner of research.

Dynamic assessment is leveraged upon this.

There are many definitions. Seeing as the literature concurs with both these statements, it is decided upon to use the author’s own definition, seeing as it is as good as any other! Intelligence, simply put, *is the evolved ability¹³² to survive on our planet and this includes the degree to which we can adapt to changing circumstances.*¹³³ There is no teleological side to this statement, it is not that intelligence evolved to some point to assist in adaptation, but due to selective adaptation and successive fit within environmental press, genetic combinations that have survived in their current form are those that were selected “for” in evolutionary terms (Terman, 1921). We are not verbally acute so as to pass SAT’s; we are verbally acute because the genetic-environmental press had it such that this combination was able to successfully exist within successive generations and it just so happens that SAT’s are able to pick up such intellectual adaptation. In fact this definition comes very close to a definition of learning potential as cited by Van der Aalsvoort and Lidz (2002) who state that learning potential overcomes what a definition of intelligence seems to have evaded, namely, how environmental demands influence the adaptive capability of the individual and concurs with Hamers, Hessels and Pennings’ (1996) definition of learning potential as the capacity to adapt to new situations by drawing on past experiences. Sternberg (1996b) also draws attention to the difference between intelligence as measured by conventional IQ tests which do fairly represent future academic achievement and later success in life (even though it has been

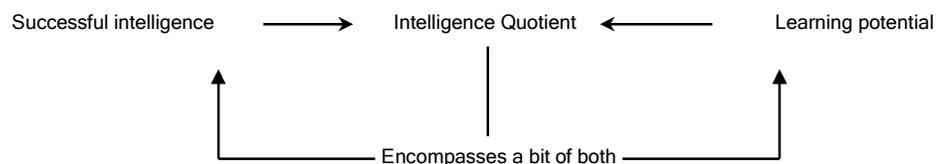
¹³¹ Bereiter’s definition is perhaps the catchiest one, stating that “intelligence [is] what you use when you don’t know what to do” (in Jensen, 1998b, p.111). Now, the effort comes in where one needs to capture the process of doing something when one does not know how.

¹³² In agreement with Jensen’s (1982b) determination of intelligence as a product of biological evolution. But this is obvious - what else could it possibly be?

¹³³ This includes the ability to adapt to a new job, a new country, a new planet; in short anything that will ensure our survival beyond that which attests to our physical survival (Sattler, 2001 in Gregory, 2004). This definition also allows for coverage of animal intelligence, both vertebrate and invertebrate (Menzel & Bicker, 1987) and can be linked to isomorphic renderings of theoretical constructs as already commented on in this study and as Schafer (1982, p.183) said over twenty four years ago; “the electrophysiologically *adaptable* brain should be the behaviourally bright brain” (own emphasis). Intelligence is most likely a by-product or exaptation of our evolutionary change (Kanazawa, 2004), a spandrel in the words of Stephen Jay Gould. Should environmental press favour less intelligent beings then our levels of intelligence many generations hence will most likely be less intelligent. There is no end point of greater intelligence, there is no teleological argument here. Put simply, evolution is change; it is human beings who decided on whether to construe such change as “good” or “bad”. There are no moral codes, no ethical standards; just reality as it has been for millions and millions of years. Is this a relativist take on the process of adaptation? Perhaps some might see it so, which is contrary to the views espoused in this chapter thus far. Yet, it is this stark reality with which many find difficulty in comprehending, especially in the social sciences. Note that the term “ability” lacks clarity and is itself suffused with loose conceptual constructs, for, is an ability linked to general intelligence (at least within the psychometric paradigm)? Or can an ability exist without recourse to correlational links to general intelligence (as evidenced within multiple intelligences?) (Howe, 1996). The definition given above treats the concept “ability” as a layman’s interpretation or an intuitive practical definition. Adaptability is core in this definition because it results primarily due to learning, both at global behavioural and neuronal levels (Posner & DiGirolamo, 2000). Learning is key to dynamic assessment; thus linking the commonality is not that difficult to envisage. The idea of adaptability can also be partially linked to Wechsler’s original definition of intelligence “... to deal effectively with his environment” (Ryan & Lopez, 2001) as well as to Stern’s notion of intelligence being defined by the manner in which novelty is dealt with (Pascual-Leone & Johnson, 2005). Broad-based adaptability also encompasses what the PPIK theory of intelligence asserts is intelligence; namely, intelligence as process, personality, interests and intelligence as knowledge (Ackerman & Beier, 2005). This is appealing but the swift pull of the spiralling vortex can already be felt as one is sucked into yet another “nice” theory which is simply too broad to be rendered useful.

posited that IQ tests are founded on naïve theory of intellectual functioning; Brown, Campione, Webber & McGilly, 1993)¹³⁴ and successful intelligence which he regards as the ability to profit from past experience. Adaptability is key in Sternberg's (1997b) understanding of intelligence and successful functioning and intelligence as the ability to learn (Ones, Viswesvaran & Dilchert, 2005) is hardly much different from the defining criteria for dynamic assessment which is simply more process orientated in this regard. In this vein then, successful intelligence as a theoretical construct comes closer to learning potential as theoretical construct. One could perhaps envisage a continuum with IQ bridging far left and far right constructs such as depicted in figure 10 below and is similar in nature to figure 11 which contrasts lay-conceptualisations of intelligence with those of expert opinions. The author does however disagree with the sentiments conveyed by Greenspan and Driscoll (1997) who state that adaptive functioning is sometimes utilised to counterbalance IQ referring to aspects of personal competence other than what is traditionally understood to be IQ. It is emphasised therefore, that adaptive functioning¹³⁵ too has manifest variance; some are better able to adapt than others in certain situations (also by creating and shaping their own environments; Sternberg, 1997c) and so on and this is evidenced in developed countries where adaptive increases in cognitive ability is promoted by environmental factors which are not as prevalent in developing countries (Barber, 2005). Adaptive functioning in our evolved past had resulted in a plastic brain which, some say, caters for an environment quite dissimilar from the one we currently inhabit (Strauss, 2005) although the counter argument here is that the very uniqueness of human brain plasticity makes our adaptations so much the better. Trainability, which is a key point within dynamic assessment concerns, is yet another link that can be fostered in bridging the gap towards intelligence as is evidenced in the literature, those with higher IQ's are more receptive to training and are hence more "trainable" (Lynn & Vanhanen, 2002). This is clearly seen in much of South African dynamic assessment research. *G* has also been proposed to have evolved as domain-specific adaptation to a narrow sphere of "evolutionary novelty" which has since proven its worth due to the novel-rich world in which humans currently find themselves resulting in *g*'s now "general" status (Kanazawa, 2004). This ties in with the above statement regarding developed countries which indeed have more novelty impinging on their brains. Kanazawa's (2004) theory does however fly in the face of the generally accepted understanding of *g* as general underlying mechanism as it proposes *g* as merely one module in the evolved brain's repertoire of modularised functioning and does not build on empirical support¹³⁶ (Borsboom & Dolan, 2006). Perhaps it ties in with Sternberg's triarchic theory in some manner as this theory has not evidenced a general factor pervading the sub-tests of the ability tasks but rather specific abilities (Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996) so perhaps it is worth looking towards modularised *g*. The latter authors cogently argue for *g*'s hypothetical nature as source of individual difference and not as a reified mechanism of mind. This notion is reiterated throughout this study.¹³⁷ Hambrick's (2005) evidence illustrates that *g* does not necessarily account for expertise in task performance or on-the-job performance as much as domain knowledge for instance which accounts for greater variance explained as predictor of performance. The accumulation of knowledge in general but in particular task-specific knowledge thus plays a large role in determining success at the task. However, the retort to this (in support of *g*) could be that *g* underlies this very ability to accrue information more efficiently. It seems that everywhere one goes, *g* is sure to follow.

Figure 10 A possible alignment of IQ within the continuum of successful adaptation as differentially measured



Adaptability could mean many things to many researchers and in keeping with Hansen's (2003) link between intelligence and learning ability a case could be made for linking the above broad definition of dynamic assessment, "[they] become better learners in other words, they become more intelligent" (p.60). Donald (1997) views adaptability at the neuronal level noting how neural plasticity has allowed for human cultural development of language and writing which, he states, is not a module housed within the brain but reflective of the brain's changeability to its environs. Moreover he argues that cognitive "fundamentals" are

¹³⁴ For instance, certain sub-tests utilised within intelligence tests such as inductive and deductive reasoning tasks are in fact not unified in terms of the underlying construct they purportedly measure (Wilhelm, 2005). This thesis does not look into the matter of the nature of the tasks utilised in test batteries and upon visiting some of the literature concerned seems to be yet another contentious area of debate within the field.

¹³⁵ As Greenfield (1998) states, everyone has the ability to adapt and she refers to this as "panhuman genotypic intelligence" (p.81), but it is not this which intelligence tests measure.

¹³⁶ One could alternatively also argue that the factor analytic tradition was too heavily data-driven and theoretically weak (Taylor, 1994). So which do we need? Or at least, what criteria should be used to determine the veracity of a theory; its data or its theoretical grounding? This discussion will be taken up in chapter 3 when the method of science is discussed in greater detail.

¹³⁷ Will there ever be a resolution? Perhaps we are once again asking the wrong questions or at least asking them in the wrong way. Can natural science methodology aid us in this endeavour at arriving at some sort of resolution? Here, a plea is made for psychology to deviate and follow natural science methods of explaining away occurrences. It is anyone's guess as to how successful such an endeavour will prove to be.



not necessarily biologically universal but unique to specific cultures. Berry's (1998) opinion on intelligence also pivots the notion of adaptability but in this instance it is the adaptability to culture which is essential in determining group survival. Different cultures vary on their responses to different test items, thus "phenotypic intelligence varies from culture to culture" (Greenfield, 1998, p.81). It also veers closely to the definition of Feuerstein's modifiability of the individual (Schur, Skuy, Zietsman & Fridjhon, 2002) and buffers the notion of cultural adaptation and deprivation as paramount to intelligence functioning as opposed to a strict and narrow construal of intelligence per se as utilised within Euroamerican psychology (Berry, 1998; Gardner, 1998) or as perceived by society in general (Carroll, 1998). Via the meaning of adaptability, learning as a process-oriented concept also veers away from constraining itself in reified terms as the static definition of intelligence is often viewed (Feuerstein & Kozulin, 1995). Assessment of adaptive functioning within various contexts other than those assessed within schooling contexts reveal holistic pictures of general functioning which can easily shed light on functioning within specified narrow contexts (Oakland, 1980). There are many research traditions within the intelligence field emanating from across the world with numerous "mini" traditions in each. There are remarkable similarities between some traditions evidencing coalescence of thoughts and thus indicating that there are indeed universal concepts and criteria of "intelligence" (Sternberg, 2004b). Although layperson definitions of intelligence also form part of many cultural systems and appear very different in nature to one another (Baral & Das, 2004, Mpofu, 2004) far more testing carries on unabated informally than does so within controlled testing environs (Weinberg, 1989).

Natural intelligence and artifactual intelligence as described by Glaser (1998) in a manner resembles Vygotsky's lower and higher order thinking where performance within a given cultural setting is normal but skills need to be taught for higher order processes as required by schooling. It is the latter which is problematic within culturally diverse and deprived individuals and usually not the former. Of course this implies that the latter test for only school-like subjects, which it does, more often than not. Hence, are the correct questions being asked? Colvin (1921) stated many years ago that he was, in principle, in agreement with the above-mentioned definition but considered it too broad as it encompasses instinctive as well as learned behavioural adaptation to the environment. This is true but evolutionary adaptation to the environment is no less an indicator of intelligence than anything else yet once again criticisms are lodged at this broad notion of defined intelligence (Cowan, 2005). Intelligence research is perhaps problematic for the reason that we might very well be asking the wrong questions¹³⁸ (Estes, 1998) and it must be noted that definitions are only as good as their utility in explaining away aspects pertinent to the intelligence debate (Zigler, 1998). Utility value is perhaps the most obvious in naturalistic or everyday settings where intelligence is valued according to a number of "lay" criteria, that when studied closely, reveals its similarity to expert definitions of intelligence (Kail & Pellegrino, 1985). Yet, Derr (1989) and Sternberg (1979) warn against the admixture of both lay and informed views of intelligence definitions as this matter is one for science and not for ordinary discourse. Derr (1989) also posits that some conceptual confusion could well dissipate if such considerations were given their due. Unfortunately, common understandings of intelligence have not yet filtered through to informed conceptualisations as previously envisaged (Turnbull, 1979). Figure 11 illustrates the extent of considerable overlap between lay and expert opinions on intelligence and overlays a few concerns and areas of connectivity to our evolved selves. Figure 12 which follows illustrates Greenspan and Driscoll's (1997) content model of personal competence which itself draws on Sternberg's research into lay and expert opinions regarding intelligence. Note the overlap between social competence, intellectual competence within the "everyday" sphere and the academic sphere. As is evident, this represents a continuum approach towards the understanding of intelligence and what it means to function intelligently in life in general. Overlaid are aspects of importance to dynamic assessment's understanding of how intelligence functions. The authors attempt to capture within one model the overlapping areas concerning intelligence and personality; an area already addressed within some dynamic assessment initiatives.

¹³⁸ Could it be that the situation will resolve in a manner surprisingly simple yet effective? Perhaps our questions are not so much incorrect as too complex? Possibly not but it is something over which to ponder.

Figure 11 Lay and expert opinions about intelligence (Gregory, 2004, p.142; Sternberg, 2004c, 2004e). Concerns about interpretations

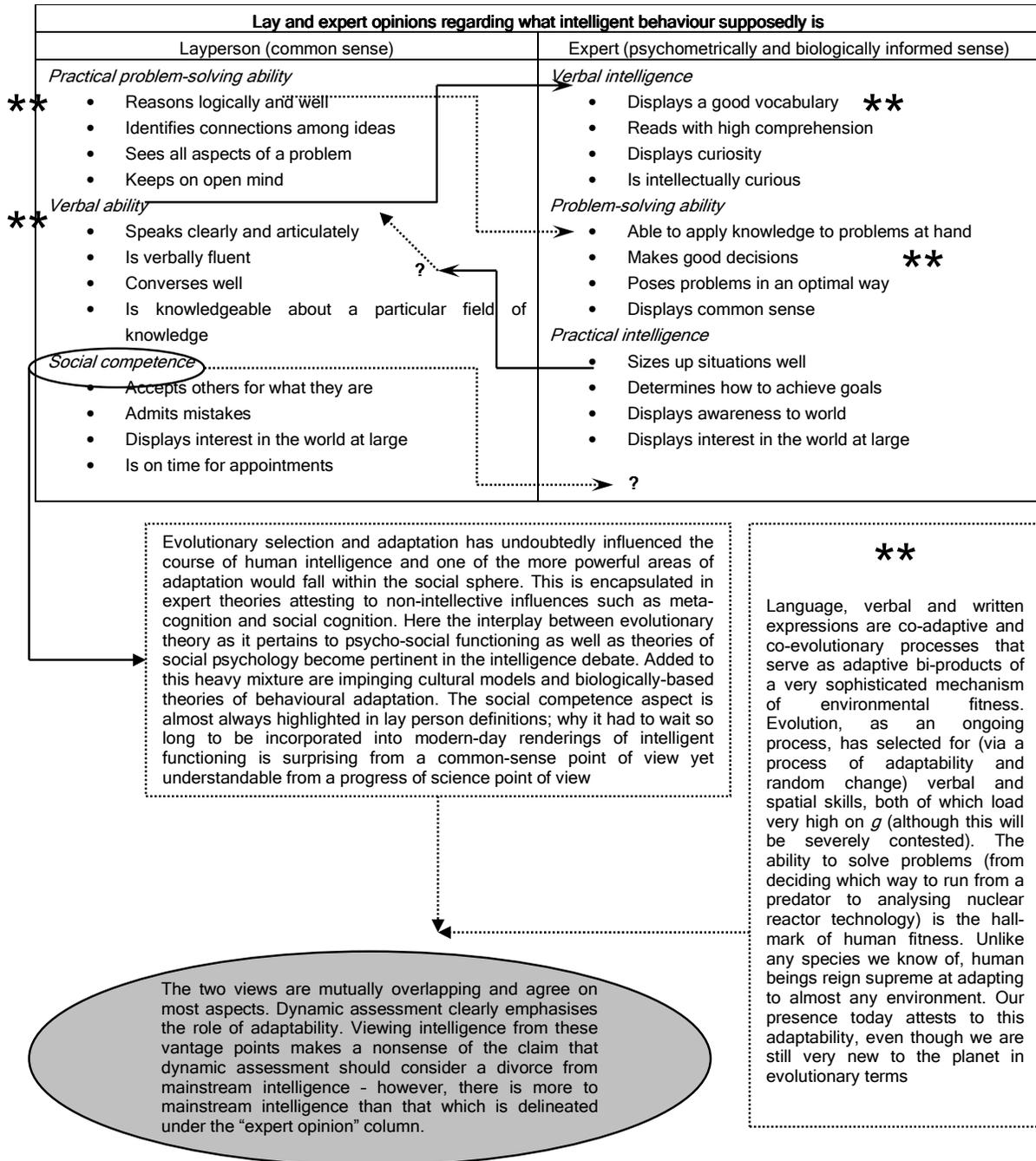
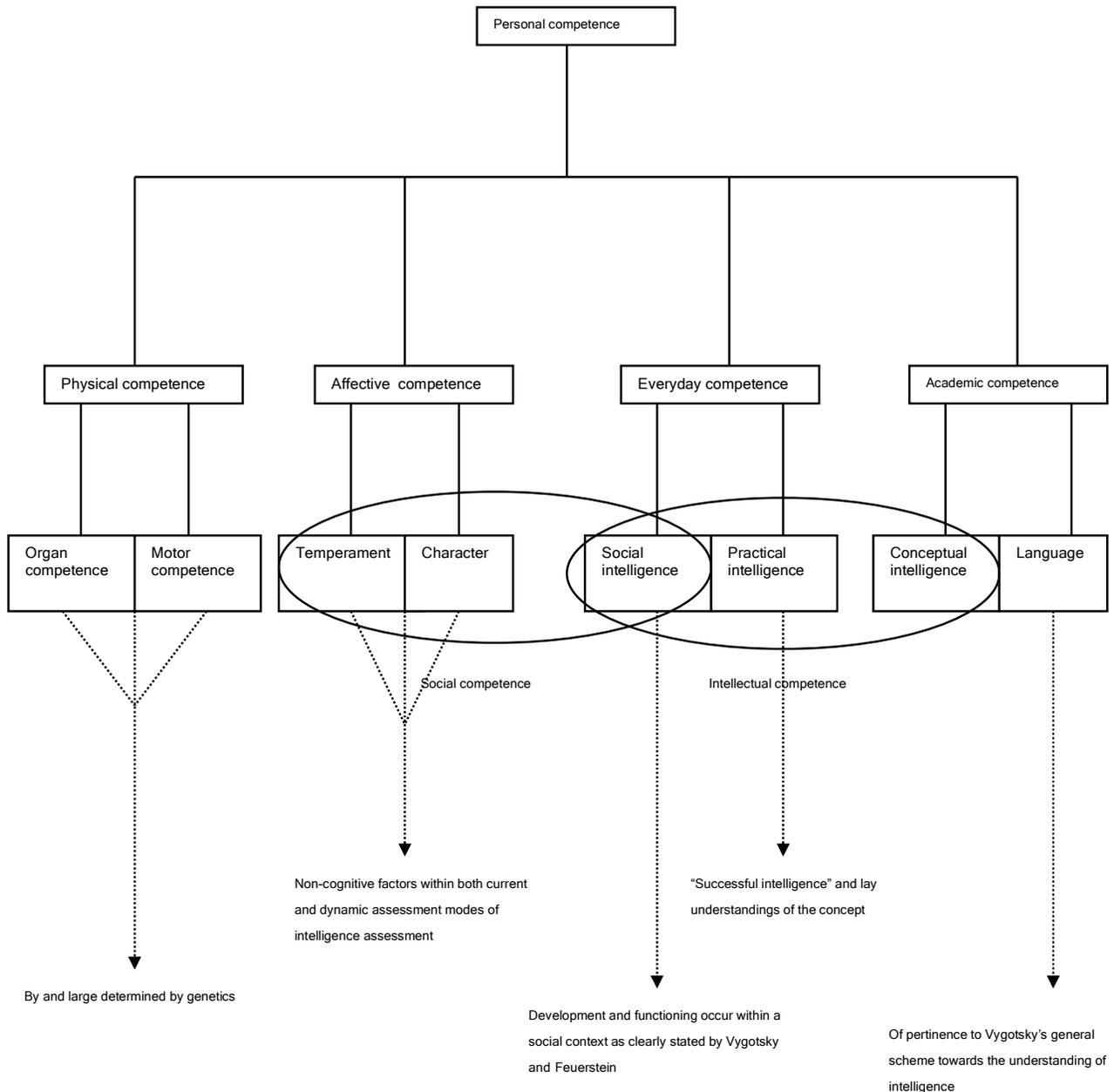


Figure 12 Greenspan and Driscoll's (1997, p. 133) content model of personal competence with attending aspects of importance to dynamic assessment



If we look at figures 11 and 12 it is immediately evident that one aspect stands out: that of practical or social intelligence which plays a large role within evolutionary theories of intelligence and survival (ability or proclivity to detect cheaters and so on cf. Cosmides and Tooby in general). It is also the construct within intelligence research which most appeals to intuitive ideas of intelligent functioning (Matthews, Zeidner & Roberts, 2005). Kanazawa (2004) firmly states the case for *g* as evolved modular ability and posits the following regarding *g*:

1. Intelligent (high-*g*) individuals are better able to solve problems than less intelligence (low-*g*) individuals, *only if* the problems are evolutionarily novel
2. Intelligent (high-*g*) individuals are no more able to solve problems than less intelligence (low-*g*) individuals, *if* the problems existed in the EEA [environment of evolutionary adaptedness] and are thus evolutionarily familiar

Hence, the case for adaptability and "everyday" or social intelligence has merit which can be tapped by dynamic assessment theories and models more so than current mainstream assessment which concentrates on only a specific sub-set of intelligence-eliciting tasks. However, and this is a big however, it is these very sub-sets of tasks which are manifestly evident and obviously so in the world we now inhabit. If you want to be successful, one does need high *g*-loading abilities for specificities and this does not seem as if it will change any time soon.

2.8.2 Approaches towards the study of intelligence

Utility and parsimony (see chapter 3) are hallmarks of good theories and intelligence theorising is an area of investigation which needs severe delimiting or theoretical excision in order to prove successful (Snow, 1998b). All theoretical programmes (for instance information processing, learning, factor-analytic and cognitive developmental) are by their natures limited to and by their areas of investigation and points of departure (Li, 1996).¹³⁹ Intelligence theories explain different things about the same phenomenon, with some theories tending to seek structure, others seeking causes of such structure whilst others emphasise function and thus results in different conclusions (Carroll, 1994).¹⁴⁰ In accordance with the above definition, is the embedded notion of intelligence as problem-solving ability as one necessarily has to solve various problems in the game of adaptation (Wenke, Frensch & Funke, 2005). Along with the assessments of problem-solving myriad other aspects tie in with intelligence assessment, such a cognitive speed, perception, attention and memory, hence the diverse array of intelligence sub-componential research areas. Is it perhaps possible to study intelligence without the use of the concept itself (Grigorenko, 2004b) seeing as it poses such problems? The concept “intellect” (reason) and “intelligence” have followed different historical paths and can be separated within the mainstream intelligence arena, at least within the predominant traditions. Russian concepts of intelligence have fused the two rendering a definition of intelligence without recourse to the necessary definition so espoused within Western traditions. Soviet psychological history and Vygotsky’s movements within have already received brief attention above, so it is not surprising to find that intelligence research as well has received varied interpretations (Grigorenko, 2004b).

Factor analytic explanations of intellectual structure can inform and be informed by other manners of viewing intellect, so no one view can be said to represent all there is to intelligence. Behavioural genetics can at least attempt a partial explanation of how and why factor analytic structures form in the manner they do (via intelligence assessment and general modelling of cognitive growth; Cherny, Fulker & Hewitt, 1997); the father of whom is Galton (Jensen, 1997) who also introduced twin and adoption method studies (Bouchard, 1997). Given the time and context during much of early intelligence research historical development, is it really surprising that statistical envisioning of intelligence predominated? (See chapter 3 for more on the context prevailing at the time). Psychometrics aims to measure and quantify whereas more biologically attuned models of intelligence seek to provide explanation about developmental change while behavioural genetics attempts to account for heritability changes throughout life and how this impacts on or is impacted by g (Ceci & Bruck, 1994). See figure 15 for a depiction of these views. The plea for subsequent melding of approaches can only but benefit the discipline although Sternberg’s (1997) warning cannot be ignored where it is stated that the biological approach towards intelligence has yet to offer lucid approaches towards the integration of biological models of learning and how learning occurs within practical settings but the statement from Brody (1992) also cannot be disregarded, that intelligence is a heritable trait. Note here, however, that dynamic assessment translates “trait” into “state” and hence moves away from the immutability concept to one of modifiability (Feuerstein & Feuerstein, 2001). This implies that there are biological correlates of intelligence with findings from inspection time studies to event-related potential studies informing different aspects of information processing (Fernández-Ballesteros & Colom, 2005; Li & Kunzmann, 2004). “It is relatively easy to discover the biological correlates of intelligence but it is relatively difficult to determine the *causal* relationship between a biological measure and intelligence” (own emphasis) (Brody, 1992, p.215). Researchers still remain within a theoretical vacuum regarding intelligence theories even though technological advances have forged ahead (Eysenck, 1994). Looking more closely at the delineation of these four paradigms of intelligence explanation (considered the prevalent ones but by no means the only paradigms), it can be seen that dynamic assessment philosophy is more firmly entrenched within the learning and cognitive developmental paradigms. Figure 13 depicts these views. Figure 14 illustrates the realms of theory and substantive realities and the placement of intelligence and dynamic assessment within them. Note the schism between the two realms. This notion is continuously echoed throughout this study. As Sternberg (1994c) succinctly notes, if conventional intelligence tests only predict between five and ten percent of various life measures of adjustment and success where has the other ninety to ninety-five percent disappeared to? Clearly this hypothetical construct cannot adequately account for reality as it is experienced.

¹³⁹ The same type of delineation of research angles can be seen within neuroscience endeavours where brain functioning can be mapped according to sequential anatomical, functional or organisational frames of reference (Chugani, Phelps & Mazziotta, 1994).

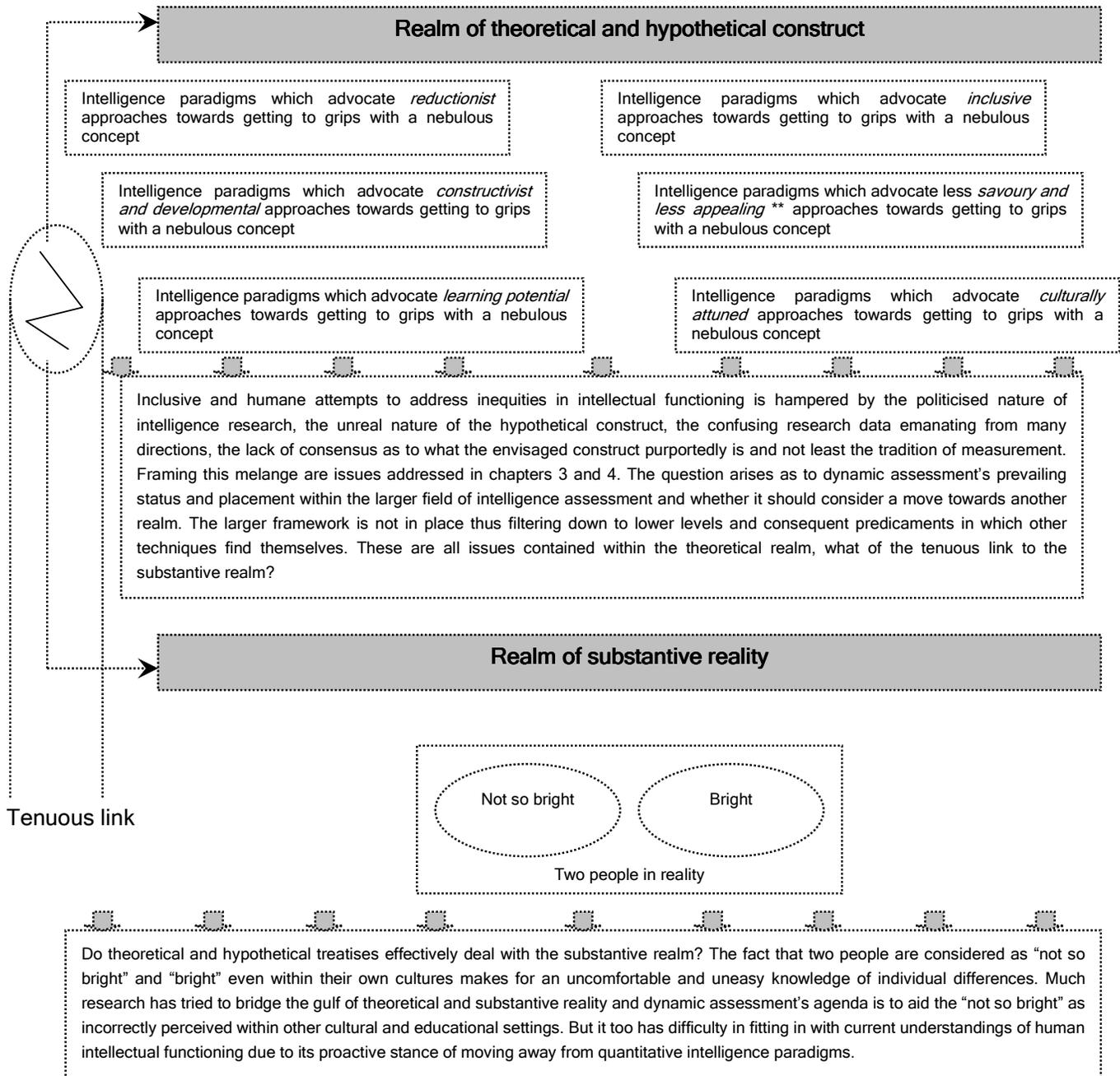
¹⁴⁰ The number of frameworks used to study intelligence in fact almost mirrors the number of definitions available! Another simplistic manner of investigating intelligence would be to divide theories and models into what Weinberg (1989) refers to as “lumpers” and “splitters” referring in turn to g -based and multiple intelligence based approaches. Perhaps the manner in which we choose to study intelligence is itself reflective of initial tacit assumptions.

Figure 13 Paradigms within intelligence explanation (Li, 1996) with problems overlaid

Some current prevalent influential paradigms and their attendant problems (<i>with Sternberg's 1997 divisions in brackets</i>)	
<i>Information processing paradigm (psychometric)</i>	
o	Sternberg's triarchic theory - broader and more encompassing than its predecessors in its inclusion of social and contextual factors. The notion of learning is not, however, considered an independent construct
<i>Learning paradigm (anthropological and computational)</i>	
o	Butterfield's components of intelligent action - in which learning and the information processing paradigm are both accorded framework status. For instance, once a knowledge base becomes exhausted in terms of aiding in skill acquisition, learning is said to occur, where new executive routines are learned
o	Brown and Campione's learning potential - change construct is problematic because its foundational construct is itself problematic. Dynamic assessment fits in here as well as below under cognitive-developmental models
o	Snow's six aspects of intelligence - learning is adaptive and changes according to the requirements of the context. Learning differs not only between but within people depending on the context
o	Schank's artificial intelligence - recursive argument ¹⁴¹ inherent yet can be extremely useful. By considering learning, understanding and explanation without the attendant non-intellective aspects via a process of information gathering, learning very much hinges around adaptive functioning calling on past experiences in informing present decision-making
o	Perkins's learnable intelligence - if intelligence is learnable then we all learn in differing ways, which brings us back to the original concern of individual difference research, so one might just as well change the research label from intelligence to learnability. Has some common concerns with dynamic assessment
<i>Factor analytic (psychometric and biological)</i>	
o	Jensen's <i>g</i> - as has been stated repeatedly, the author's preferred realm of discussing intelligence but still too narrowly defined to be useful in the practical world (at this point in time, perhaps the future will be more conducive as technology improves)
o	Also known as the laboratory-based approach towards intelligence (Gardner, Kornhaber & Wake, 1996)
o	Eysenck's biological basis of intelligence - in keeping with the above, the author seeks to forge cross-disciplinary expertise for intelligence, dynamic assessment and biological basis of behaviour and intelligence. Yet this too is not practicable at present
o	Horn's psychometric construal of intelligence theories as being a mixture of various approaches towards understanding intelligence - who has tried to make a case for both multiple and unitary forms of intelligence. Which has resulted in our asking whether those ill-posed dichotomies are really useful within the social sciences? See chapter 3 for more
o	Ackerman's Radex model - visually similar to Feuerstein's cylinder model yet different in its conceptualisation
<i>Cognitive developmental (epistemological and systems)</i>	
o	Piaget's cognitive-stage theory - a group of theories now known to be inaccurate and contested (in detail not necessarily as a whole) yet powerful in its descriptive novelty - dynamic assessment fits in here as well as above under learning models
o	Glaser's cognitive efficiency theory - conceives of natural and artifactual intelligence with the former being inherently learned and the latter being taught. Similar overtones can be seen in the theories of dynamic assessment models. How one performs is dependent on what one has learned but does this answer the question of the nature of intelligence?
o	Zigler's social competence - the non-intellective is stressed in this research model such as motivation. Increasing awareness of non-cognitive aspects is a current trend within intelligence research and has formed part of the dynamic assessment paradigm since its origins
o	Gardner's multiple intelligences - a modularised approach in determining the nature of intelligence which ironically by its very arguments for modules has highlighted the role of governing <i>g</i>

¹⁴¹ Very similar to Dennett's views concerning qualia and studies relating to this topic. "It's not hard to see how philosophers have tied themselves into such knots over qualia. They started where anyone with any sense would start: with their strongest and clearest intuitions about their own minds. Those intuitions, alas, *form a mutually self-supporting closed circle of doctrines*, imprisoning their imaginations in the Cartesian Theatre" (own emphasis) (1993, pp.369-370). The inherently recursive or closed circle is perhaps the biggest draw-back when one views the workings of the brain from a computer-based perspective.

Figure 14 Intelligence within two realms

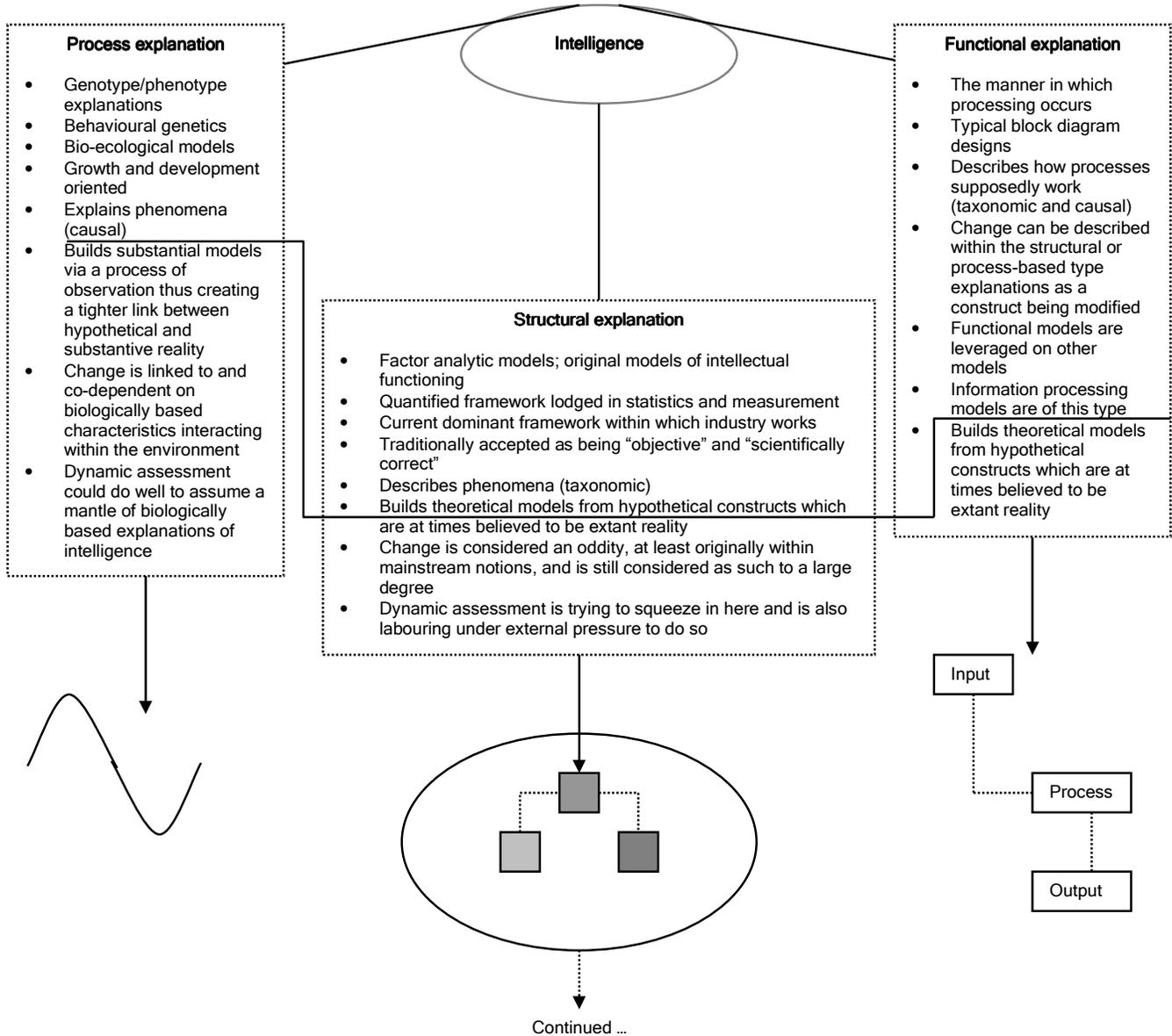


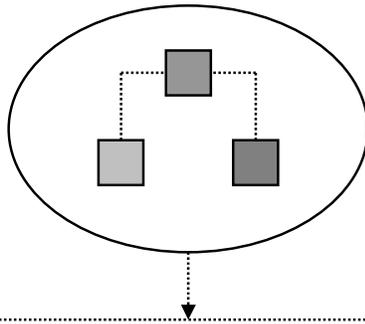
** however unpalatable and currently socially unacceptable it is to mention social inequities, one simply cannot turn a blind eye to stark reality which has, in the past, does so currently and will most likely do so for many generations to come, illustrate that not all are allotted equal propensities and opportunities in life in general. This does not in anyway entail a need for a eugenic type resurgence of any kind but it does behove the researcher to acknowledge that such differences are manifestly obvious but need to be tactfully approached (Loehlin, 1992). Simply because some compensatory education programmes may not have yielded great gains in the past (Jensen, 1969) does not mean that mediatory steps should not be taken to try address low performers; as large scale compensatory programmes cannot necessarily be equated with dynamic assessment for instance. One need only look at the recent literature (Hunt & Sternberg, 2006; Jensen, 2006; Sternberg, Grigorenko & Kidd, 2005; Templer & Arikawa, 2006) to view the still-current and heated debates surrounding intelligence and race. In keeping with tenets developed in chapter 3, the author fully upholds the sentiment advocated by Detterman (2006) in his editorial review policy for "Intelligence" which is worth citing word for word, "I believe that it is important that controversial ideas have access to the pages of this journal. Without a forum for the resolution of controversy, controversy will not be resolved and science will not advance. If a journal does not advance science, then what good is it? All it can do is fortify the status quo" (p.iv).

Figure 15 Different approaches towards the study of a single issue

Do not know what "intelligence" is but can surmise from what we do already know. It is a construct-fuzzy notion

It is one thing and many things. It is *g* in turn made up from many *s*'s. It is not a unitary concept but somehow we always find it lurking in the data





Elaborating on the structural explanation

The structure of intelligence and its key historical representatives (Brody, 1992, 2004, Davidson & Downing, 2004)

- Cattell's theory of fluid and crystallised intelligence which resembled Hebb's theory of intelligence A (biological) and B (cultural). Fluid is largely genetically determined and decreases with age whereas crystallised increases with age in tandem with experience. Cattell's theory grew from the factor analytic work of Thurstone and his primary mental abilities and consisted of second order factors as well. Horn modified Cattell's theory in more recent years and utilised cognitive measures to derive the second-order factors as opposed to Thurstone's factor analytic analyses of primary mental abilities - note here the difference between theoretical and substantive measures! These initiatives sought to move away from Spearman's global dominant notion of *g*. Humphrey re-analysed Cattell's data and showed that the two factors could be subsumed within a global factor and was in keeping with Vernon's hierarchical notion of intelligence (note the swing of the pendulum)
- Guttman modified factor analysis and developed the Radex theory in which complexity and content differed for various tests which could be displayed and conceptualised as a simplex and circumplex model and is visually appealing
- Guilford's theory illustrated his idea of intelligence (non-*g*-based) as being made up of three dimensions; operations, contents and products. Along these dimensions operated five areas, cognition, memory, divergent production, convergent production and evaluation which could be applied to four content areas; figural, symbolic, semantic and behavioural (the author often thinks of Feuerstein's model when viewing Guilford's model). Guilford's model has been criticised for being too cumbersome and complex especially given the data he utilised which was seemingly unresponsive of his own theory
- Gardner's theory of multiple intelligence can be slotted into the structural category although it can and has been accommodated under cognitive developmental (epistemological and systems) categories (see figure 13 above). This categorisation depends, of course, on the literature. Gardner has garnered support for his view on intelligence from various areas such as brain damaged patients, non-average performing individuals such as savants, expert performance analyses, experimental psychological tasks and from psychometric findings

2.8.3 Situating aspects of intelligence measurement

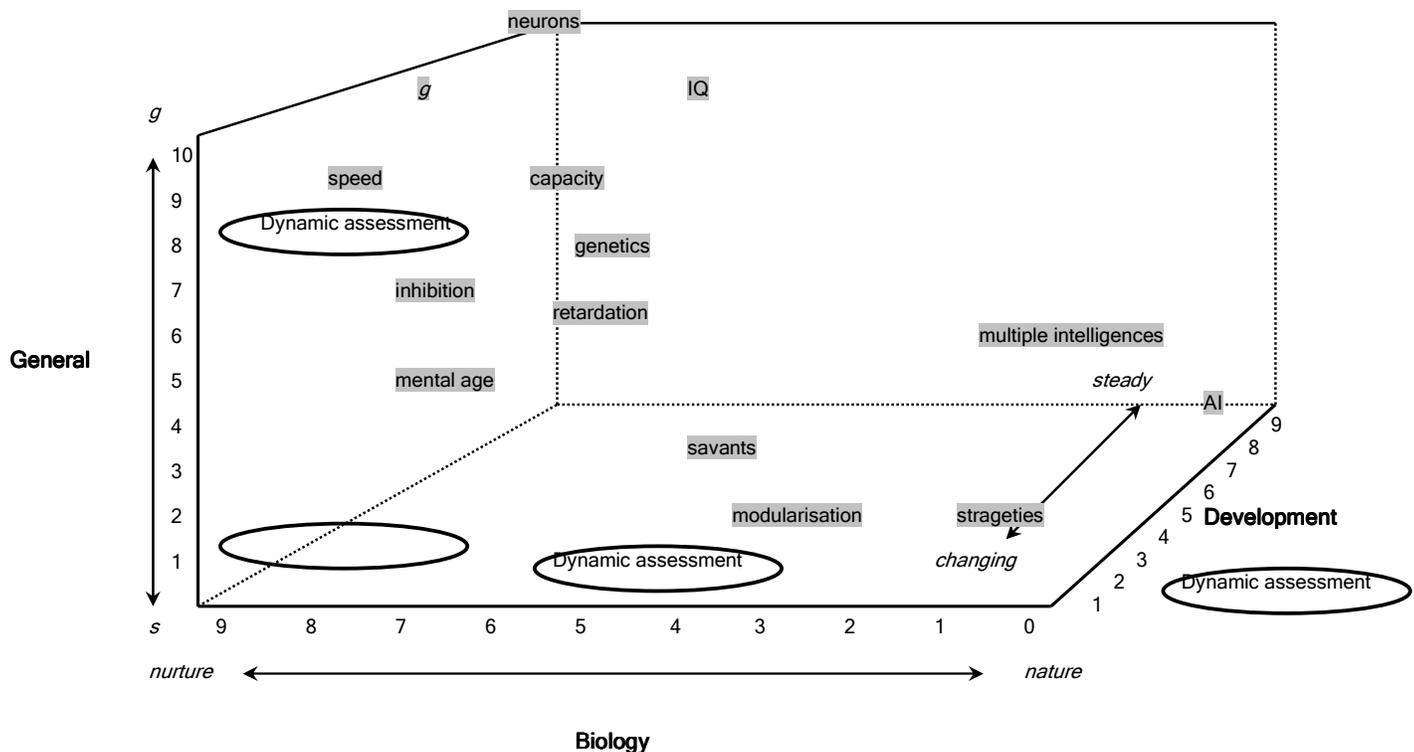
Our intentions are of course academic and scientific and the preoccupation with the nature of intelligence is, it seems, quite typically human behaviour, or is it? It would be wrong to state that it is only the Western countries that pay particular attention to intelligence but it would not be wholly incorrect to state that they do emphasise the notion rather heavily, yet most cultures evidence a recognition of the notion of "intelligent" behaviour (Woodcock, 1998). Having preceded this section with a brief digression into Russian psychology and the role played by Marxist thought may lead one to ponder the role of capitalism and how this has shaped the West's eugenic past and one-time much maligned obsession with levels and degrees of intelligence pivoting notions of retardation and normality (Hodapp & Zigler, 1999). Intelligence and as some would have it, learning potential (Taylor & Richards, 1990), is not a unitary concept (despite the authors' leaning towards reductionist research traditions¹⁴²) (Carroll, 1994; Horn & Noll, 1994; Hunt, 1992; Miller & Vernon, 1992). Somehow it seems to appear unitary at times (or at least very general) and possesses a life of its own, one whose trajectory is almost impossible to pin down even with the most skilled and computationally sophisticated operations; its scope is, in essence too broad (Baltes, 1998) and attempts to engage with a definition from this viewpoint is seen by some to be futile (Horn, 1998b). The need to refine this large base is understandable if one is grapple with a complex phenomenon (Eysenck, 1998). In similar vein to what will proceed below when change based item response models are considered as possible solutions to the gain score issue within classical test theory, it can clearly be seen that much intelligence argumentation takes place within the arcane realm of sophistry where the deployment and

¹⁴² It is difficult to accept that one's preference for a definition rests on something akin to Eysenck's (1998) "error-free transmission of information through the cortex" (p.71). This definition is no doubt very plausible and most likely correct in its circumscribed task of assessing the transmission of error and information through the brain but there just has to be more to intelligence than that! However, as Horn (1998) states, much current biological research can serve an overwhelmingly important role for instance in trying to determine why fluid intelligence decreases with age. Is it possible that decreased blood flow to certain areas results in the loss of relatively quick functioning? Perhaps lifestyles, he adds, should be assessed to help reduce the loss of functionality brought about from the normal aging process.

implementation of sophisticated tools allows special entrance into the labyrinthine and detailed world of nothing other than a face validated extant reality. This is all there is when one considers what intelligence is. We simply do not know. All our tests are based on an initial assumption (face validated of course) of what it is we presumably think we want to measure (measurement: another complicated and contrived notion which will be dealt with in chapter 4) and early intelligence research was predicated almost entirely on measurement (Valencia & Suzuki, 2001). Measurement instruments aid our endeavours but accounting for unobservables within intelligence makes the task infinitely more complex (Humphreys, 1998). To state this so boldly in a treatise dealing with the explicit placement of dynamic assessment within intelligence is tantamount to heresy. But heresy it will have to be. This is undoubtedly the author's views and the following will deal with more scholarly treatments of what intelligence is considered to be and dynamic assessment's place within it.

As described above, the field of intelligence research is so vast, that it would not be too far off the mark to state that it has one of the longest, richest and most confounded histories in psychology (physiological psychology is included here) and a search for a model which can account for this rich domain is nigh impossible. However, to atone for this rather academically ungracious attitude, figure 16 below is presented which at least encompasses the major aspects within intelligence assessment as presented throughout its officially documented Western-favoured history (1904, the year of Binet and Spearman's practical and theoretical contributions; Brody, 1992; Cattell, 1998) as well as its general preoccupations today. This model also fulfils the dual function of allowing one to place dynamic assessment alternatives within it.

Figure 16 Anderson's (1999, p.5) terrain of intelligence in which three principle axes are delineated according to which placement of various forms of intelligence are situated



Unfortunately this figure does not bear a true resemblance to the one of Anderson (1999a) whose illustration shows the three-dimensionality of the various placements in terms of gradations.

Here, in the figure above, the placements have been situated on a two dimension map. Dynamic assessment has been placed along each of the three axes as decided upon by the author and not Anderson (1999a). Regarding the "general" or structural axis, dynamic assessment assesses for skills that are both domain general and domain specific for which there are batteries developed to assess for such skills. Feuersteinian dynamic assessment mediates general skills and prefers to veer away from scholastic type activities as intellectual potential is more easily accounted for within generalised cognitive and metacognitive skills possessed by all alike regardless of cultural context (Mandel, 2002; Paour & Cèbe, 2002). Domain-specific or curriculum-based assessment ties in very closely with specific skills required in certain contexts hence dynamic assessment's placement at both poles of the continuum on the general axis. The truest predictor of performance within a highly specific skill is to assess for precisely that skill (Horn, 1998a) which of course does not define intelligence in any way, merely the skill to perform adequately

the task assigned. If you want to assess piloting skill in an Airbus 380 then it would be best to assess for this skill in an Airbus 380 regardless of test bias (in fact it is strongly suggested that the test be very biased indeed)! On the other hand, assessing for general cognition for placement in educational programmes is rather more general in scope and hopefully less biased and the requisite roles played by specific abilities in the former is of far greater importance than general intellectual functioning as expected in the latter (Snow, 1998a).¹⁴³ Regarding the “biology” axis, dynamic assessment practitioners and researchers view both nature and nurture as equal contributors to the development of intelligence on average even though research within both the social environment and behaviour genetics realms are often at odds (Scarr, 1998a). However, various specific traditions align quite closely to one of both poles. This is merely an average representation. Regarding the third axis, “development” dynamic assessment tends to view the individual as changing, which is perhaps its defining feature within intelligence. Due to its acknowledgement of the necessity to co-exist alongside static intelligence assessment philosophy it is not placed at zero or one but rather at two along this axis. Traditional conceptualisations of intelligence are plotted to the upper left of the figure where dynamic assessment would most likely lodge itself at the bottom left. Recall that dynamic assessment is a manner of assessment and is not a theory of intelligence,¹⁴⁴ however, it is so closely aligned with intelligence research that the two are often difficult to pry apart. Dynamic assessment is not a theory or model of intelligence but a model for the assessment of changing individuals which encompasses notions of intelligence development. This rather fuzzy area is one of the many aspects of this field with which some may find it difficult to work. Anderson’s model is notable for its inclusion and integration of both high and low level approaches towards the study of intelligence, for *g* it is at once recognised but cognisance is taken of development and growth noting that adult intelligence and child development are not necessarily represented by the same construct (Gardner, Kornhaber & Wake, 1996).

2.8.4 Omnipresent issues within approaches towards intelligence research

Various factors play in on the study of intelligence and have an overwhelming influence on the basic philosophy underlying anyone’s basic idea of what intelligence is supposed to be. Issues yet to receive consensus:¹⁴⁵

(i) *Developmental issues* are still not sorted out (is development stage-based or punctuated with large scale changes or is it a continuously smooth progression of change?). This impinges on measurement theory (chapter 4) as well as the gain score issue within classical and modern test theory where models have and are being developed which try to navigate around the problem of what the construct of intelligence or potential in fact is. Measurement is paramount to the discussion on intelligence because this is the manner according to which we describe psychometric intelligence within psychology and how and why we measure and subsequently takes us back to fundamental science practice within global science (see chapter 3). The bridge between sound theory and measurement seems to have broken down (assuming it was ever intact in the first place) (Styles, 1999) as theory alone will not suffice as adequate explanation of intellectual behaviour (Hunt, 1995). Quantification plays no larger a role within psychology than in the area of intelligence hence the very pressing need to establish and understand more fully the rules and regulations governing what is and is not measurable and if measurement is in fact possible at all. Prior erroneous assumptions of “physical to psychological” led pioneers down a path (not to be snubbed) which was to diverge and crack at certain points along the way (Styles, 1999). The need to bridge the cognitive-psychometric gap is also a feature of newer models of modern test theory which tries to grapple as best it can with constructs evidencing change and so borrow concepts from the cognitive realm and fuse it with the psychometric realm. Intelligence cannot be divorced from the context in which it routinely appears and so environmental and developmental concerns cannot be partialled out of the search for adequate assessment tools (Ackerman, 1994). Bio-ecological theories of intelligence and their psychometric counterparts are at times viewed with suspicion in terms of their renderings of *g* and the role this should or does in fact play within assessment (Ceci & Bruck, 1994). Nevertheless, the current trend is to veer towards developmental explanations of behaviour and intelligence (Wahlsten & Gottlieb, 1997).

(ii) *Biological issues* are also still not sorted out although in fairness to most researchers, much modern literature seems to have abandoned the 1970’s dichotomy as a weak or even bad one at that and is now emerging as nature via nurture (Hay, 1999). It is a great pity yet also a great example of how human behaviour has been so troubled with misconstrued genetic heritage debates. Right questions wrong methodology vs. versus wrong questions right methodology! Will the two ever converge?

¹⁴³ It just so happens that *g* is in fact a better predictor of pilot and navigator training than lower-order factors (Miller, 1990). But hopefully the principle of what was stated above has been noted!

¹⁴⁴ Although reference is often made to models of dynamic assessment as theories of intelligence (Rothman & Semmel, 1990).

¹⁴⁵ It is not the author’s intention to be pessimistic about the field of intelligence research (otherwise a thesis would not be written) but the contention is that the wrong questions are being asked of the field. Research simply cannot deliver on the questions being asked. Perhaps they are being formulated incorrectly for it is surmised that in ten years, one hundred years time or even longer, if we continue to ask these questions in the manner in which we currently ask them, the development of research findings in this area is going to continuously frustrate researchers. This comes back to the general frustration with the current understanding of what it means to study psychology - are we talking about mental or physical? Hence the need to discuss this aspect in chapter 2, which is also elaborated upon in chapter 3 and 4. It seems that psychology has forever been involved in some sort of a crises. Is it not about time to move on and try to resolve the crisis by splitting the hitherto unmanageable domain of “psychology” (whatever that means). Intelligence research as well as dynamic assessment will no doubt be the better for it.

A related aside. The constraints of our architecture within computation and its similarities to the constraints of our prevailing understandings within intelligence research

Attendant to this question of the methodology/research terrain question: looking at the status of artificial intelligence research where strides continue to be made yet are severely hampered by the tools of the trade. Information, at the heart of computation, is still organised and conceptualised in a discrete manner befitting a mechanistic framework of information storage. Scientists utilise this current underlying architectural foundation as point of departure for just about everything else in the field. So at the outset one is constrained by the system in which operations take place. The problem of modelling the human brain (or any other brain for that matter) is not speed as much as it is storage capacity but also how information is manipulated – the brain is not a binary operated machine. The tools of the trade are not in keeping with the entity under discussion because it is the very entity that we cannot yet explain so how are the tools to keep pace? Intelligence assessment tools cannot keep pace with the substantive entity known as intelligence functioning in reality. Can we perhaps move out of the architecture and into completely new architecture which will thus inform the tools of the trade? No matter how sophisticated the tools of computation they are premised on a narrow construal of what it means to store and manipulate data; something is either on or off, a one or a zero. Coming down to this micro level it becomes starkly evident that our underpinnings are extremely unsuited to the types of modelling with which we seek to engage. What is the answer? And how would such a revised foundation function and what would be its characteristics? All the more reason to computationally model brain functioning from a biological perspective (and in keeping with reductionist trends, a physical description would be better, yet too narrow). Perhaps our thinking in this matter is itself too narrowly defined.

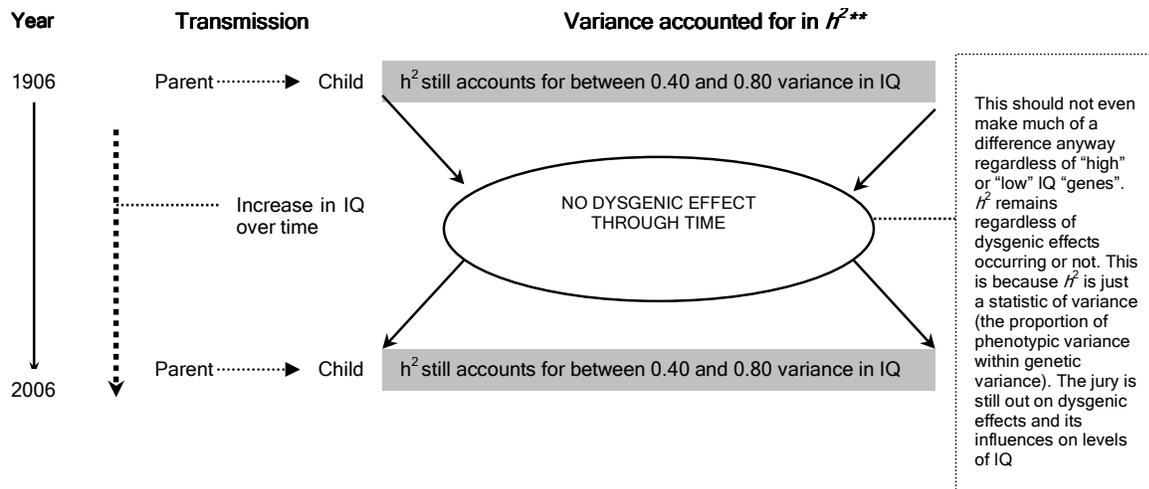
However, the physiological psychology literature is still saturated with research findings on these questions. Dynamic assessment has stood steadfast along the lines of continuous change within the individual which reflects back on the developmental issues above. It is assumed that although childhood change is rapid and stage-like, continuous change reigns throughout life, otherwise what would be the point of dynamic assessment's wonderings into adult and geriatric assessment? Similar environmental surroundings, so the logic proceeds, should result in similar outgrowths of intelligence development but the constraining effect of genetic heritage may play against this assumption. Which of the two determines the course of events? Are we asking an impossible question seeking an impossible experiment? This thorny and highly contentious debated issue is precisely so because of the melding of physical with psychological in which physical explanations are sought for psychological phenomena: the crisis in psychology and a theme which recounts its own unending and repetitive cyclical nature. As Humphrey's (1994) states, the task of psychology is after all to explain behaviour not cognition per se; all the more reason to question the sub-discipline's status. Dynamic assessment rests upon a change-based philosophy of intellect and the literature is very much divided on the question as to the differential and changing effects of both genetics and environment on the developing individual throughout life; does genetic contribution wear off or increase over time? There is evidence to cite the relevance of both (Hay, 1999) and the source of much of these results emanates from twin studies (Brody, 1992; Charles, 1973; Horn, 1998a; Reznick, 1997; Reznick & Corley, 1999; Rodgers & Rowe, 1987; Stankov, 2004;) and family constellation set-ups (Galbraith, 1983; Jensen, 1998b, Locurto, 1990; Loehlin, Horn & Willerman, 1997; Scarr, 1998a; Valencia, Henderson & Rankin, 1985; Wilson & Matheny, 1983). Yet substantial research has posited the case for heritability estimates to increase over time with a concomitant decrease in shared environmental influences¹⁴⁶ which also encompasses community influences on cognition which is mediated by parental genotypes (Chipuer, Rovine & Plomin, 1990; Coon, Carey & Fulker, 1992; Jensen, 1997; Loehlin, Horn & Willerman, 1997; Petrill, 2005; Plomin, DeFries, McClearn & McGuffin, 2001). The crux of the argument here is the role of variance accounted for through genetic predictors and not the role of genetics per se, and paradoxically, as age increases and fluid scores of *g* decrease the cultural environment comes into its own as supportive aid (Lövdén & Lindenberger, 2005). The case is indeed very strongly in favour of intelligence being resultant from hereditary more so than on environment. This is particularly the case of high-risk environments where genotype-environment interaction highlights the role of heritability in intelligence estimates (Asbury, Wachs & Plomin, 2005) especially tests which load higher on *g* (such as verbal tests; Gignac, 2006) which generally account for higher heritability variance (Wainwright, Wright, Luciano, Geffen & Martin, 2005). One need only peruse the literature in this regard to agree with this statement (Fletcher, 1991). Moreover, and something which is continuously reiterated in this study, is the recognition of the effect of genetic heritage on socialisation theories, which is mentioned far less in comparison to socialisation theory and its effects on behavioural genetics (Scarr, 1997, 1998a). Yet one cannot ignore the data which attests to the well-known Flynn effect and how mere exposure to technology and education results in elevated levels of IQ or ΔIQ (change in IQ per decade; Jensen, 1998b) (Blair, Gamson, Thorne & Baker, 2005; Flynn, 1998). It

¹⁴⁶ The proportion of variance accounted for by genetic traits when compared to the proportion of variance accounted for by the rest of the criteria. This does not mean that "intelligence is inherited to the degree of 50%" rather, intelligence test results and their associated variance is accounted for by the variance within the genetic traits variance to a ceiling of usually 50% accounted-for variance. It is imperative that these two sentiments are not equated (which they often are, at least tacitly within some literature).



must be recalled that IQ tests assess for those aspects commonly found in technological and educational areas! One should not confuse this issue. Figure 17 puzzles over increasing IQ scores as this has philosophical implications as far as this study is concerned in terms of construct validity for both intelligence and dynamic assessment.

Figure 17 The puzzling increase in IQ and the resultant implications for construct validation



** recall that h^2 is a quotient stating variance accounted for by generational contribution which is genetically transmitted but genes are resultant from evolutionary genotypical-phenotypical influences; so the environment did indeed have very much to do with selection of genes within populations. One can almost declare the irony inherent in the “genes only” argument. The opposite of dysgenic effects would be eugenic effects where the traits are indeed carried through successive generations (Jensen, 1998b), hence the once popular eugenics movement

h^2 still accounts for the same variation in IQ regardless of the IQ score itself. Something is pushing up IQ and it cannot be concluded that more IQ-savvy individuals are being born due to choosy mating strategies (as there are fewer high IQ parents today in comparison to low IQ parents). Hence, genetic influences are not pushing up IQ levels. The nature of the tests has remained the same. Therefore:

1. same h^2
2. same test
3. different IQ

what has changed in the interim?

4. the environmental influences perhaps (cognitive, technological, nutritive and so on)
5. therefore, IQ is in part dependent on environmental influences
6. but IQ is supposedly a pure measure of a theoretically stable construct unadulterated by environmental conditions
7. so we had better re-look our
 - a. notions of theoretical and empirical constructs
 - b. ideas of immutable IQ measures

however

8. twists and turns pervade the argument and include the following
 - a. variance in gains in test-retest scores are not accounted for by g (for the sake of the argument it is assumed that g = IQ for the moment)
 - b. the increases in IQ are accounted for mostly by the lower half of the bell curve where decreasing variance has been evidenced
 - c. dynamic assessment predicates are founded on manifest change brought about by enhanced environments (via mediation predominantly) and usually although not exclusively concentrate on those in the lower half of the bell curve
 - i. dynamic assessment should thus affect this section of the curve more so than the right hand side (which it does)
 - ii. dynamic assessment predominantly intervenes at the level of the environment
 - iii. could it be that dynamic assessment results too have increased over time due to the same factors resulting in higher IQ scores? Due to sub-test similarity it should but due to reliance on mediation as variable this relation may be difficult to detect
 - iv. as is argued elsewhere in this study; IQ simultaneously does and does not equal learning potential
 - v. if dynamic assessment's concern with environmental aid is strong; then there should be concomitant increases in scores on dynamic assessment batteries too (hypothetically testing immigrants in 2006 and 1906 via dynamic assessment may well evidence gains)
 - d. once environments are homogeneous in terms of exposure dynamic assessment will likely fall away as the only variance yet to be accounted for will be h^2 (if the logic above is followed and applied here too)
 - e. it can also be argued that IQ and dynamic assessment methods are measuring the same environmental influences; added to this is the one common test used in both static and dynamic assessment - the Raven as best current indicator of fluid g (IQ)
9. IQ's have increased but h^2 has remained stable so it can also be argued that IQ does still measure a stable construct after all; a shift to the right for the entire curve (and it possibly is not so, merely a decreased variance in the lower end as already mentioned) does not logically imply that IQ is not assessing for a stable trait (g). Is the variance decreasing for those in the right hand side as well?



(iii) *General and specific intelligence* conceptualisation is perhaps the most contentious issue (McArdle & Woodcock, 1998) for which a partial solution has been advocated above in chapter 2 and is developed in chapter 3. The discipline of psychology should split into various streams as competing methodologies and philosophies are incompatible and have as a result only confusion and dissent. This is not good science practice. Physiological research (minimal cognitive architecture and neural speed research which is not knowledge-dependent and hence cannot serve as a definition for intelligence but may well indicate a correlation with it; Guttman & Levy, 1991) should continue unabated (however unpalatable to the more systems-oriented researchers who prefer to witness development and change as emergent). Notwithstanding the advantages of just such a tradition, systems orientated models of change and development need to continue within their own stream. Emergent properties are by their nature inherently unpredictable,¹⁴⁷ Three main views which have permeated the disciplines' writings concerning intelligence are monarchic, oligarchic and anarchic models (Sternberg, 1998). A single global monarchic form of intelligence is most likely correct in formulation but is inadequate when practically explored. Oligarchic concepts view intelligence as consisting of broad factors (which is also most likely correct as it really only is a sub-set of the former with most of the published test batteries available addressing fewer than a hundred primary dimensions; Horn, 1998a). Lastly, anarchic models, which view intelligence as incredibly multifaceted (also correct to a point but very unwieldy and so specific that one tends to get lost in the morass of abilities which seemingly goes against the tide of parsimony within science research - see chapter 3). Scarr (1998a) emphasises the role of theory within socialisation and behavioural genetics research where competing theories cannot be said to predict and control for everything as it would thus not be testable nor could it be inconsistent with itself (Styles, 1999; Torff & Gardner, 1999). Of course the very same critique can also be levelled at a grand framework which seeks to encompass everything in life! A bio-psycho-social-genetic-environmental model is equally as absurd in detailed reality but less so in conception. The debate surrounding *g* is infinite and links up to the notion of construct validity as discussed in chapters 3 and 4. Dismissing the construct debate and forging ahead with measurement technique is a rather pointless activity yet it seems to consume much of measurement psychologists' time. Horn (1998a) states with regard to *g* that it has yet to meet three pivotal requirements for a suitable scientific notion:

- The requirement of a similar factor - that is *g* should be computable across different groups
- The requirement of a unifying principle - that is how lower order factors tie back into *g* (and for which *Gf-Gc* theory posits no unifying *g*; Wilhelm, 2005)
- The requirement that *g* should relate to other variables in a consistent fashion (here we see the construct validation argument)

But a theory of *g* can also attest to the following (Bowman, Markham & Roberts, 2002):

- The existence of a positive manifold - intelligence tests correlate in a lawful fashion as first evidenced by co-variation noted by Spearman (Anderson, 2005b; Neisser, 1998); high scorers on verbal tests similarly score highly on spatial and numerical tests (Borsboom & Dolan, 2006). *G*-based theories are the single most persistent theories in the intelligence literature (Blair, 2006; Morgan, 1996) and one has to consider this itself an indication of its validity in some form or another; after all, general ability "represents a broad construct that underlies non-specific information-processing efficacy" (Ackerman, 1988, p.290). Whether or not such positive manifold can be equated with psychometric *g* (recall that it is itself a statistical concept) is debateable (Heitz, Unsworth & Engle, 2005).
- The stability of *g* across test batteries - regardless of factor rotation, the underlying constructs remain; although one can argue that factor analytic approaches are merely one type of statistic which require support from structural and nonfactor analytic evidence (Flanagan & McGrew, 1998). Also most tests would correlate as many are inevitably constructed within one reigning paradigm
- *G*'s real-world utility value is great (this is clear even within layperson interpretations and understandings of intelligence; see figure 11 above). Regarding the utility value of intelligence tests per se, studies have evidenced that of the top ten utility value ranked standardised tests within the educational counselling domain, four were intelligence batteries (Esters & Ittenbach, 1999) which most likely alludes somewhat to the generality of *g* given its statistical appearance across tests
- *G* evidences meaningfully empirical correlates (clearly a pro-Jensen view¹⁴⁸ which is contested by many researchers)

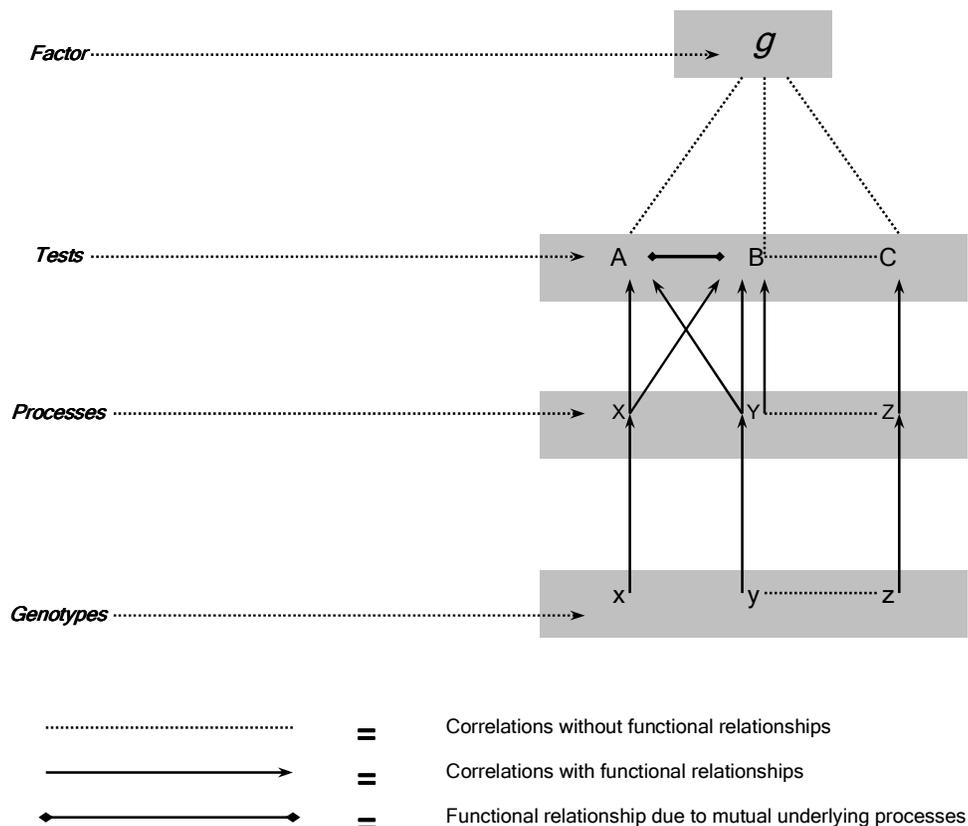
The counter-arguments against the above are detailed in critique and offer tantalising arguments to the contrary nature of many *g*-based claims. *Gf-Gc* theory and models derivative of this theory attest to *g*'s somewhat partial existence yet also affirm its fluid nature when attempts are made to capture it. Perhaps one may refer to *g* as evasive, similar to visual illusions where concentric circles appear to move but upon looking directly at them they appear stationary.

¹⁴⁷ The author can imagine the backlash of criticism to this statement.

¹⁴⁸ Arthur Jensen. Mogens Jensen, a firm advocate of dynamic assessment should not confused with Arthur Jensen.

Recall that g is a positive manifold psychometric construct (statistical, think Binet) not a biological one (not completely validated as one just yet and not proven as a substantive theory, see chapter 4 for more on the difference between substantive and hypothetical constructs; think Galton) (Bowman, Markham & Roberts, 2002; Detterman, 1994b; Hunt, 1998). Jensen and others contest this (1998; Brand, Egan & Deary, 1994; Deary & Smith, 2004; Demetriou & Papadopoulos, 2004; Li, 1996) by favouring g 's biological nature (Baron, 1985) with suggestions that the prefrontal cortex is very important in g -related tasks, although not exclusively so (Duncan, 2005). However, Jensen has in the past maintained g 's non-reified nature (1969). The issue surrounding g ¹⁴⁹ as both a statistical and psychological construct is contentious (Lohman, 1996) after all "how do we proceed to endow g with meaning?" (Borkowski & Maxwell, 1985, p.221). Jensen's (1982b) depiction of the relations between g and tests, processes and genotypes in the paradigm of reaction-time research is illustrated in figure 18 below. Recall that genotype is a result of a continuous process of modification brought on by phenotype-genotype interaction and mutual influence which is not shown in this figure. Reaction-time research (RT) evidences differences in all populations including sub-normal populations which yield differential reaction times within groups so that correlations between RT and g are commensurate with level of intellectual functioning. Jensen (1982b) is at pains to emphasise that RT is not necessarily a component of g but that it relates to it is evident. RT is able to discriminate between low and high IQ populations which attests to its functionality and utility. Learning potential too is able to discriminate within a narrowly defined range as evidenced in the early studies conducted by Milton Budoff and his colleagues in the 1960-1970's with educable mentally retarded and educable educationally retarded individuals (Budoff, Meskin & Harrison, 1971). Both these groupings were originally considered one group of uneducable individuals. Their work was also concerned with biopsychological substrates which were far from irrelevant in their furnishings of additional information into interactions between other levels of descriptors within cognition (Snow & Lohman, 1984).

Figure 18 Jensen's (1982b, p.270) depiction of g 's relations to tests, processes and genotypes in the reaction-time paradigm



This typifies the swing of the pendulum within constructions of intelligence; "the psychometrics of intelligence tests and the psychology of human intelligence" (Lohman, 1997a, p.372). Here it can be seen that the struggle between different approaches towards one area can lead to vastly different conceptions; for within the structural understanding of intelligence "much is known

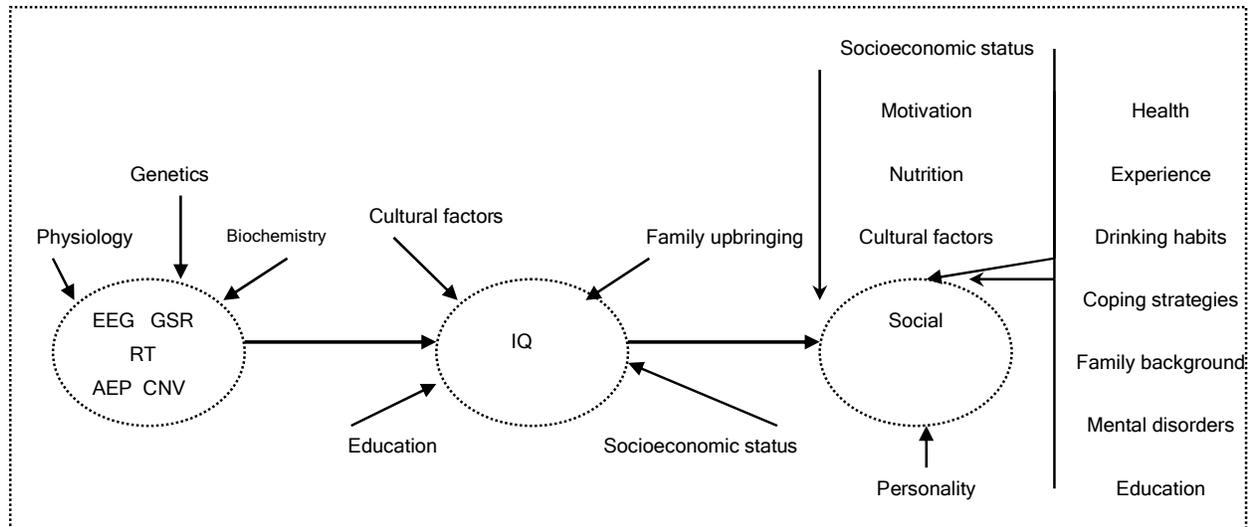
¹⁴⁹ G arises from correlations, pure and simple. This implies its statistical rendering. However, it seems to arise more often than not thus indicating (not stating, merely indicating) its substantive nature.

about the taxonomy and predictive validity of human intelligence differences, [but] there has been relatively little progress in understanding their nature” (Deary, 2001, p.127). *G* is perhaps assessed during some instances and not others although logic would dictate that a governing or underlying pervasive structure such as *g* should indeed be manifest at any level. The degree to which modules in the brain can function without the need to inform themselves of other modules (multiple intelligences or modular theory; Gardner, Hatch & Torff, 1997) (“relatively autonomous intellectual competencies”; Krechevsky & Gardner, 1994, p.287; Wellman & Gelman, 1992) is no justification for the fact that *g* operates solely in isolation from other areas in the brain. What we cannot see is the nature of the underlying mechanism at work within structures in the brain and behaviour although evidence suggests that both neocortical and the more evolutionary primitive areas of the brain developed specialist localised interconnected sets providing support for modular-like intelligences (Barton, 1997). Looking at the issue from a biological or psychometric perspective colours the very conclusions that are drawn. Hence, say some, the need to revisit the issue from a systems perspective where the global emergent nature of the set-up is assessed. This brings one back to the case for reductionism and hence the need to branch off in different directions in order to piece together the fabric of brain functioning and behaviour. The author will not belabour this point any further, but suffice it to say that *g* most likely exists in some form (Detterman, 1998a) (perhaps not psychometrically feasible in some instances but this is a methodological fault on the part of human researchers and not evolutionary processes that are unwilling to yield their secrets!). Butterfield (1998) betrays his allegiance to a *g* form of intelligence when he succinctly states that all intelligent systems learn, but more intelligent systems learn more yet he does emphasise the ability to learn (including metacognition) as crucial in locating for intelligence some adequate defining criterion. This brings us firmly into the terrain of levels of description.

2.8.4.1 Levels of description

Explaining levels or hierarchies of description (Gleitman, 1985) are necessary if researchers (those trained within the behavioural and natural science spheres) are to adequately account for or at least offer tentative explanatory models of behavioural and testing outcomes. This is so because each level necessitates level-specific theories and assumptions based on physical and structural functioning of the brain and behaviour (Changeux & Dehaene, 1994; Changeux & Konishi, 1987; Chipman, 1986; Davis & Anderson, 1999; Hunt, 1998). Hunt (1992) refers to three levels regarding theories of thought; namely, the physiological (brain-based), architectural (the way in which information is processed) and knowledge-representational (semantic concepts) levels. Champions of such approaches have long advocated the need for more integrated analysis of the system (Churchland, 1986; Churchland & Sejnowski, 1996; Neufield, 2002) and although not embarking on a dynamic systems analysis of what it means to be human, a favoured approach within this dissertation is to adapt both a view from above and below (high and low levels). After all, each level of analysis offers its own insights into the workings of a domain of interest or research (Stillings, Weisler, Cahse, Feinstein, Garfield & Rissland, 1995) with the subsequent recognition that no single level reveals the whole story (Klvington, 1986). However, the focus within any one discipline for obvious reasons neglects to study the detail of another, which is one such method of globalised science practice. Ignoring any view is acknowledged to be dangerous (Thelen, 1994) and as such inclusive co-investigatory procedures should be undertaken within the study of intelligence and specifically dynamic assessment interventions, as it is so often the favoured technique utilised within sub-populations of patients requiring specific treatments as discussed above. No anxiety should accompany any specific research agenda housed within its own tradition, for as Neisser (1997) states “we need not fall prey to post-modern cynicism, and we are in no danger of being eliminated by our neuroscientific friends” (p.258). Eysenck’s (1986) model is similar to the notions of cross-boundary disciplines aiming towards a fuller more inclusive understanding of intelligence and is depicted in figure 19 below.

Figure 19 Eysencks' (1986) model of the relation between biological, IQ (psychometric) and social intelligence



Neurophysiological, anatomical and behavioural studies of the human brain at once permit and firmly bar the study of the organ at various levels, ensuring that the lone researcher will not quite be able to grasp the full functioning organ at one level only (Gardner, 1998). Hence the need for multiple-angled studies stressing multiple points of investigation into one phenomenon at various points in time (Chugani, Phelps & Mazziotta, 1994; Huttenlocher, 1994; Nowakowski, 1994; Oyama, 1994). Epigenetic growth within the individual comprises so multitudinous a network of reactive genetic responses to impinging environmental stimuli (or lack thereof) that selecting matched age cohorts within studies designed to test for any factor, for instance, is immediately prejudiced against a host of cross-pollinating variable reactions which occur at differing times for different people. Reducing to a singular notion a genetic complement within human beings and ascertaining the nature of genotype will not readily reflect anything of use as epigenetic growth is by its nature phenotypical expression based on mutual inclusion of genotype and phenotype (Ceci & Bruck, 1994; Jensen, 1997) which are themselves governed by the environment¹⁵⁰ (Hurtford, Joseph, Kirby & Reid, 1997). Boggled down in such a mire of information necessitates the narrowing of focus within studies purporting to study the "intelligence" or "improvement" of subjects over time (pre vs. posttest designs) and space (in various cultural settings). The way out of just such a doomed scenario would be to isolate methodologies and goals, but to what cost? Anderson's (1992) view of intelligence is that both low and high level accounts are really working towards the same construct but are doing so in markedly different ways.

Psychometric and laboratory-based approaches towards assessment are often seen to be **low-level** explanatory systems which are concerned with three dominating aspects:

- The increase of cognitive abilities with increasing age
- The stability of individual differences throughout development
- The co-variation of cognitive abilities on a broad range of psychometric tests
- In sum: low level approaches are concerned with intelligence as biologically (genetically) determined attributes of the nervous system (Anderson, 1992, p.2)

Dynamic assessment models, notably Feuersteinian-based ones interface well with bio-ecological models from intelligence research and straddle the link here

High level approaches towards the understanding of intelligence foster the notions of:

- Specific abilities in addition to a general ability
- The universal cognitive mechanisms which are characteristic of everyone
- In sum: high level approaches are concerned with intelligence as governed by culture and is an experimentally driven attribute of cognitive functions (Anderson, 1992, p.2)



Dynamic assessment focusing on improvement over time does so from a behaviourist perspective, a psychological angle lacking in perhaps a more detailed analysis of the amount and type of change undergone at a more neuroanatomical level. For instance, although far more evident and pertinent at very young ages, stages of brain maturation will simply not allow for certain expected resultant behaviours from interventions. This is due to the maturation of the brain at any one particular stage and its allied neuroanatomical development (Diamond, 1994). Who or what is to say that delayed development or inadequate stimulation at critical periods counter efforts to increase scores on psychometric tests? Interventions at age six might only be reciprocated by increased scores at a later age, even though age-relevant tasks are mediated for this age group. This being wrongly interpreted however, might lead unquestioning investigators to assume that intervention was a failure due to task-intrinsic factors or that the model of development and so forth. The reason will not necessarily be found on a behavioural level but well within the neural level.

Regardless of the explanatory level reason offered, constraining and/or inclusive experiential stimuli will also play an overwhelmingly large role in determining (not always dictating) the next step in the progress of the individual, whether behaviourally or neuronally (Johnson, 1994a). As such, this too has to be included in the conceptual framework when considering the what/when and how of interventions within intelligence settings. Maturation (think Vygotsky's learning) and experience (think Vygotsky's instruction) are co-evolutionary (Churchland & Sejnowski, 1996) and there is no one-to-one linear relation between the two (Greenough, Black & Wallace, 1994) displaying a continuum of cause and effect between maturation and environment. One need think only of the interaction and difficulty of pin-pointing effects of education on IQ and IQ on education. Does having a higher IQ elicit education-seeking behaviour or does increased education access stimulate higher IQ? (Retherford & Sewell, 1989). Gene expression can be altered by exposure to environmental stimuli (Kennedy & Dehay, 1997) and gene expression can result in certain experiences being sought; for instance over-production of neurons results in their eventual pruning and thus allows for more practicable allowance of neuronal density in the developing brain (Huttenlocher, 1994; Nowakowski, 1994). This blending of levels should take place when answers are not forthcoming from within other allied levels.

Questions at one level will ineluctably lead to queries bombarded at higher and lower levels within the same area of description (Churchland, 1986) notwithstanding the now-tiring debate between radical materialism, epiphenomenalism and dualistic interactionism as to the constituents of mind/brain (Eccles, 1989). How models are built determine in part their explanatory power, as such, functionalist models will assume linked functions of the various tiers or hierarchies of study as mentioned above¹⁵¹ and structuralist approaches assume that the way in which a system is built yields different information. To allow one a seat at the table of viewpoints does not necessitate adherence to any one particular approach to the study of intelligence for instance. Casting a conceptual net over any viewpoint will result in being cut off from other fertile areas of pursuit. The question raised is whether it really is a good idea to forego old notions of narrowly focused conceptual fields of enquiry when the full landscape of possibilities is open to use by broadly focused areas. Bridging this gap will allow more confluence within seemingly disparate areas of interest (when in fact these "disparate" areas are at times quite closely aligned). Melding two or more areas of focus (the level of analysis) can only be more extensive as opposed to not allowing some sort of coalescence,¹⁵² Intelligence assessed from psychometric and structural approaches allows for aspects such as attention and memory (Messick, 1996) to be viewed as correlatory to a general level of intellectual functioning and is perhaps conceptually easier to study from this vantage point than perhaps a biological point.

Both micro and macro systems, as illustrated by the neuronal and behavioural levels of explanation respectively offer glimpses into functioning which simultaneously also limits our understanding of the complete unit. It would appear that a four dimensional analogy would suffice to elaborate on the predicament in which we find ourselves, being unable to explain events at one level without sacrificing the modes of explanations at another level. In order to visualise a hypercube in our three dimensional existence, we are forced to visualise its three dimensional shadow, a somewhat frustrating experience, allowing access to what it may look like and even allowing us the luxury of pursuing it's shape within the mathematical realm (Rucker, 1986), but never affording us the opportunity to probe it closer. So it is with the various levels of description which are necessary to invoke as alternative manners of gaining access to how the brain functions, but not allowing us the similar luxury of being presented with the full picture. It lies tantalisingly close within our grasp but our meta-theoretical framework seems lacking in a fundamental ingredient. An overarching conception? What would it look like? Without pseudo-characterising brain studies as quantum systems,¹⁵³ the situation is very much akin to only being allowed to know the momentum or position of an electron at any one time, but never both! Is there not perhaps a way in which the system can yield all information at various levels without ceding

¹⁵¹ Even though these conceptions are merely to make life easier for the researcher, it is not evident that nature has in fact arranged matters so that we may view them through functionalist/structuralist frameworks. Our shaping of the framework and continual reshaping allows us to pursue the same questions in different contexts as well as ask different questions within one context. Whether this is good or bad is not the issue as it is unlikely that it can be construed as good or bad as this is value-neutral territory.

¹⁵² To each his own. At our peril! Remaining focused but forsaking the need to correlate or map onto other levels of description is a dead-end.

¹⁵³ An oft-quoted critique levelled at Penrose (1999), in which it is speculated that cells "somewhere deep in the brain" (p.517) are assumed to function as quantum systems. However, it does seem as if Penrose is merely utilising a provocative thought experiment to illustrate his point.



necessary information? This endeavour resonates with intelligence research within dynamic assessment, as change is evident but the level at which it occurs is difficult to ascertain. Once the change has occurred it is nefariously difficult to trace its origins. The experiential and/or native domain is offered as arenas in which change is orchestrated within the human. How sure can researchers be that change is in fact induced via environmental stimuli and not as a result of an epigenetic flowering? (Karmiloff-Smith, 1994). This selectionist view, which accommodates both the preformist (enlarging of an already nativist or formed organ/cell) and empiricist views (the blank slate or constructivist view) is ever-more sanctioned by both neuroanatomists and psychologists alike (Karmiloff-Smith, 1994; Nowakowski, 1994). Although subject-area specific jargon is used to describe similar concepts (empiricist vs. constructivist), the re-directed attempt in understanding the full system posits a more inclusive methodology towards the study of how the brain adapts, functions and changes over time. Dynamic assessment's emphasis on process as opposed to attainment (same score different paths getting there) is thus perhaps closer to a more modern understanding of intelligence assessment after all (Glaser, 1998; Snow, 1998). Yet process analysis is still much poorer in information than the traditional conceptualisation of intelligence as content-based (Humphreys, 1998) with current static-based intelligence assessments making headway in allowing for the assessment of underlying cognitive processes rather than abilities (Harrison, Flanagan & Genshaft, 1997).

This is the main point within this section, namely that "change" evident at the behavioural level need not be evident at any other level, hence undermining the efforts in inducing such change. The permanency of any such change needs to be investigated at other levels beyond the macro behavioural. Perhaps it is helpful to designate levels of description within this pre/post test set-up so as to materialise efforts at distinguishing where in fact the so-called changes are taking place, at what age, at which location in the brain and how manifest this change is. Not all manifest change is however permanent and neglected change at the micro level may be of more use at certain critical points in development as well as within assessments aimed at dynamically testing people across domains. Figure 45 (chapter 3) illustrates Madsen's (1988) conception of a similar stance taken on the relation between model and experimental object (the brain) and fits well within the Churchland and Sejnowski (1996) depiction of the understanding of physical sub-systems within the larger system of the brain.

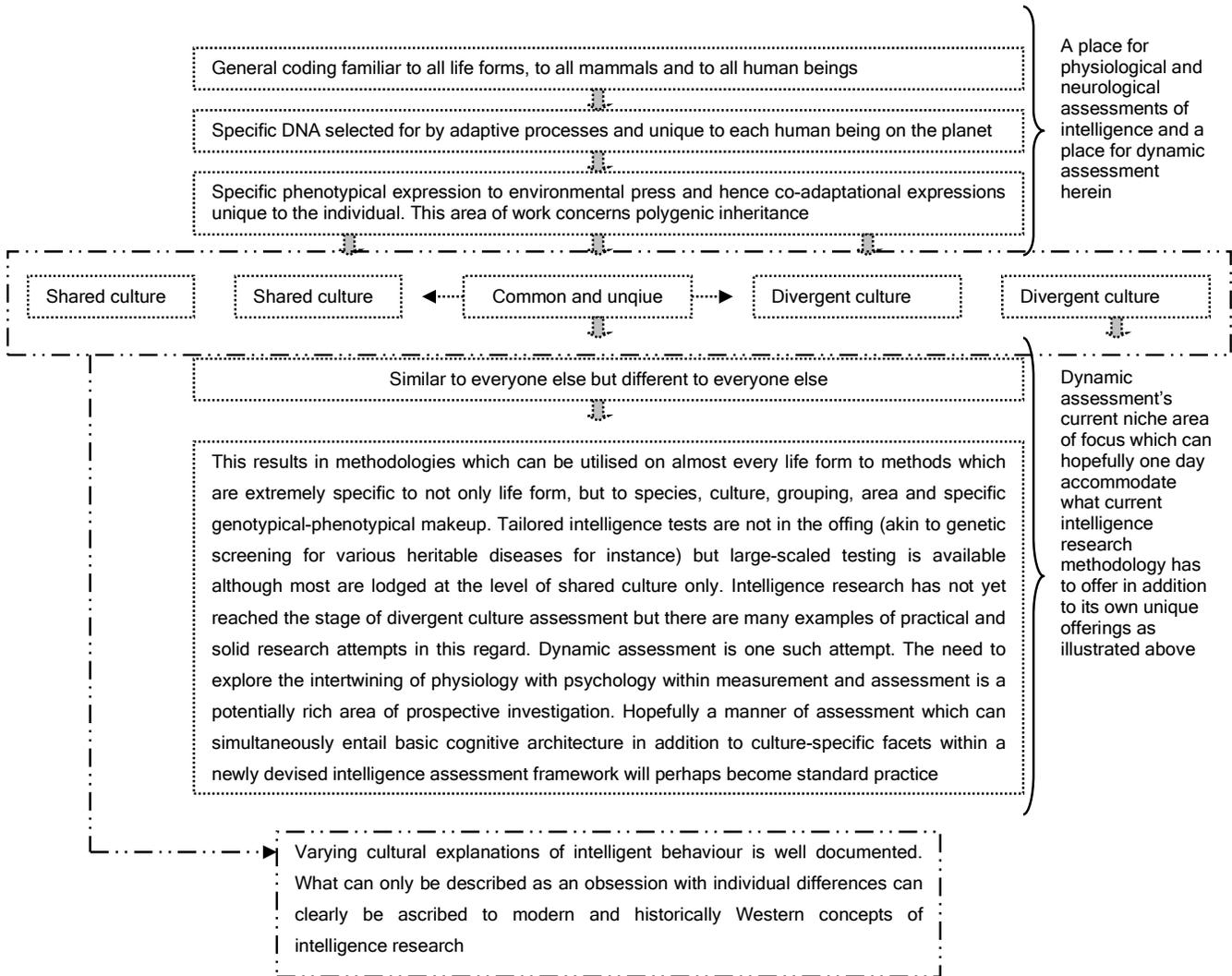
(iv) *The issue of culture* which is prevalent in dynamic assessment research comes to the fore but on a scale further down the continuum. The need for some to place emphasis on intelligence and the manner in which this is done differs radically from culture to culture as well as the emphasis placed upon individual and group differences (Davis & Anderson, 1999). Intelligence can be considered so narrow a feature of existence, that pigeon-holing it into specific contexts is about as accurate as one can get states Berry (1998) whose cultural concerns are pervasive when attempting the study of intelligence. Intelligence is appreciated only as it pertains to the culturally understood meaning of what it means to act intelligently and how society is likely to treat people (Detterman, 1998b; Pellegrino, 1998). Such an argument can only go so far in today's world where the global arena is pervasive and will most likely result in even greater homogenisation. Human beings can be simultaneously viewed as being enclosed in and enclosing various layers of social and individual functioning. This can be viewed as follows in figure 20 below. If the aim of dynamic assessment is to make available a context for the opportunity to improve on current level of functioning so as to address resident yet non-manifest performance, can one state then, that it's aim is to improve performance? If this is the case, then it is not a short leap between performance in general and performance on an intelligence test for instance. This smacks of the rallying call for attempts to raise intelligence (Spitz, 1999) which, if alignment with a biological perspective is taken becomes more untenable as a research effort. Unless, of course, biological predisposition can be moulded of which phenotypical consequences bespeak. This is a particular instance of where we might perhaps be asking the wrong questions or utilising the incorrect methodology for the question. It is unlikely that *g* is malleable but its various tendrils including metacognitive reserve are perhaps malleable. If *g* is indeed not malleable where does this leave intervention programmes included among them of course, dynamic assessment and its attendant mediation strategies? What exactly is being mediated? Cognitive skill, metacognition and behaviour which are then construed as non-*g* factors (because *g* is by default unalterable just as DNA coding is unalterable¹⁵⁴) in which case the effort is useless (far transfer usually does not result for instance but near transfer is possible, as often cited).

The magnitude of the transfer of skills learned as a result of mediated learning within global dynamic assessment depends on the nature of what is being mediated; metacognitive information may transfer further than specific skills such as spatially demanding skills (Snow, 1998a). Closely aligned to transfer is the related concept of maintenance; the degree to which information is retained and applied in novel instances but is itself ascribed to intelligence (Campione, Brown, Ferrara, Jones & Steinberg, 1985). In other words, the higher the initial estimate of intelligence the greater the likelihood of far transfer and increased maintenance which brings researchers back to the point of origin in terms of intelligence and potential. The point is that depending on how the nature of intelligence is construed the nature of mediatory intervention should be tailored accordingly but due to the uncertainty still prevalent within intelligence in psychology, leaps forward will most likely be hindered by undeveloped understandings of what and how mediation works. Perhaps the reason as to why far transfer so rarely occurs is

¹⁵⁴ For now the author will ignore recent research attempts to molecularly alter genes.

because the underlying mediated or remediated trait is not altered which if argued back again means that the target of dynamic assessment interventions are targeting *g*-type skills. However, this does not close the gates for remediation and mediation research because interventions can at least ameliorate the effects of low performance due to any number of reasons (Anderson, 1999b; Horn, 1998a). The question is how best to go about doing this. Cattell (1998) cites research which indicates that fluid intelligence suffers as a result of damage to the brain regardless of the area of injury whereas crystallised intelligence is more resistant to such change, which loops back to the former statement of mediation affecting non-cognitive or metacognitive skill.

Figure 20 From the specificities to the generalities of intelligence assessment - where dynamic assessment can invent its place





A number of evidential criteria as to why vertical and horizontal depictions of intelligence are correct are offered and include results from brain damaged individuals which attest to the sole functioning of areas of the brain which are then able to accommodate for the loss (Nettelbeck, 1999). Research results emanating from the loss of functioning in the brain as well as from mental retardates has proven a fertile area of knowledge gathering within intelligence research (Horn, 1998a). The fact that savants are able to perform the feats they can; the research emanating from experimental psychological tasks; the support from psychometric measures; a history of development which can identify end-state performances; the reflection of evolutionary processes and their resultant consequences; operational sets of information processing activities and the propensity for the symbolic encoding of information all manifest in identifiable separate modules or intelligences (Reznick & Gardner, 1999). When conceptualising for global intelligence research and multiple forms of intelligence, along with other modular theories and g -theories, would it not seem possible to state that all these models are merely snapshots taken at different angles of the very same construct? Fluid (cognitive mechanics) and crystallized notions (cognitive pragmatics) of intelligence, g and s versions of intelligence, socio-environmental and behavioral genetic considerations and much else besides inflames the agenda with facets which seem almost to burst at the seams (Li & Kunzmann, 2004). A finite number of sets of intelligences can at least constrain the system in which all sets work together in a complex system (Detterman, 1998a) which, as with Gardner's multiple intelligences brings us back to omnipresent g (Li, 1996). None of the models are wholly incorrect and none are fully inclusive and encompassing of all there is to know, but one gets the feeling that that much research within these so-called opposing traditions are in fact looking at the same construct from different views. Each has much to offer but the attempts to glue together this tapestry is where the difficulty arises.¹⁵⁵ The concept of Anderson's (1999) minimal architecture infused with growth of developmental modules (akin to Fodor's modularity of mind which does not account for development of these modules) is an instance of an attempt to assemble various views on intelligence. Increasing development of modularization is a manner of side-stepping this issue. The fact that g underlies supposed modular functioning is contested as modularity presupposes distinct functioning capacities and performances (Nettelbeck, 1999) but if the view of g is taken to be one of substantial underlying foundation then this conception of modularity has to be revised. Not all approaches are however greeted with their original degree of enthusiasm as many early notions of information processing for instance have long since outlived their utility (Reed, 1997).

2.8.5 Yesteryear and today – how far have we come?

The journal of educational psychology convened a panel of intelligence research experts in 1921 to collate various opinions regarding the nature and future of intelligence and its research endeavours; entitled "intelligence and its measurement". One will read with humour and humility the aspects highlighted by these venerable scholars of over 85 years ago for much of what they say has hardly changed in the interim yet there have been staggering improvements on a number of fronts, not least concerning physiological understandings of brain functioning. Embretson's (2004) comments on the future trajectory of research within this domain echo similar sentiments and places particular emphasis on the fact that the future bodes well for the further refinement of techniques and concepts developed during the latter half of the twentieth century.

"The value of a test score is its value in prophesying how well a person will do in their intellectual tasks" (Thorndike, 1921, p.125). Given the specific weighted task at hand, states Thorndike, will allow for better detailing of a person's functioning on that task but to search for perfect correlations between specificities and to relate this back to a general level of intelligence is almost impossible unless each separate task is assessed and as we know there are too many sub-tasks to assess. This very same issue still poses problems in the defining of intelligence today (Baron, 1998). Just as there were serious misgivings as to a general factor in 1921 so too are their similar misgivings decades later (Rumel, 1921). The awareness and clarity already present so early on in the history of intelligence testing in terms of differentiation between the substantial and hypothetical nature of the construct is testament to the slow progress that has been made (Pressey, 1921). The end goal of assessment was and still is a practicable one; in other words seeking to aid versus seeking to characterize; the former being of central concern within dynamic assessment. Sufficient time and practice expenditure (assuming unlimited time and resources) will undoubtedly improve upon naturally possessed faculties. In order to ascertain intellectual functioning, the correct weighting, sampling of persons as well as domains will need to be accurately computed. After this is complete we are still no closer to defining what it is we are looking for. Thorndike (1921) concluded his contribution by stating that more research was needed! Terman's (1921) understanding of intelligence centered around the proportion or degree to which abstract thought made up intelligence functioning. Scaled utilization of intelligence tests was deemed a notable advance in assessment but still the need to consider intelligence in its multifaceted make-up is a theme within Terman's conceptualisation of intelligence. Simply put, passing a test of intelligence means nothing other than the fact the test was passed. Freeman (1921) concurs and emphasises the importance of an inclusive

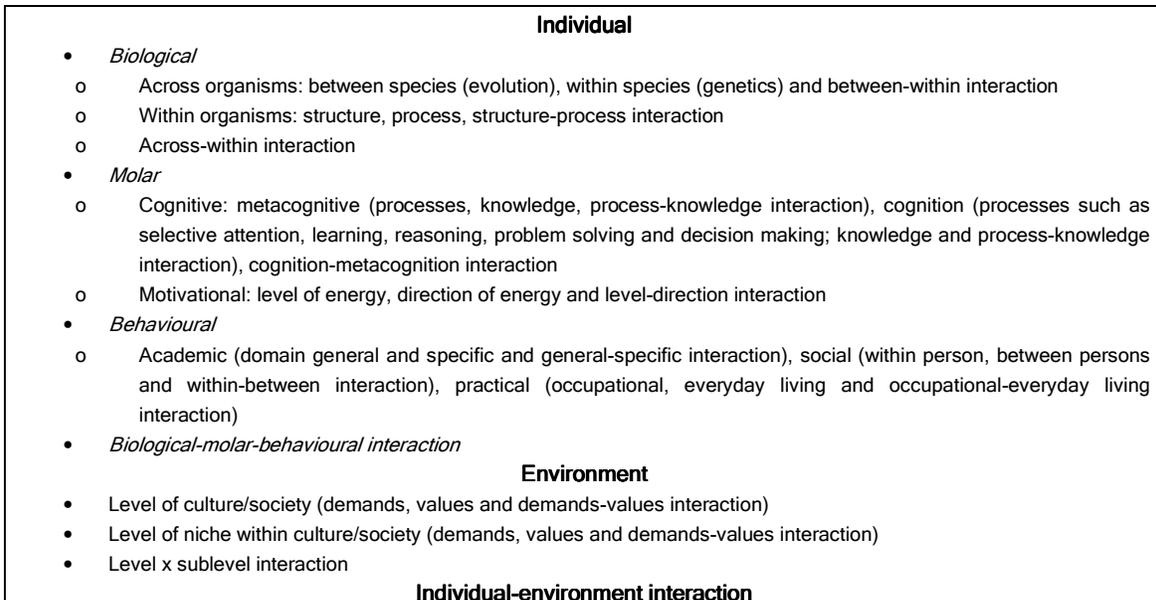
¹⁵⁵ Abbott's (1992) Flatland perhaps offers the easiest analogy here: two dimensional beings are unable to conceive of three dimensional beings and so exhibit frenzied fits upon the sight of three dimensional characters in their own two dimensional world. Instead of seeing three dimensions (a circle) they perceive merely a planar section. Taken from different angles, an intelligent flatlander will be able, in time, to piece together the myriad images into one whole unit. Can intelligence research not attempt the same?

definition of intelligence stating that an individual can be bright yet unintelligent. Depending, of course, on how one views the context.

Fully inclusive tests obviously do not exist and the multiplicity of tests available in 1921 led Colvin (1921) to lament the vast array of domains available for testing and he was also in agreement with the notion of intelligence not being a unitary factor. The author's own stated definition of intelligence given above is perhaps a definition that strikes an immediate intuitive appeal for lay people and scholars alike, for is adaptation not the hallmark of human life? Adapting to niches within the physical and cultural environment has enabled human kind to successfully navigate through time and space. Adaptability was proffered as general definition on intelligence by many scholars (Pintner, 1921) and includes the degree to which one is modifiable as Pintner (1921) states "... this leads us back to the general modifiability of the nervous system" (p.139). Is this not what dynamic assessment advocates? Brown and Campione (1998) pick this up in their assessment of the changes (if any) that intelligence definitions have undergone since the symposium. What has really changed in the intervening 85 years since statements such as these were uttered? The retort to the argument of general adaptation to the environment is that it hardly proceeds beyond general differences which is not the task set aside for individual difference researchers requiring finer levels of discrimination between individuals which is typically the area with which intelligence research concerns itself. Yet group difference intelligence research plays a primary role in synthesized research results across the spectrum and is perhaps even more contentious than individual difference research (Wittmann, 2005). Pintner (1921) himself recalls the "horror" of masses of intelligence research publications emanating during the first decades of the twentieth century and noting that a movement of "back to basics" resulted from the plethora of research results. Here we find ourselves back to basics 85 years later (Scarr, 1998b). There is clearly something amiss.

Over sixty years later, another panel of experts convened to discuss the nature of intelligence and the progress made since the 1921 symposium (Sternberg & Detterman, 1988). A number of loci of intelligence was considered as framework axes and included three main strata: the individual, the environment and the individual-environment interaction. The individual axis includes the biological, molar and behavioural levels. The environmental axis includes the level of culture/society; level of niche within culture/society as well as level x sub-level interaction. Each of these in turn describes a number of sub-categories within themselves and is testimony to the structured development and nature of intelligence research since the early decade of the twentieth century (Sternberg, 1988). Figure 21 illustrates this in-depth framework and considers much of what constitutes essential ingredients in the intelligence debate.

Figure 21 Loci of intelligence (Sternberg, 1988, pp.4-5)



Estes (1998) maintains that the oft-repeated question of what intelligence is, is rather old in terms of expecting an advanced answer and prefers to ask the question of where intelligence is now as opposed to 1921. Rather than attempt to continue with a question with no answer, it would perhaps be best to pursue other avenues of information pertaining to intelligence other than within the traditional grounds of psychology such as artificial intelligence for instance, an aspect not considered in 1921 (Schank, 1998). The major defining features between the construal of intelligence in 1921 and current renditions, is the former's emphasis on equating testing with intelligence understanding (although not exclusively so) and current conceptions of intelligence as far broader in scope and less preoccupied with the predictive nature of intelligence and more concerned with the nature of the



construct (Sternberg & Berg, 1998). Although this second point of concern is particularly contentious as doing away with intelligence measures of any sort is highly likely to reduce the predictive validity of a substantial number of intelligence and selection test batteries (Wilhelm & Engle, 2005). Similar debates abound in their current form much as they did in 1921 such as the unitary nature of intelligence (Detterman, 1998b) although there are moves towards regarding aspects such as metacognition and the like.

2.8.6 Cognition

Intelligence research is often subsumed within cognitive studies which includes, among other areas, artificial intelligence, neuropsychology, computational intelligence, cognitive science, social cognition, cognitive education (and specifically its relation to dynamic assessment; Lidz, 1992b), cognitive psychology and the increasing awareness of cross-disciplinary research¹⁵⁶ (Fetzer, 2002; Harnish, 2002; Jordaan, 1993; Potter, 2000). It has been questioned as to whether this grouping is necessary or even correct (Hunt, 1998) as the boundaries are often blurred (Boden, 1988). As with intelligence, cognitive psychology also appears fragmented without an agreed upon all-encompassing definition of a model or theory of cognition (Conway, 2005) which is cause for concern as the cognitive mechanisms underlying intelligence are considered the tools through which intelligence research progresses. Dynamic assessment is a manner of psychological assessment thus placing itself in the arena of cognitive psychology which is itself placed within the realm of cognition. Cognitive studies thus include an impressively wide array of focus areas including neurocognitive, behavioural, psychological, neuroanatomical and computational research (Ochsner & Kosslyn, 1999). Housed under the realm of cognitive science, dynamic assessment should lend itself to the rigours of the cognitivist approach which itself has opted for a more natural science inclination towards study and research (Strube, 2000). Evaluating the future of cognitive science warrants a level of analysis encompassing all aspects of cognition and in so doing affords an ecological frame of reference, through which cognition's past and future can be evaluated. This ecological level of analysis inherently claims that no one particular field of investigation seeking ultimate explanation can in fact carry out such a mandate, yet it would seem that current endeavours in neuroscience are attracting seemingly larger advocates of biological/physiological explanations of cognitive functioning. Neisser (1997), a proponent of just such an approach asserts that an either-or approach to the study of cognition will result in limited understanding and renditions of what is purported to occur cognitively. Asserting that the strides made thus far in terms of brain functioning within neurology provide lucid and at times contrary evidence to that provided by more traditional cognitive theories, does not necessarily imply the future abandonment of psychological theories or the usual model-building approaches towards cognition (Donald, 1997).

As a research science, cognitive psychology cannot hold sway over the entire discipline of intelligence with which it is intimately linked predominantly due to the fact that very much the same problematic issues beset the field as they do with intelligence: the transient link between theoretical and empirical construct. Jensen (1998b) refers to this matter thus "the vehicle is not the construct; the construct is not the vehicle" (p.309) which in sentiment can be traced back through to the level of empirical construct and this is derived from the measurement instrument.¹⁵⁷ Within the last fifty years, psychology has been the continuous product of a confluence of disparate enterprises culminating in broad-ranging theories and models and owes much of its cognitive repertoire to the research areas dominated by, among others, Chomskian linguistics, neuroscience, computational theories and ethology (Potter, 2000). The history of intelligence assessment is bound intimately with the history of cognition and the two can hardly be delineated as distinct enterprises which makes for the demarcation of these sub-disciplines very tedious. Most burgeoning areas within cognitive science (cognition) occurred within the twentieth century (Posner & DiGirolamo, 2000) after foundational psychometrics had been laid, hence the need to briefly consider cognition. The psychometric model which emphasises the measurement of individual differences encompass theories which are based on patterns of relationships among test scores. The information processing model typically emphasises precise theories of how cognition occurs; and specific tasks rather than broad generalisations are studied. The cognitive developmental approach emphasises commonalities between people rather than individual differences. All three have overlapping similarities but major differences characterise their unique approaches (Kail & Pellegrino, 1985). Perhaps greater attention can be paid to instructional influences on cognition and the brain and how dynamic assessment can be positioned to tentatively test within this approach (Ashman & Conway, 1997; Beatty, 1995; Friedman & Cocking 1986; Walters & Gardner 1986).

This very short digression into cognition serves only to place or situate intelligence research as sub-discipline. There are numerous overlapping areas which concern intelligence and some areas which do not share similar agendas. Hence there will be no discussion, bar a fleeting reference to the very broad field of cognition. Where some insight is necessary however, is within the notion of bridging principles already discussed above in section 2.4.5.3.2 and to be discussed again in section 3.4.2.1

¹⁵⁶ At times, these terms are used synonymously and interchangeably and at other times are used to delineate differing meanings.

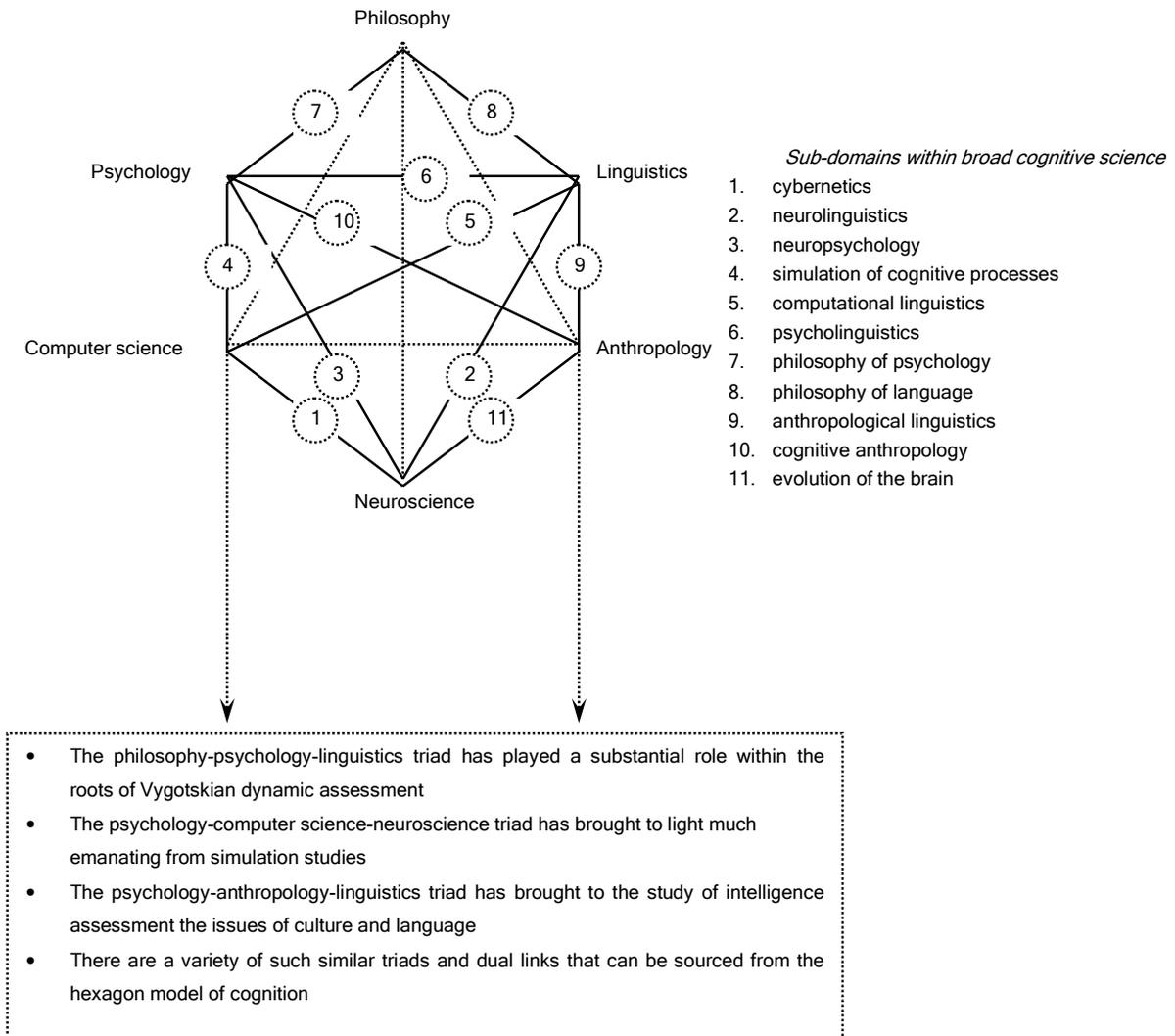
¹⁵⁷ Jensen (1998b) argues cogently for the stability of psychometric *g* even though it has been shown that IQ levels are on the rise (performance IQ not verbal IQ) with a concomitant decrease in scholastic test results. This poses somewhat of a problem when thought of in terms of both instruments' *g*-loadedness. Jensen's argument is that *g* has not changed, but rather the vehicle through which such conclusions are drawn. Since measurements are dependent on the instrument, these measures will thus change. Hence, *g* is manifest through instruments which evidence measurement. There are areas of concern for all three aspects as is argued in this thesis.

below. The link between isomorphic renderings of psychological-neurological understandings of brain/behaviour is a recurrent theme within this thesis. The notion pops up once again in this discussion within cognition. Various theories and models of cognition allow for the brain to be modeled and understood in terms of systems representing mental states whilst other models construe the brain as very much brain-like, such as connectionism or parallel distributed processing (PDP) for instance (Jorna, 1995; McCauley, 1998) which models the physiological brain in a similar manner (Green, 2001). Research into the connectionist models of cognition came to the fore in the 1980's (Green, 2001). The connectionist computational theory of mind and the digital computational theory of mind are both positioned under the more generic computational theory of mind which itself is viewed as part of the older representational theory of mind (Harnish, 2002). The confluence of many sub-disciplines within intelligence research may seem overwhelming in their capacity to discover much more about human functioning than in any era before the present and the need to link research findings across disciplines could be construed as a bad move. Max Planck¹⁵⁸ once referred to the acceptance of ground-breaking science as a very slow process, only being accepted by the new generation as the older generation die off - perhaps we are in the midst of a new paradigm within intelligence research and are not yet aware of it. According to Hunt (1997a) cognitive science is simply a revisionist attempt at looking at the intelligence paradigm, a paradigm which Woodcock (1998) suggests should start to reduce the lag behind cognitive science. Once again, any view necessarily has to be constrained lest the domain of focus becomes overwhelmingly complex resulting in too large a theoretical conception of what is being studied.

There has been a progressive move away from behaviourist renderings of the "mind" (positivism) towards the cognitivist view of brain (realism) where the former denies access to what it cannot sense and where the latter, via the hypothetico-deductive method of science investigation (see chapter 3), attests to the brain's further discovery (Harré, 1997). However, the situation reflective of the contextualised/decontextualised development models noted in the psychological literature (Piaget and Feuerstein for instance) is reminiscent of the parallel developments within the cognitive sciences in which various aspects are considered decontextualised and others contextualized. This above-mentioned move away from behaviourist towards cognitivist renderings of human functioning is often viewed as a standard dichotomy (Spiker, 1989) within the broader field of cognition but may well be a normal path of progress along which the discipline of cognition and cognitive psychology travels. The so-called cognitive revolution may be a misnomer and was moreover confined to the United States (Bechtel, Abrahamsen & Graham, 1998; Mandler, 2002). Nevertheless, along with a renewed approach towards the understanding of cognition-in-context (situated cognition) is the similar move towards ecologically valid understandings of human performance within educational and social environs especially in the area of cognition and learning contexts (Glaser, Ferguson & Vosniadou, 1996a) and away from laboratory-based results of human performance. This is much in keeping with current dynamic assessment trends and tenets representative of sociocultural approaches towards education and its attendant technologies (Dillenbourg, 1996). Learning takes place within interactional contexts and not in isolated situations (Vosniadou, 1996b) and even if there is no human mediator present, the system of tools mediates much of what is learned. Piagetian developmentalist notions are understood within an abstract framework within which development occurs (domain independent) as opposed to Feuersteinian notions of culturally dictated or collaborative frameworks (domain independent) (Mendelsohn, 1996). Vygostkian notions have made their presence felt over a very broad spectrum of application within education and education technology such as the ZPD's transference into scaffolding concepts (De Corte, 1996; Day & Córdón, 1987; DiLalla, 2000; Dillenbourg, 1996; Kanselaar & Erkens, 1996; Mayer, 2001). Donald (1997) insightfully depicts this relation in his comments on how artificial intelligence, cognitive psychology and Chomskian linguistics was supposedly to have supplanted its decontextualised forbears, but ended up speculating in very much the same vein. Time and again, researchers are confronted with levels of description upon which their opinions and data need to be leveraged for anything to make sense. No sub-domain as yet managed successfully to answer all questions, least of all because we simply do not know most of the answers, but more so due to this inherent limitation within the scientific method. Nevertheless, science and cognition within it forge on in what Donald (1997) hopes will be representative of an integrative approach and not a reductionist one. Figure 22 illustrates the various domains within cognition research and how most, if not all, have become subsumed within the larger domain of intelligence assessment including dynamic assessment.

¹⁵⁸ In fact Planck (1858-1947) hailed the year 1900 as the year of transition between classical and modern physics with the discovery (origin) of quantum physics. Quantum theory, itself revolutionary, had yet many years of development ahead of it before it was accorded status of "new paradigm".

Figure 22 The cognitive hexagon (Pylyshyn, 1983, p.76): cognitive science's boundary transcendence



Such constraints however, bring with their own limitations such as the utilisation of the computer metaphor as analogy for the brain. One major limitation in understanding how the brain functions and learns when viewing it within the computer analogy is its dependence on a recursive argument and the fact that the brain's biodynamic architecture is very much unlike the architecture of a binary representation of information (Pascual-Leone, 1997). It is more like the neuropsychological models based on neural nets (Anderson, Silverstein, Ritz & Jones, 1977; Cilliers, 1990) emphasised over thirty years ago. The computer metaphor is a product of our thinking and is of course merely a snapshot of how we think we operate and as such we proceed to base and compare findings to the computer analogy thus further narrowing our understanding of how the brain functions, after all we are basing our functioning on a model that is partly derived from this very functioning, so then is it not a limiting analogy of which to make use?¹⁵⁹ As Broadbent (1958 in Shotter, 1997) states "we are trying to model man, and yet there is a man in the model" (p.324). Computational psychologists are not computer-model-mind followers necessarily but do share some similarities with computer-analogy theorists of perceptions (Boden, 1988) as to how the brain functions such as the

- Adaptation of a functionalist approach to the study of the brain in which psychological process are viewed as procedures
- Conception of the brain as a representational system
- Link to neuroscience in terms of understanding the functional relations via neurology; i.e. how logical operations are carried out within a neural network framework

¹⁵⁹ Likewise with the pioneering work conducted in artificial intelligence, but here at least one is not confronted with a back-to-forward view of the brain in terms of the computer metaphor. Linear functioning artificial systems are of course not in the same league as the functioning non-linear brain (Warwick, 1998)

However, computational systems struggle to map the learning and development process as it is not even fully understood within the biological realm of study. When positions such as these are proffered it is done so often from a meta perspective in order to frame the argument. That there are similarities with intelligent functioning in humans such as the processing of relations for instance is not in question (Halford, 1999). Yet such a completely solipsistic argument allows one no-way out of such a muddle and though hard-pressed to ignore the argument in favour of just such a view even thinking about a way out is considered as time better spent pursuing other avenues. Such a view will not aid in the discovery of how the brain develops and changes, and a healthy dose of such solipsism is taken under consideration but is not an avenue this thesis seeks to wind its way through (Bates & Elman, 1994).¹⁶⁰ Formal computational thought models are extremely useful but only as one of many nested metaphors for describing psychological states where progress on a number of fronts can be made within simulation environments and then brought to bear on reality.

By our computer analogies, we are only depicting a narrow conception of how we think we operate, which is to be expected, as the partial reasoning behind this exploration is to further elucidate our own functioning, which is then reflected back to us for more interpretation and so on. This computer-reductionist approach is referred to as the hard cognitive science assumption,¹⁶¹ an assumption which has been criticised for being too reductionistic and not encompassing of more “subjective” approaches towards such study; but for which for instance Pascual-Leone has offered a tentative remedy¹⁶² (Pascual-Leone, 1997). Although critical of the hard cognitive science approach, Pascual-Leone does state that computer metaphors involving programmes and theories serve to illuminate how certain cognitive functions occur but mental processing itself encompasses this computational approach as well as broader, more explicable theories of how the mind works (including his metasubjective explanations). In other words computational processes (as described by the hard science metaphor) only yield partial glimpses into the nature of the mind’s psychological functioning and although it can and should be utilised (in describing psychological processes) it should be used in conjunction with the more complete armoury of metasubjective tools¹⁶³ to study these processes. Notwithstanding these reservations it must be noted that computer science has influenced theory and methodology within the information processing approach (Reynolds, 1987). This includes the strides made within other areas of investigation where computational simulations elucidate (and at times are able to predict) closed and open system behaviours¹⁶⁴ (Dennett, 2004), and are capable of serving as an interface between traditional cognitive assessment and neuroanatomical models of functioning (Neufeld, 2002). Norman (1988) explicitly states that the architecture of the modern digital computer was a result of humans’ conceptualisation of how the mind in fact operated and to view the computer metaphor as a limited comparison may well be founded, but that it was an original¹⁶⁵ conceptualisation of “mind” must similarly be kept in mind (d’Ydewalle & Denis, 2000)! However, it would seem that research into [computer] engineering and psychology are synergistic in terms of both fields contributing knowledge to the other (Gardner, Kornhaber & Wake, 1996), pointers if you will, in the directions into which certain sub-disciplines should enter. Vygostkian sign or semiotic tool and role of language as sign system has a modern-day ring to it in terms of computational analogies and how computers are based on a system of signs (mark manipulators) such as programming languages but sign utilisation within human language and computational language are fairly far apart in many respects. Within this framework of symbol manipulation and semantic theories of language one arrives at the static difference between computers which manipulate marks as opposed to humans who process signs. Another framework dictates the algorithmic rule-based system which computers follow as opposed to human non-algorithmic thought processes and is subsumed under the rubric of dynamic differences. Lastly there is the conception of the new paradigm of simulationist thinking which comprises simulations, emulations, replications and emulations via neural networks, connectionist and artificial intelligence research (Fetzer, 2002) which is also being increasingly utilised in statistical analyses of behavioural data (Baker & Richards, 1999; Maree, 1998). Figure 23 illustrates the three framework conceptions.

¹⁶⁰ See Kleene (1967) who states that according to the Church-Turing thesis there are some values which a machine cannot compute; in other words a recursive function in which something or some machine refers back to its own description (Church and Turing concluded very similar sentiments within the same year, 1936, yet they did so independently; Copeland, 2002). Kleene’s own recursive function theory was applicable in the determination of the range and limits of axiomatic systems in mathematics (Mahoney, 1997). Gödel maintained that a logical system might be consistent but never complete and Alonzo Church maintained that it was impossible to devise any algorithm which could demonstrate which theorems were provable and which were not (Rucker, 1987; Turner, 1967). Of interest is the fact that Gödel’s metamathematic proof was axiomatised in terms of arithmetic; in other words he was able to provide a proof in the same language for which he was trying to find a proof (Butterworth, 2000); see chapter 4 for more information on axiomatisation.

¹⁶¹ Here, the hard cognitive science assumption refers to both the older concepts of AI (artificial intelligence) and the newer versions represented by connectionist simulations (Pascual-Leone, 1997).

¹⁶² Pascual-Leone intends remedying this approach by creating for this new confluence of hard and soft approaches to cognitive science a new lingua-franca which is informed by a metasubjective theory; a metatheory which is a “... theory of the observer, but of an observer carefully modelling subjective and objective processes of the ‘psychological’ organism, not of the computer” (1997, p.79).

¹⁶³ Included here are aspects of psychological functioning that have as yet not been adequately simulated by computers, such as affect, consciousness, development and so on (Pascual-Leone, 1997) although with the aid of sophisticated computational intelligence software, ‘archaic’ programmes such as ELIZA (developed in the late 1960’s) are progressing rapidly in terms of the modelling human behaviour (Reynolds, 1987).

¹⁶⁴ Here reference is made to British mathematician John Conway’s “Game of Life”, “an oversimplified model of determinism” (Dennett, 2004, p.36) developed in the 1960’s.

¹⁶⁵ See Norman (1988) in which he states that a host of researchers and academics from disparate disciplines were already engaged in the development of the digital computer ca. 1940.



Figure 23 Three conceptions of computer-mind analogies (Fetzer, 2002, pp.xiii-xvii)

Static difference		Dynamic difference		Simulations	
<i>Argument 1: Computers are mark manipulating systems, minds are not</i>		<i>Argument 2: Computers are governed by algorithms, but minds are not</i>		<i>Argument 3: Digital machines can nevertheless simulate thought processes and other diverse form of human behaviour</i>	
<i>Premise 1</i>	Computers are mark-manipulating systems, minds are not	<i>Premise 1</i>	Computers are governed by programmes, which are causal models of algorithms	<i>Premise 1</i>	Computer programmers and those who design the systems that they control can increase their performance capabilities, making them better and better simulations
<i>Premise 2</i>	These shapes, sizes, and relative locations exert causal influence upon computers, but do not stand for anything for those systems	<i>Premise 2</i>	Algorithms are effective decision procedures for arriving at definitive solutions to problems in a finite number of steps	<i>Premise 2</i>	Their performance capabilities may be closer and closer approximations to the performance capabilities of human beings without turning them into thinking things
<i>Premise 3</i>	Minds operate by utilising signs that stand for other things in some respect or other for them as sign-using (or "semiotic") systems	<i>Premise 3</i>	Most human thought processes including dreams, daydreams and ordinary thinking, are not procedures for arriving at solutions to problems in a finite number of steps	<i>Premise 3</i>	Indeed, the static and the dynamic differences that distinguish computer performance from human performance preclude those systems from being thinking things
<i>Conclusion 1</i>	Computers are not semiotic (sign-using) systems	<i>Conclusion 1</i>	Most human thought processes are not governed by programmes as causal models of algorithms	<i>Conclusion</i>	Although the performance capabilities of digital machines can become better and better approximations of human behaviour, they are still not thinking things
<i>Conclusion 2</i>	Computers are not the possessors of minds	<i>Conclusion 2</i>	Minds are not computers		

A refreshing alternative to the hardware computer model of the mind (which hardly does justice to the electro-chemical make-up of the brain) is one which is lauded by Greenfield (1997) as a "molecular symphony" (p.82), which, as a real description of how the brain functions, can not as yet be fully described by the prevailing computer models of the brain. Of course it must be kept in mind the level of description we wish to describe (Hofstadter, 1980); be it intelligence models of how human brains recall memory or compute logical analogies, or neural architectures of how the brain physically functions, which encompasses neurochemical explanations. How thought is conceptualised in the brain can be ascertained from tracers in the blood in the brain or can be determined by more global models of thought which are currently the preferred mode of explicating how the brain in fact does function in psychological intelligence models. Which level of description does justice to the manner in which human beings think? At which level should remedial programmes be pitched in order to overcome or redress imbalances in thought processes? At which level outside the system should one even begin to engage in trying to remediate? Environmental or genetic? Hence the need for a more engaged and inclusive dialogue which should be favoured in the new century. This continual lament (Wilson, 1999; Royce, 1973) for an innervated inclusive research programme which will probably still be argued for in decades to come. Until then, the level of description at which remedial programmes are levelled will have to content themselves with lack of information about other levels and an inundation of information from its own chosen level of description. Shall we refer to this all-inclusive level as a 'thick description'?

In sum, the hard cognitive science approach is and should be utilised as a means of explanatory metaphor when analysing mental processing but due to the recursive nature of this tool (in terms of it lacking full self-description) should not be used exclusively within the psychological realm and needs to be combined with more traditional (yet not stale) soft science approaches. This has a bearing on how theories of mind, theories of intellectual functioning and ultimately theories of dynamic assessment are operationalised. Initial research into devising computers though, was based and constructed on intelligent behaviour, information gleaned from research into intelligent systems. In a rather simplistic analogous model, human information processing can be equated with computer processing at a very basic level. Perception in the form of input (visual or audio) is analysed at various levels in humans and in terms of binary digits in computers (although the input may present itself in a variety of forms, the computer compiler translates the varied input into machine readable code), information is then temporarily stored in "primary memory" (Craik and Lockhart, 1972) where it is housed till the immediate task is completed (which may be seen to superficially resemble RAM¹⁶⁶ in computers). Our long term memory storage can similarly be equated with a computer's hard drive (or any removable storage device).

The information processing paradigm or metaphor although having found its predominant field of application to be that of, among others, pattern recognition, logical problem solving and perception (Broadbent, 1987;¹⁶⁷ Kenrick, 2001) has woven its way into more diverse psychological areas of interest including clinical, social, developmental psychology as well as human-computer interaction (David, 2004; Posner in Broadbent, 1987; Reynolds, 1987). Dating the initial presence of any discipline or sub-discipline is not an exercise that can be exacted with precision and certainty, although it is generally recognised that researchers such as Chomsky, Newell, Simon and Miller were at the forefront of this new information processing approach to the understanding of human behaviour and whose pioneering work started to gather momentum in the 1950's (David *et al.*, 2004; Harnish, 2002). Later, Neisser's 1967 text on cognitive psychology helped solidify this emerging field within psychology (David *et al.*, 2004). The study of cognition having begun in the mid-nineteenth century using experimental techniques may be crudely said to have evolved into the information processing approach as the two are at times used synonymously (Reynolds, 1987). Housing the information processing approach within the larger framework of cognition as opposed to seeing the former as a progression of techniques within the latter remains debateable. The move towards mapping brain function onto brain structure witnessed the rapid growth of neuroscience (Harnish, 2002), the level of description more easily attuned to physical reality but because so much remains hidden from view, information processing models often interpret the working model from information that is already known. As has been mentioned a number of times, levels of description are inherently narrowly construed but to effect a cross disciplinary programme, all levels need to be viewed when considering human behaviour in its totality. Most of the above levels of description have one common thread running throughout their research paradigms and that is to represent brain and behaviour in some fashion. Hence, most are representational theories of mind (Harnish, 2002). Digital computational models of the brain are subsumed within general computational models themselves serving under the broader model of representative models. Figure 25 illustrates the representational nature of both neural and mental structures which loops back to the brief discussion on isomorphism within neurological explanations of psychological behaviour (see chapter 2 section 2.4.5.3.2. figure 2 and chapter 3 section 3.4.2.1 figure 34). The explanation of psychological states via neurological models and the subsequent invoking of bridge principles can effect an understanding of brain working and does so within the broader field of cognition under which is housed intelligence and dynamic assessment. The two above-mentioned figures illustrate how behaviour is translated into data via semantic linkages and various empirical and bridging principles as well as rules of correspondence which results in hypotheses. In keeping with the so-called bridge principles discussion, figure 24 depicts Simon

¹⁶⁶ Random Access Memory

¹⁶⁷ Reprint of Broadbent's original 1958 work.

and Wallach's (1999 in Strube, 2000) six criteria of empirical adequacy in order for cognitive models to correctly map from and between computational modes to human behaviour modes.

Figure 24 Simon and Wallach's (1999 in Strube, 2000) criteria of empirical adequacy for correspondence expanded upon by the depiction of the role of bridging efforts

Cognitive model paralleling real life behaviour - six ideal bridging or correspondence criteria	
Product correspondence> Overall performance
Correspondence of intermediate steps> Processing in the model parallels separate stages in human processing
Temporal correspondence> Computation expense parallels reaction-times patterns
Error correspondence> Same error patterns in model and in experimental data
Correspondence of context dependency> Comparable sensitivity to known external influences
Learning correspondences> Identical learning curves

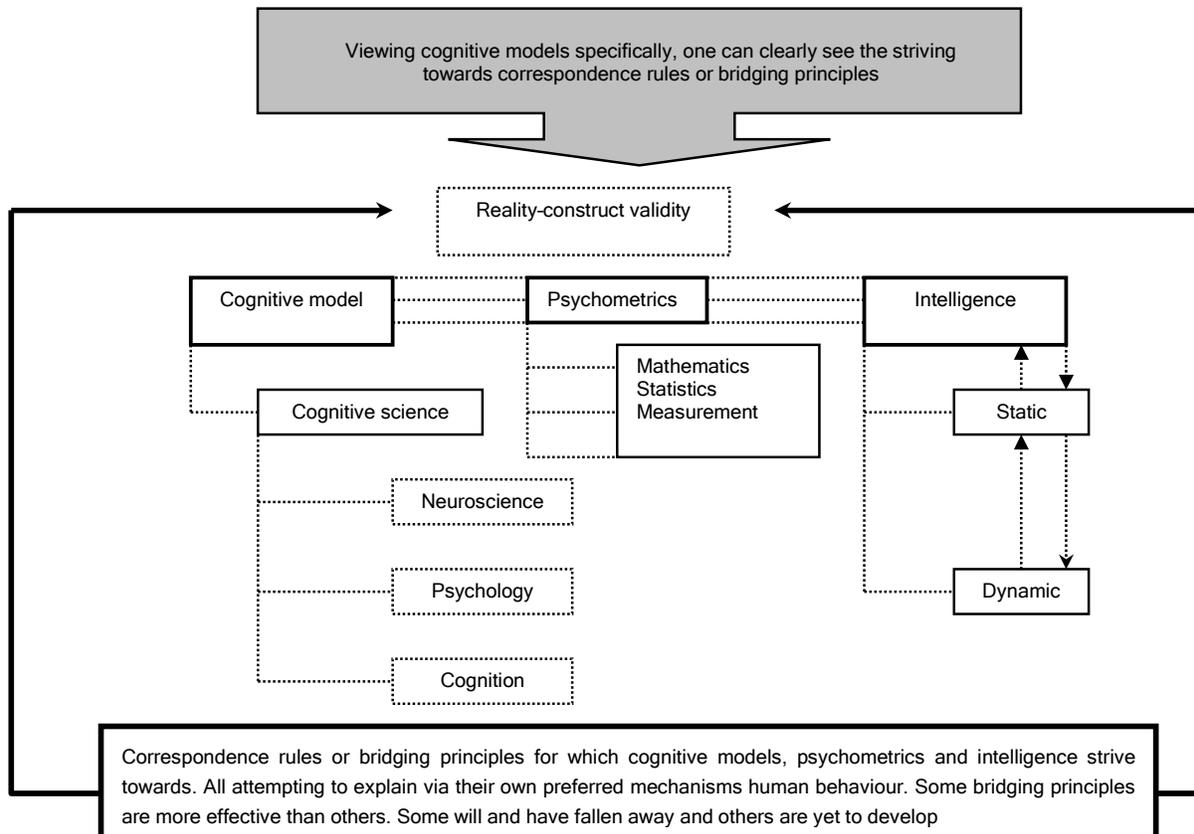
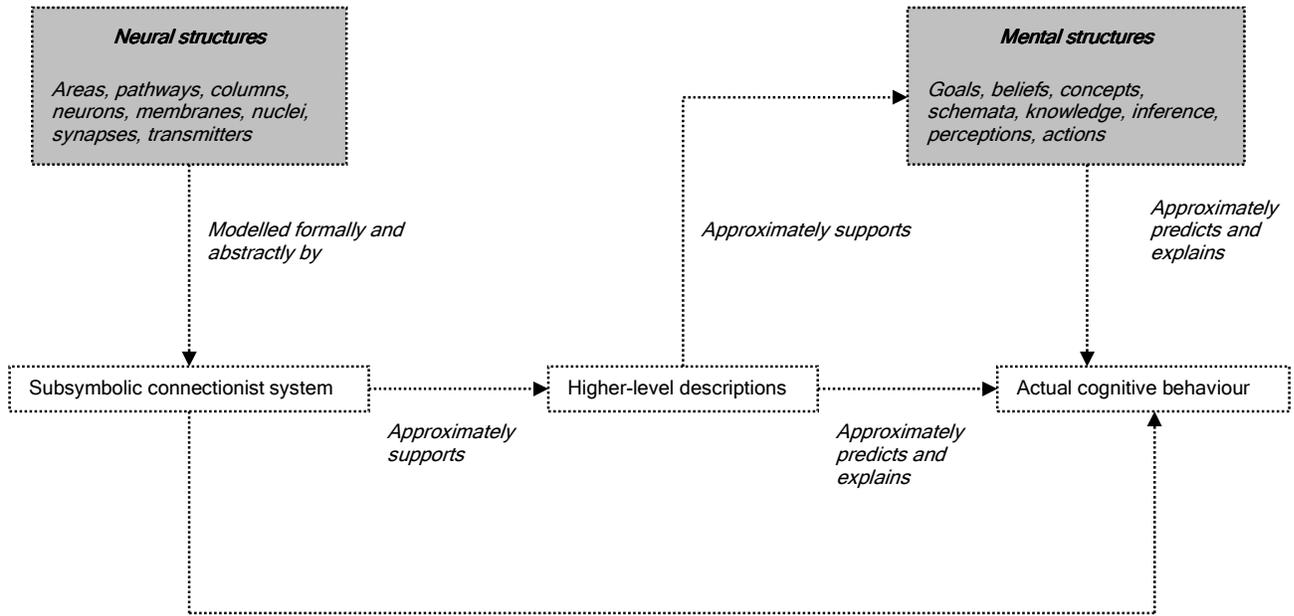


Figure 25 Neural and mental structures in the sub symbolic paradigm (Haugeland 1997 in Hamish, 2002, p.347).



Note the similarity within this cognitive domain illustration and figures 2 and 24.

Computer-assisted assessment - moving steadily into the twenty first century and a niche for the further expansion of dynamic assessment?

The number of dynamic assessment instruments currently available using the computer as medium is steadily increasing (Jacobs, 2001). The efficacy of using different media when attempting to teach in any pedagogical setting has been studied for a number of years (Greenfield, 1987) and has placed emphasis upon the integration of cognitive psychology principles into these media of instruction (Mayer, 1997, 2001, 2002, 2003; Mayer & Anderson, 1992; Mayer, Dow & Mayer, 2003; Mayer, Fennell, Farmer & Campbell, 2004; Mayer, Heiser & Lonn, 2001; Mayer & Moreno, 1998, 2002a, 2002b, 2003; Mayer & Sims, 1994; Mayer, Sobko & Mautone, 2003; Plass, Chun, Mayer & Leutner, 1998). It has for instance been shown that varying cognitive strategies for processing information come to the fore when information emanates from different media such as print, audio or visual media (Mautone & Mayer, 2001; Moreno & Mayer, 1999, 2000, 2002). Dynamic assessment engages individuals within what is understood to be a co-construction (scaffolding) of reality and this principle in turn is utilised within various multimedia learning environments (Herrington & Standen, 1999). Is there really a marked difference between performance using dynamic assessment techniques which have as their medium the traditional paper and pencil format as opposed to performance when utilising computers for instance? The very nature of dynamic assessment is one of learning to learn within any specific assessment situation, so, the argument may proceed, if learning is affected by the medium of instruction, it is feasible to state that the media used in dynamic assessment setting should influence the nature of learning or at the very least strategies employed in the processing of information. Every medium has inherent biases which are prohibitive when using certain information processing strategies, however, it is these very biases which allow for other strategies to be used in turn when assessing for alternate strategies.

- *Print media*

The traditional format of instruction and assessment works well under circumstances where the information that is given requires the imagination of the reader in order to complete the expressed experience but does not promote the use of visual movement (Greenfield, 1987) an area which may be more suited to media which are able to carry the task (such as computers and or television).

- *Audio*

One aspect found in dynamic assessment that does in fact make use of the audio medium in assessment is the hints methodology used in testing-the-limits as means of eliciting correct information processing strategies. Of note is the fact that when analysing audio presentations dialogues and figurative language are understood better as opposed to receiving the information from a visual source (Greenfield & Beagles-Roos, 1983 in Greenfield, 1987). Administering assessment in the testing-the-limits manner might well be more beneficial using an audio format as opposed to a visual one, however, this depends heavily on the aspect that is being assessed, certain types of reasoning strategies may be more easily understood using visual media when encompassed within a testing-the-limits scenario. Each test and sub-test will thus have to be screened for best choice of medium taking into consideration the specific processing strategy that is being assessed.

- *Television*

Information regarding action is best sent via media utilising three dimensions as this information will be better understood in context. Action information is better recalled by subjects when it is sent via television as opposed to receiving it via audio transmission (Greenfield, 1987). Computing technology has progressed to the point of allowing for mastery and surpasses that which is offered via television, hence the greater concern with computers as transmitters of action information as opposed to television.

- *Computers*

Computers are at once exceptionally useful exponents of understanding brain functioning, their explanatory power is evident in the many areas in which they are used to derive analogous functioning of the human brain that it has become increasingly difficult in today's technologically oriented environment not to make use of computing power when understanding the process of learning and of information processing. The role of the afore-mentioned media cannot however be questioned as to their efficacy in terms of their pedagogical roles, however, when computers are able to allow for assimilation of all these media in one, the usefulness of this particular medium becomes evident.

Simulation techniques are able to offer researchers a platform in which "complex systems having multiple, interacting, dynamic variables" can be assessed (Greenfield, 1987, p.19). Whether the translation of pencil-and-papers assessments to assessment on computers makes a difference is one question, but the manner in which this translation is carried out is another. This "mindless" translation between media may well be an area to watch closely (McArthur, 1987).

2.8.7 Non-intellective factors

Non-cognitive factors play a paramount role in intelligence,¹⁶⁸ life/job success or achievement (Cattell, 1963; Grossberg & Gutowski, 1987; Howe, 1987; Watkins & Mauer, 1994), school-based performance as well as dynamic assessment predictive validity and often come to the fore during the assessment situation of average and learning disabled students (Budoff, 1987a, 1987b; Hamers & Sijtsma, 1995; Haertel, Walberg & Weinstein, 1983; Kormann & Sporer, 1983; Miller, 1998; Revelle, 1987; Short, & Weissberg-Benchell, 1989; Resing, Ruijsenaars & Bosma, 2002). The latter group, for whom dynamic assessment is a particular choice of assessment technique (Lauchlan & Elliott, 2001) cannot be treated as a homogenous group and they are often unfairly and disproportionately represented in educable mentally retarded groups (Hamers, Hessels & Pennings, 1996; Harrison, Singer, Budoff & Folman, 1972; Schlatter & Büchel, 2000; Short & Weissberg-Benchell, 1989) just as culturally disadvantaged individuals are not similar in all respects. Hence dynamic assessment's concern for the individual within the group. It has been evidenced that sample groups treated with dynamic assessment interventions may not appreciably differ on IQ tests, learning potential scores or on school marks but differences between such groups becomes more manifest when non-cognitive aspects such as cognitive style, motivation and attitudes are assessed (Babad, 1977). This highlights the need to perhaps be more inclusive of such traits in dynamic assessment interventions as well as more static mainstream assessments. Criticisms of mainstream assessment are directed towards this missing aspect within tests where non-intellective factors are simply not considered (Tzuriel, 2001) and the efforts involved within cognitive and structural modifiability attest to the increase in performance of those who change from impulsive to reflective cognisers (Carlson & Wiedl, 1992).

The purported success of psychology's measurement endeavours in years past which resulted in the perception of greater access to educational resources has played out in precisely the opposite manner for many minority groups (Sewell, 1987). It is these groups who stand to benefit more so from dynamic assessment interventions than their higher functioning counterparts within the dominant cultural group (Lidz, 1987b). *G*-based research accounts for intelligence but does not necessarily reveal the whole story of scholastic achievement (Nettelbeck, 1999) though the author tends to veer towards *g*-based performance. In other words, *g* severely underwrites global functioning no matter in what state of life an individual happens to find themselves. This does not negate the argument from polygenetic and phenotypical reactions to prevailing environments. Can phenotype be considered a norm of reaction or a response to unalterable genotype? The phrase "norm of reaction" was already coined as far back as 1909 by Woltereck and explains the changes undergone within an organism due to interaction of genes and environment ("genetic" does not refer to stagnation; Valencia & Suzuki, 2001). The greater the environmental resources the greater the phenotypical range of values (Ceci, Rosenblum, De Bruyn & Lee, 1997). Genetic inheritance works via additive and non-additive influences; the former being the result of the additive influence of genes which in turn influence the phenotypical response whereas non-additive influences are caused by differential influences of one or more genes at various locations (Brody, 1992, Grigorenko, 2004a) which are thus unique to each individual. It makes a mockery of genetic heritability research to be recast in eugenic or naïve consistency theory terms (where "heritability" in the past was construed as stable and unchanging; Sato, Namiki, Ando & Hatano, 2004) but acknowledgement of the overwhelmingly large role played by this form of interaction is necessary for an effective future programme or tradition on intelligence research. One can attempt to redress phenotypical responses through the use of programmes which seek to foster healthy "IQ-boosting" environments (also via micronutrient and general nutrient supplementation, Arija, Esparó, Fernández-Ballart, Murphy, Biarnés & Canals, 2006; Brody, 1992; Colom, Lluís-Font & Andrés-Pueyo, 2005; Eysenck & Schoenthaler, 1997) and can achieve movements up (or down) the IQ scale but one is always imprisoned within the norm of reaction. Moreover it has been evidenced that nutrient supplementation really only benefits lower IQ performers resulting in decreased variance of this population and not really affecting the higher performers in the same manner (Colom, Lluís-Font & Andrés-Pueyo, 2005). The inevitable politicisation of these contentious issues results in a jaundiced view of the attempts to increase the unchangeable. Figure 26 below illustrates the certainty in the calculation regarding heritability and genetics and their subsequent influence on intelligence.

¹⁶⁸ It is this author's opinion, however unsubstantiated this claim may be, that non-cognitive factors may supersede intellectual powers in some instances. Determination, perseverance, will power and the like are more powerful in certain contexts than levels of intellectual functioning. Affect influences cognition and cognition influences affect (Messick, 1996). Of course this does not deflect attention away from what is genetically a predisposed likelihood of potential functioning.

Figure 26 Heritability, intelligence and its calculations (Grigorenko, 2004a)

The not-so straightforward understanding of heritability and intelligence

- The relation between phenotype (P), genotype (G) and environment (E) can be expressed as
 - $P = G + E$ (“in the genetic analysis of multiple phenotypic measures obtained on each individual, [a] basic structural model is invoked repeatedly for each phenotype, so that the system of linear structural equations defining a multivariate system [is characterised by this relation]” (Volger & Fulker, 1988, p.480)
- However, G and E must yet be divided into their respective components, namely,
 - Additive effects (A) from the combined differential effects of two alleles within and between genes; dominant effects (D) resulting from interaction between the alleles; epistatic effects (I) resulting from the interaction between different genes, environmental effects shared by all family members (S) and non-shared environmental effects (N). Thus, the general components of a phenotype can be expressed as
 - $P = (A + D + I) + (S + N)$
- However, the above expression does not take into account individual differences, or variance (V) within a population. The total variance is the phenotypic variance and is expressed as
 - $V = V_G + V_E = (V_A + V_D + V_I) + (V_S + V_N)$
- G and E components may be correlated leading to a value of V_P twice its value in the above expression; twice the covariance of (G)(E) or $2Cov(G)(E)$. Also there may well be interactions between G and E which is represented by $V_G X_E$
- The above expressions have decomposed via variance the heritability coefficient which is the indicator of the contribution of genes to the variation in a trait, namely, phenotypic variation
- Heritability narrowly construed is defined as V_A/V_P and heritability broadly construed is defined as V_G/V_P
- The above is, as all expressions in this thesis, purely abstract

Cautionary announcement ...

- Behavioral ecology, behavioral genetics, sociobiology or however the area of interest is known is a population-based approach towards the quantified understandings of genetic/phenotypic interaction
- Intelligence research emanates from a very focused tradition of individual difference research
- **Hence population-based research ≠ individual difference research (the coefficient of heritability is not indicative of between population difference and moreover views population differences not individual differences)**
- Behavioural genetics research norms cannot be directly imported into individual difference norms
- Eugenics research did just this resulting in many unsavory activities
- Modern-day understandings take into account the above and are moving towards inclusive research (culture, bio-ecological models and the like) but
- This does not mean that behavioural ecology and individual difference research cannot work in tandem

And lest we forget ...

- There are four mechanisms of genetic influence that have to be considered when discussing and debating the issue of heritability and intelligence, namely:
 - Mutation - random changes in both coding and non-coding genes
 - Random genetic drift - over time alleles change in frequency due to random sampling error which decreases as the population grows
 - Gene flow (genetic exchange) - combination of genetics from two different populations results in the offspring resembling both parent populations
 - Natural selection - resulting from the success of breeding within an environment, usually those that have adapted well (which is of course co-determined by the above three mechanisms)

(Sternberg, Grigorenko & Kidd, 2005)

Motivation, anxiety, metacognition,¹⁶⁹ persistence, impulsivity, temperament, confidence, personality factors (such as openness to experience for instance) (Gignac, 2005; Matthews, Zeidner & Roberts, 2005) and self-esteem (particularly pertinent to mediated learning experience and the learning experience in general) (d'Ydewalle, 1987; Elliott, Lauchlan & Stringer, 1996; Glaser, 1987; Hansen, 2003; Haywood, Tzuriel & Vaught, 1992; Kozulin, 1999; Meijer & Elshout, 2001; Paour & Cèbe, 2002; Resing, 1997; Samuels, 2000; Skuy, 2002; Zeidner, Matthews & Roberts, 2004) as well as higher order metacognition such as person appraisal, task demands and strategy; (Howie, 2003) and the rather confusing aspect entitled cognitive or learning style which is partially culturally linked (Du Toit, 1990; Geldenhuis & Waterston, 1998; Lidz, 1987b; Owen, 1992; Richter, 1992), among other characteristics, can lead to under-estimates of individual abilities and are generally linked to intelligence in some form and are encompassed in a model referred to as task-related beliefs. It has been evidenced that it is not only capacity which allows for cognitive and educational growth but the concomitant and subsequent realignment of knowledge structures which necessitates at least rudimentary understandings of one's own metacognitive processes including memory, comprehension, learning, linguistics and communication (Benjafeld, 1993; Biggs, 1985; Borkowski & Konarski, 1981; Das, Naglieri & Kirby, 1994; Feuerstein, 1972; Foster, 1986; Halford & McCredden, 1998; Sharratt & Van den Heuvel, 1995; Van Ede, 1995). Metacomponents also feature heavily in Sternberg's triarchic theory of intelligence, an instance of an intelligence theory from an information processing perspective which seeks to link factors beyond conventional ideas on intelligence variables (Sternberg, 1997b) and a theory which has great potential in linking back up to dynamic assessment predicates. Expectations of failure or success can also greatly alter the outcome of test or intervention (Bethge, Carlson & Wiedl, 1982; Castelijns, Van Werkhoven & Stevens, 2002; Lategan, 1991; Pressley, Van Etten, Yokoi, Freebern & Van Meter, 1998; Resnick & Nelson-Le Gall, 2000; Stankov, 2004; Yussen, 1985). Feuerstein's theory of mediated learning experience specifically is viewed as an instrument of metacognition in that it offers both didactic and explanatory indices of individual behaviour (Birnbau & Deutsch, 1996). Interestingly, Brand, Egan and Deary (1994) make a compelling argument for the influences of intelligence on personality and of course this raises a whole new research debate.¹⁷⁰ Non-intellective factors such as intrinsic motivation, locus of control and achievement motivation impact on the test situation, learning task and on test results (Boekaerts, 1988; Cordova & Lepper, 1996; Tzuriel, Samuels & Feuerstein, 1988).

Baron (1998), in his attempts to define intelligence, distinguishes between what he refers to as capacities and dispositions, the former tied to heritage and the latter being an aspect over which one has more control hence the need to work in synchrony. Metacognitive strategies come to the fore in certain dynamic assessment batteries where extensive training focuses in on this very aspect of intellectual functioning and not necessarily specific cognitive skills (Losardo & Notari-Syverson, 2001) although metacognition research has generally been confined to learning and memory research (Hertzog & Robinson, 2005). There is documented evidence for the neurological explanation of executive functioning of this sort which is yet another compelling reason to consider certain reductionist approaches towards the study of metacognition within dynamic assessment (Case, 1992; Davis & Anderson, 1999) as it has been evidenced that frontal lobe damage severely impedes metacognitive reserve (Hertzog & Robinson, 2005). Being aware and knowledgeable of one's own problem solving abilities, cognition, knowledge and control over the cognitive domain (Bruer, 1998) means that a challenge is more easily overcome if strategies utilised are better suited to the person's cognitive repertoire and includes encoding information, selecting the appropriate plans for action and identifying aspects that may impede progress (Davidson & Sternberg, 1998). This self-regulatory mechanism is particularly evident in expert approaches towards solving problems in certain domains when compared to the solving of problems by non-experts (Glaser & Bassok, 1989); which can tie back into dynamic assessment's teaching strategy where previously non-expert individuals become more knowledgeable about their own functioning as well as becoming aware of necessary task strategies. As Luria (1994) simply yet aptly points out, the young child does not possess the same memory capacity as an adult, mostly due to the fact that the child does not know how to use memory and has less to do with the fact that memory capacity exists (i.e. potential). These aspects, along with the ability to transfer these skills to other settings are considered paramount in dynamic assessment (Schlatter & Büchel, 2000), and as many interventions are conducted with children, metacognition takes on a leading role in how such knowledge of one's own skill develops; an area still being researched (Carr & Biddlecomb, 1998). The inability to self-regulate cognition within university entrants in South Africa is of particular concern as these students encounter difficulties with both cognitive and metacognitive backlogs (Craig, 1990). Older populations (over the age of 65) also make use

¹⁶⁹ Metacognition has only really become of academic interest since the 1970's (with the coining of the term "metamemory" by Flavell in the 1970's; Campbell, 1993; Garner, 1988) but has already been explained in various cognitive models such as information processing, cognitive-structural, cognitive-behavioural and psychometric models or paradigms (Yussen, 1985). Its richest area of applicability is usually found within developmental psychology, (Nelson, 1998) hence its pertinence to dynamic assessment. However, as a notion it has a rich history predating modern times and stretching as far back as antiquity (Gleitman, 1985) and was clearly identified by Binet and Spearman as paramount ingredient in intelligence (Brown, Campione, Webber & McGilly, 1993). Metacognition has also been considered a separate approach within the "learning to learn" and "thinking about thinking" paradigm co-incident with the more dynamic assessment-like approaches which emphasise executive functioning of the person's thinking style (Bondy, 1987; Griffey & Claxton, 1997; Lloyd, 1995).

¹⁷⁰ Thinking that one is intelligent to a degree further inspires one to tackle tasks otherwise not approached resulting in a mixture of positive and negative experiences but possibly motivating one to succeed at ever higher levels of functioning. However, the flip side to this argument would be that one would have had to possess the knowledge of one's intelligence (thus being intelligent in the first place) in order to make this assumption. Yet another instance of blurred cause and effect. Their rationale and research results evidence that intelligence could well be viewed as a personality trait (Morgan, 1996), in which case the manner of assessing for intellect could differ quite substantially.



of self-regulated learning and as the population ages (at least in many Western countries) application of models which are currently focused on children will need to be increasingly shifted to older individuals about whom not much is known in terms of identification and assessment of learning difficulties (Dunlosky & Hertzog, 1998; Samuels, Lamb & Oberholtzer, 1992). Recall dynamic assessment's application within older populations. Aspects such as bravado, mischievousness, and obedience for instance correlate in both negative and positive ways with performance on various dynamically assessed tasks (Budoff, 1987a) and are heavily integrated into mediatory models of dynamic assessment (Jensen, 1992). Intelligence can be viewed attributively where judgments about ourselves and others are in fact what constitute intelligent behaviour (Goodnow, 1998) and such a view resonates with social aspects of cognition and intelligent functioning as identified by Vygotsky for instance. Intellectually disabled individuals, for instance, report low levels of emotional worth as opposed to their more able counterparts and general behavioural style is considered by some to be independent of learning ability (Budoff, 1987a). Academic achievement is so much more than a quantified score on a test (Frisby, 2001) and as one reads the literature, it seems that this concept is well understood and acknowledged by many educators, psychologists and practitioners; yet the wheels of testing grind away slowly without much of this progressive thought seeping through at a fast pace (Suzuki, Ponterotto & Meller, 2001).

However, movements within assessment such as dynamic assessment, curriculum-based assessment as well as portfolio-type assessments are making inroads within various educational settings where co-constructed knowledge and teaching is occurring in the classroom (at least this is the principle as identified on paper¹⁷¹) (Lidz, 2001; Suzuki, Ponterotto & Meller, 2001). Granted, time and money are overwhelmingly important, perhaps most important in the game of assessment, but ultimately it is the individual testee who pays the price of inadequate assessment.¹⁷² Learning potential assessment approaches through their various mediatory styles are able to effectively cope with much of these adverse side-effects evidenced during mainstream intelligence assessment (Desoete, Roeyers, Buysse & De Clercq, 2002; Klauer, 2002; Meijer, 1993; Resing & Roth-Van der Werf, 2002; Ruijsenaars, Castelijns & Hamers, 1993).

2.9 Summary

Dynamic assessment is couched within intelligence research which is subsumed within the larger domain of psychology. In order to cast dynamic assessment in its proper light and in order to effect the meta-theoretical framework to be developed in chapter 3, it proved necessary to not only include a discussion on the fundamentals of this manner of assessment but also to plumb the depths regarding one very particular and very important chapter in its history and origin; that of its Soviet origins. Dynamic assessment is a perennially intuitive and appealing framework around which assessment issues have constantly interacted in some form or manner whether it was explicitly or tacitly admitted as such. A defining feature of this manner of assessment is its origins in chiefly non-Western countries where mainstream Western methods of assessment have either been banned or negatively looked upon due to the resulting appearance of low functioning individuals or due to reigning political ideology which via grand philosophising was not only considered deviant but dangerous. The need to somehow assess and assist (its key characteristics) low performing people necessitated an alternative approach which was both grounded in some sort of theoretical approach and amenable to wide-scale testing. Various themes emerged from a few core ideas that germinated during the early half of the twentieth century yielding parallel forms of assessment ideas that were both geographically and temporally separated yet shared a common theme stretching across these dimensions - a concern for fairness in assessment and a sustained effort to help those in need of assistance in a manner befitting more clinical-like modes of interaction. The process of engaging in matter to be learned and assessed is deemed a superior indicator of learning potential as opposed to static-based modes of assessment and is more in keeping with the valid construct of learning; a construct that static modes of assessment do not seek to engage with or measure in any way. Product-based approaches and their results manifest after material has been learned. In contrast, dynamic assessment is witness to the process that takes place during the learning phase. In allowing for situations which instigate change and thus direct change into the zone of next development, dynamic assessment assumes that such zones exist, although the width of such zones depends on the individual as well as what is being assessed. This brings into focus the need for domain-specific versus domain general assessments. The finer the skill and more complex the task the lower the efficacy of dynamic assessment's mediatory approach. This merely highlights one

¹⁷¹ The South African department of Education's initiative in the so-called "outcomes-based education" in which the student's learning process is increasingly emphasised as opposed to emphasis being placed on product-based assessment alone. Universities are still more attuned to the older style of assessment and when one deals with thousands of students, methods other than standardised testing are hardly much of an option. However, at tertiary education levels the intellectual levels attained are higher and are more easily reached than is the case with high school students for instance due to an increasingly restricted range of intellectual achievement. Hence, standardised testing may not pose as much of a problem. However, (and there is always an "however") the situation is not as straightforward as all that in South Africa for instance, where most universities have to implement some form of alternative assessment for those students who can be regarded as disadvantaged in terms of having had access to lower quality education. Time will eventually rectify this as more students attend school and facilities improve and so on. But until such a time is reached (unlikely to be reached in this author's generation) issues pertaining to alternative assessments still need to be addressed.

¹⁷² Perhaps it is ultimately both the individual and society who pay the price. Is it perhaps not better to rectify a problem at the outset as opposed to waiting for someone to travel through the system and eventually land up being supported via other means many years later? This is a very difficult debate indeed: for whose responsibility is it? The individual or society at large?



of a number of paradoxes currently pervading the territory. One reason advocated as to why this may be so is the lack of construct validity evident in so many other areas within psychology and is not only problematic for intelligence research.

Coming to grips with dynamic assessment requires knowledge of its historical and philosophical roots. These tendrils are varied and only a brief look into Lev Vygotsky's Russia transpired in an effort to place one origin of this manner of testing. In essence, Vygotsky did not denounce mainstream testing and was forced to reckon with unpalatable dictatorial powers which coerced so many of his colleagues into other areas of research. Perhaps, in hindsight it can even be said that had it not been for these circumstances, the model from which much has grown might never have come to light. Attendant to this, though, was Vygotsky's seemingly own under-emphasised notion of the zone of proximal development which was not a fully fledged theory. It has since inspired many researchers in the West however to continue in similar vein.

The field of intelligence assessment is aligned in different ways along different fronts with each attendant alignment seeking evidentiary results in support of *a priori* assumptions. Many of the results offer alluring conclusions which augurs well for all the various alignments. Intelligence research has offered a galathea number of results with more than its fair share of contradictory theory and evidence for the past one hundred years. Some notions have remained the same whilst others have been seriously revised. The staggering amount of contradictory conclusions from these studies borders on the bewildering and the scenario is not much different for dynamic assessment. There are trends within mainstream assessment which seek to amalgamate newer models of intelligence which consider and place greater emphasis on non-cognitive variables, the role of phenotypic-genotypic interaction effects as well as more emphasis on process than product, although the extent to which this has seeped into most test batteries is not reflective of the gains made in acknowledging this. Likewise, dynamic assessment's tenets and founding fathers, although grounded in what has been described above, have not decried the use of mainstream intelligence testing nor have they advocated its disuse. Moreover, mainstream assessment methodology attests to greater reliability and validity whereas its dynamic assessment counterpart method cannot be said to have developed as robust a repertoire of techniques as has yet although there are new techniques propelling this area. There are areas of serious concern within dynamic assessment methods towards the understanding of change within testing and learning just as there are severe shortcomings within mainstream intelligence testing.

Dynamic assessment's concern with educability and trainability of individuals is underscored by the various assumptions regarding child development, specifically cognitive development. There is a traceable link between Vygotsky, Piaget and Feuerstein and their respective theories and models infiltrate modes of learning potential assessment with various differences and similarities. Perhaps one of the major issues within the field as it touches upon development theories is the degree to which humans are collaborators in their construction of the world as opposed to operating on the environment due to inherent timings which are genetically controlled. Currently this debate is not as strongly juxtaposed as it once was with most human development researchers agreeing that environmental press elicits or suppresses various developmental instances which are then set upon a course altered by the attendant environment. Piaget's theory was not a mechanistic treatise suggesting that the child develops only in strict accord with genetic developmental pacing but placed emphasis on the continuing interplay of both environment and genetic determination. Vygotsky's similar notion of environmental concern was more directed upon mediators in the environment which could helpfully provide enough "push" for the developing child so as to bring about growth which may not have occurred had such mediation never taken place. A continuous striving towards one's zone of proximal development results in growth and adaptation which need not cease at the end of childhood. This is reflected in the numerous areas of application within dynamic assessment.

Dynamic assessment is currently receiving much attention in the literature but this is not paralleled in practice where costs, timing and lack of training are often cited as the main reasons for the lack of inclusion of this manner of assessment, particularly in cash-strapped schools where personnel shortages and lack of funds militate against its use. There has been a consequent backswing of the pendulum in this regard with more tests being aligned along these very constraints which unfortunately results in a move further away from the initial goal of dynamic assessment. A positive feature of dynamic assessment methodology, however, is its continuing appeal to teachers and school psychologists alike who favour the approach or at least prefer to see both static and dynamic assessment methods being practised due to its all-inclusive nature and generally intuitive appeal.

The issue of placing dynamic assessment and intelligence within a meta-theoretical framework illustrates a three-fold problem; firstly, the field of intelligence, although overflowing in empirical studies and cumulative results over the past century has developed only partial theoretical models in keeping with results produced from such studies. The strain of competing models and theories brings into question the legitimacy of theoretical and empirical constructs (and attendant validity issues) within the domain of intelligence research (although this situation is not limited to this field alone). Secondly, dynamic assessment shares this concern regarding its various definitions and implementable strategies of which there are a number. It is partially aligned with intelligence research and shares a number of core assumptions but departs radically from a number of these core assumptions in other ways. The interwovenness of the two, commonly referred to as static and dynamic in turn has as a result that they usually operate in tandem but various movements within dynamic assessment history has pushed for the separation of the two in some quarters. One can consider dynamic assessment as leveraged upon intelligence or as developing in parallel



with it. Once again, this state of affairs is not limited to this particular area within psychology but characterises many sub-domains within psychology. Thirdly, meta-theory is seen by some as timely for certain areas of research and untimely for others; the placement of dynamic assessment within intelligence research within the broader framework of psychology will thus be greeted with both enthusiasm and dissent.

Successful merging of behavioural and biological correlates of intelligence and intelligence functioning is currently a trend within the field of intelligence research. This potentially fruitful avenue might well be worth pursuing in terms of dynamic assessment research. This does of course stand in stark contrast to the very reasons why dynamic assessment is not being implemented to the extent that it should be and would most likely not be feasible nor practical at this time in its course; but is something to consider for its future existence. This issue looks towards levels of description necessitated by different views on intelligence, development and growth in future endeavours to achieve potential in zones of proximal development. These developmental “leaps” are more often than not contextualised within the cognitive education domain, hence the emphasis on cognition within the intelligence discussion. Cognitive models, as with intelligence theories too are bound by reigning methodology of investigation. Depending on the nature of enquiry and tools utilised to do this, varying models come to the fore allowing for models of cognition to flower from different angles (computational, information processing and so on) which can serve and do serve as foundations for theories and models underlying dynamic assessment and intelligence. Some of these are now only of historical value whilst others are highlighted for their reality-endorsing views concerning human development (computational models which are better able to offer true-to-life renderings of brain functioning such as connectionist work). The section on dynamic assessment and intelligence draws to a close with a look at the increasingly dominant role that non-cognitive factors are starting to play within mainstream considerations of intelligence. Such concerns have been omnipresent within dynamic assessment since its inception as original notion, idea, thought, model and theory and in a manner of speaking, intelligence research has had to “catch up” to areas such as this. Dynamic assessment’s emphasis on broader culture and contextualisation of individual functioning has allowed the model to move forward and develop in this terrain and can now offer something substantial by way of theory to the intelligence domain within which it is situated. Likewise, dynamic assessment can evolve into more robust scientific method by adapting methodology from intelligence research, specifically in areas such as behavioural genetics and biological renderings of intelligence constructs. Bridging hypothetical and empirical constructs however, remains a thorny issue for both dynamic assessment and intelligence research, but not insurmountable.

2.10 Conclusion

Chapter 2 introduced basic philosophical concepts of import to this study and aided in situating the various discussions. The treatise attempts to deal with the subject matter in an objective fashion but due to the nature of human and social science research the pinnacle of such objectivity is hardly likely to be achieved with the tools at researchers’ disposal and although the toolkit within the scientific arsenal basically equips social science researchers to attempt rudimentary knowledge acquiring initiatives, there is a sense of the situations’ hopelessness in this regard. However, this does not negate efforts at endeavours to strive for as much scientific legitimacy as possible which is precisely the aim of this treatise as far as it’s underlying scientific enquiry is concerned. Couched within the above is a veiled allusion to the author’s own preoccupation within the knowledge acquisition enterprise. These opinions, although supported and argued for via cited works remain at the core of this study and permeate the narrative throughout. No serious scholarly opinion can enter into the arena of informed debate without the requisite foundations being fairly entrenched for such deliberations which is why the section leading up to dynamic assessment within intelligence was necessitated. The author entreated the reader to enter into the foregoing discussion with certain assumptions in mind and included a reference to the time and place in which the work is written, the philosophical leanings within the study as well as further discussion into various ontological and epistemological matters. Brief attention was paid to what can be considered seven crucial areas within intelligence research; the mind-body problem; consciousness, *g*-dominated vs. multiple intelligence dominated leanings towards the understanding of intelligence; emergence (irreducibility) vs. reductionism; realist vs. relativist approaches towards research; nature-nurture and static-dynamic conceptions towards the assessment of intelligence and potential. The emphasis within these discussions centred on argumentation on varying points of view. An attempt at well-rounded and fairly debated understandings of these issues was presented. The chapter focused more on introducing dynamic assessment both from a fundamentalist and historical point of view with emphasis on Vygotsky’s Soviet origins as to how this played forth in his thinking at a time in what can only be described as Russia’s turbulent past. Current awareness of dynamic assessment was perused as this informs the continuing debate surrounding its eventual use or disuse within intelligence assessment research. Lastly, intelligence as domain within psychology was discussed with particular concern for dynamic assessment’s place within the broader field. Now that the foundations have been cemented by way of introducing “dynamic assessment and intelligence” place is made for the construction of a meta-theoretical framework to house the above-mentioned concerns.

Chapter 3 Delineation of the meta-theoretical framework

3.1 Introduction

To mindfully place dynamic assessment and intelligence into an explored meta-theoretical framework, meta-theory as a concept and received view of theorising will now be considered. However, before meta-theorising can commence, surety is sought about the meanings inherent in the concept of theory and so down the hierarchical scale through to models and frameworks. These all have as a basic point of departure some sort of theory-laden assumptions of what it is these frameworks, theories and models seek to investigate. Theories, models and frameworks pervade all areas of human understanding, and are not relegated to only the natural and social sciences¹ but to the arts and humanities alike. Theory development progresses through each school of knowledge in disparate ways often resulting in heated debates between “opposing” schools of knowledge-procurement. As disciplines differ in their areas of knowledge gathering, so to do their methodologies. Once theories coalesce into more meaningful units of known facts within any discipline, meta-theory may serve as foundation upon which newer theories are anchored. It is to these issues that focus is now turned, initially as pertains to the subject matter of the history and philosophy of science and social science and then more specially as it pertains to dynamic assessment and intelligence within psychology.

Whether it is accurate or not to state that dynamic assessment finds itself in a state of crises regarding underlying aims, theory and model development, it is maintained that it is doubtful whether this field can rise to greater prominence without re-evaluation of its basic philosophical stance on a number of issues (Hughes, 1980). Philosophical issues pertinent to such enquiries include assumptions which are implicitly held but are not explicitly stated, implications of models which are incongruous with purported points of departure, fundamental issues which are not questioned,² assumptions which are believed to be commonly held by researchers within a field but are in fact not, misconstrual of terms, unclear usage of concepts, differing methodologies of inquiry (Fiske, 1986) and blurring of the boundaries of investigation into various areas of research within the field. The afore-mentioned list would, at first glance, appear harsh and unfair in the sense that given the nature of social science research, such vague, ill-defined and broad renderings of the meanings of research (Smedslund, 2000) is in fact very characteristic of this area of investigation into human beings. A contentious issue at the very least this may be, but it is one such issue for which some attempt at resolution must be made within dynamic assessment specifically if this field is to prosper.

3.2 Science

The fundamental question of how we know what we know (theory of knowledge) is the area with which epistemology concerns itself (Harré, 1988; Mouton & Marais, 1988). That any area as such exists (what there is) in the first place would be the domain of ontological discussion (Beatty, 1995; Delius, Gatzemeier, Sertcan & Wünsch, 2000; Fetzer, 2002; Hughes, 1981; Mautner, 2000; Scruton, 2004). Epistemology as a branch of study on its own does not posit forth a list of criteria delineating what is considered tools to knowledge but rather seeks to “describe the conditions which must be satisfied if there are to *be* any criteria of understanding at all” (Winch, 1970, p.21). Broadly defined, epistemology and ontology pertains to the theory of knowledge and how it comes to be that something can be known, if at all (Dennett, 1981; Faust & Meehl, 1992; Smullyan, 1981; Von Foerster, 1985).

This area is primarily concerned with the origins, principles, methods and limitations of knowledge and can be applied in very specific contexts (such as physics for instance) (Abercrombie, Hill & Turner, 2000; Eddington, 1939; Höffding, 1955; Delius, et al., 2000) although such ponderings were originally not part of the positivist agenda which seemed to claim that such metaphysical aspects were not important to theorising (Tolman, 1987).³ Philosophy of science may not necessarily question the nature of knowledge but epistemological enquiries into the nature of knowing does indeed question the very foundations upon which assumptions are made and as such the epistemological and ontological endeavour itself cannot be empirically self-validated (Hughes, 1981).

¹ One of the most succinct definitions of what science is, is that offered by Rogers (1992,p.495): “science is, among other things, scepticism of attitude” and is the closest to the definition endorsed by the author.

² Williams (2003) concurs when he states that psychology as a discipline is still struggling to come to grips with its subject matter: “we are engaged in work to which ontology is central, but we lack ontological commitment or even consensus” (p.2) yet Watanabe (2003) states that psychologists are indeed starting to acknowledge philosophical concepts “as not the objects, but the presuppositions, of empirical studies” (p.132).

³ Please note that the author is not being inconsistent here when it is stated that hers is a predilection for positivist-type philosophies, even though there is a necessary warrant for a meta-theory of sorts dealing with subject matter deemed unnecessary by these very same positivist philosophers. In a sense, it is a step further along the road from strict positivism to constructionism of a kind (although the latter is not endorsed).



By extension then, no epistemological or ontological method can be said to be valid and true.⁴ Also, no conclusions can be logically deduced from its own predicates. This is very similar to the axiomatic system of hypothetical-deductive models of explanations so often utilised within the natural and social sciences and which via a process of deducing causal explanations empirically verify (only increasing the theory's probability of being correct) or falsify hypotheses (Popper, 1965; Radnitzky & Andersson, 1978).

A relativism of sorts is not employed in this argument; but that an epistemological and ontological system cannot be self-validating is not equated with relativist philosophies, these two arguments are entirely separated. Seeking alternatives for the nomological mode of explanation does not mean that an argument succumbs to a relativist one. As highlighted in chapter 2, dichotomous views poised on a continuum is a rather unfortunate analogy to use when discussing alternatives to various issues and as such attenuations of the nomological model for instance does not necessarily result in the acceptance of a relativist framework. Dynamic assessment within intelligence research has developed within a nomological-deductive system of predicates and conclusions. As such, it has a history of measurement spanning the twentieth century which had as its foundations measurement ideals emanating from natural science models from the late nineteenth century. Unsatisfactory as this might be for some more progressive philosophers the situation can only be remedied with due consideration of what it is the various models of explanation seek to explain. Jumping from one extreme to another will only incur further philosophical animosity between the so-called ends of the explanatory continuum.⁵ The Newtonian models are not perfectly accurate (in fact they are incorrect in non-absolute space-time but for all intents and purposes on our macro-scale they prevail in today's space programmes; Greene, 2004) and neither are psychological models; which as with statistical models utilised in psychology are approximations to reality only (Browne & Du Toit, 1991). A pertinent example is the Rasch model of probabilistic measurement applied with great success and since improved upon in the area of ability assessment (Rasch, 1980) (see chapter 4). Regarding such psychological models, the mathematician Rasch (1980) states "models should not be true, but it is important that they are applicable, and whether they are applicable for any given purpose must of course be investigated. This also means that a model is never accepted finally, only on trial" (pp.37-38).

Science, as some relativists would have it, is another form of religion, a system religiously followed and applied, assuming common understanding by like-minded peers resulting in publishable accounts of knowledge-gathering activities and a system incomparable in getting to the truth of the matter. This statement wavers between truth and falsehood and assumes a middle ground not unlike the grey murky areas in which science is at times practised. In contradistinction to religion yet playing on a similar sentiment a credo may well be "they are here not to worship what is known but to question it" (Bronowski, 1974, p.362). Of course the one time dominant version of what science purported to be was that it was just another form of common-sense and that the scientific attitude was merely an outgrowth of an attitude shared by most people in the first place; a way of pursuing ideas which were rational, testable and legitimate for further study (Schaffer, 1997). There are a number of inescapable facts which cannot be swept under the carpet, and one major bug-bear for social science as a whole is that sub-disciplines are purported to be scientific, as the case may be today ca. 2006. That this will forever be the case is an historical anecdote with which future generations will have to content themselves. What trajectory psychology as sub-discipline and intelligence as further sub-discipline will follow is anyone's guess.

That it vies for current scientific status is not in question; unfortunately how this term "science" is construed and interpreted is open to question. Psychology is a fragmented (or faculty) discipline⁶ (Koch, 1985; Scarr, 1998b) just as the study of intelligence is and has been (Detterman, 1979) so much so that many a psychologist will be unable to converse with another so-called

⁴ As it is captured in its own system in which primitives are accepted in order for consequent statements to be made; of course these primitives cannot be defined within the system it purportedly explains as one of two consequences emerge, namely, definitional circularity and definitional regress in which the former merely uses the same language to define other states and these states are in turn used to define them and the latter in which new definitions will have to be continuously sought to define a new word or state and in order to adequately define the new word or state another new definition is employed and so *on ad infinitum* (Fetzer, 2002). Dictionaries are definitionally circular but we can get by with them.

⁵ It is perhaps easier for the human brain to conceptualise of a linear continuum, a line spanning the extremes of view points. Is this truly the case though? Perhaps this analogy has failed as mechanism of explanation and as such does not further the cause of viewing particular views as sharing many features. Perhaps overlapping Venn diagrams would be a better-suited alternative.

⁶ A clinical psychologist schooled in qualitative narrative techniques might not understand the importance that neural networks play in cognitive science studies of the brain; likewise, the neurobiological psychologist might be similarly at a loss to converse with the clinician about single case-studies dealing with life experiences from an ethnographic point of view. Yet the two are both psychologists; the chasm it seems is too wide for both types to be included in one single discipline (a chasm which is becoming increasingly fragmented, "because it addresses increasingly isolated subspecialties and because it fails to take into account the theoretical themes and problems that motivate the enterprise of psychology as a whole" (Slife & Williams, 1997, p.118). It is routinely accepted that disparate disciplines such as biochemistry and art history for instance have little in common and do not share the same language of research; but it seems that inter-disciplinary fragmentation too is the order of the day. Blurred as this analogy may be, the case for dynamic assessment in intelligence could find itself in a similar situation, with clinically-oriented case study approaches vying for acceptance in stark contrast to computer-mediated statistically valid and highly replicable approaches. Clearly the two approaches spring from fundamentally different philosophies. An extreme view would be a calling for the dissolution of psychology as a discipline and to "reapportion its empirical results to other sciences that carve nature more closely to its joints" (Kukla, 1992).

psychologist. In a manner of speaking, the two are at ends of a continuum which appears to be evidencing an unwholesome crack. Figure 27 illustrates just such a continuum.

Figure 27 (i) Placement of psychology as discipline in between varying poles of methodology and areas of concern (Royce, 1973, p. 16)

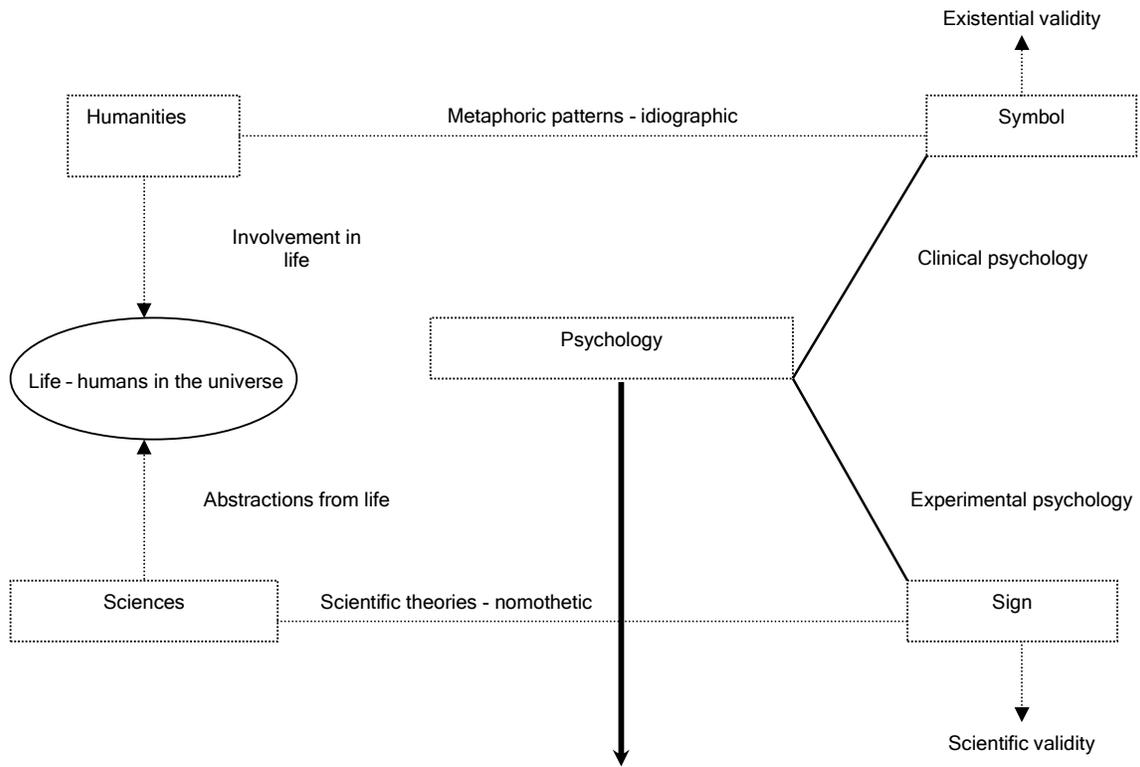
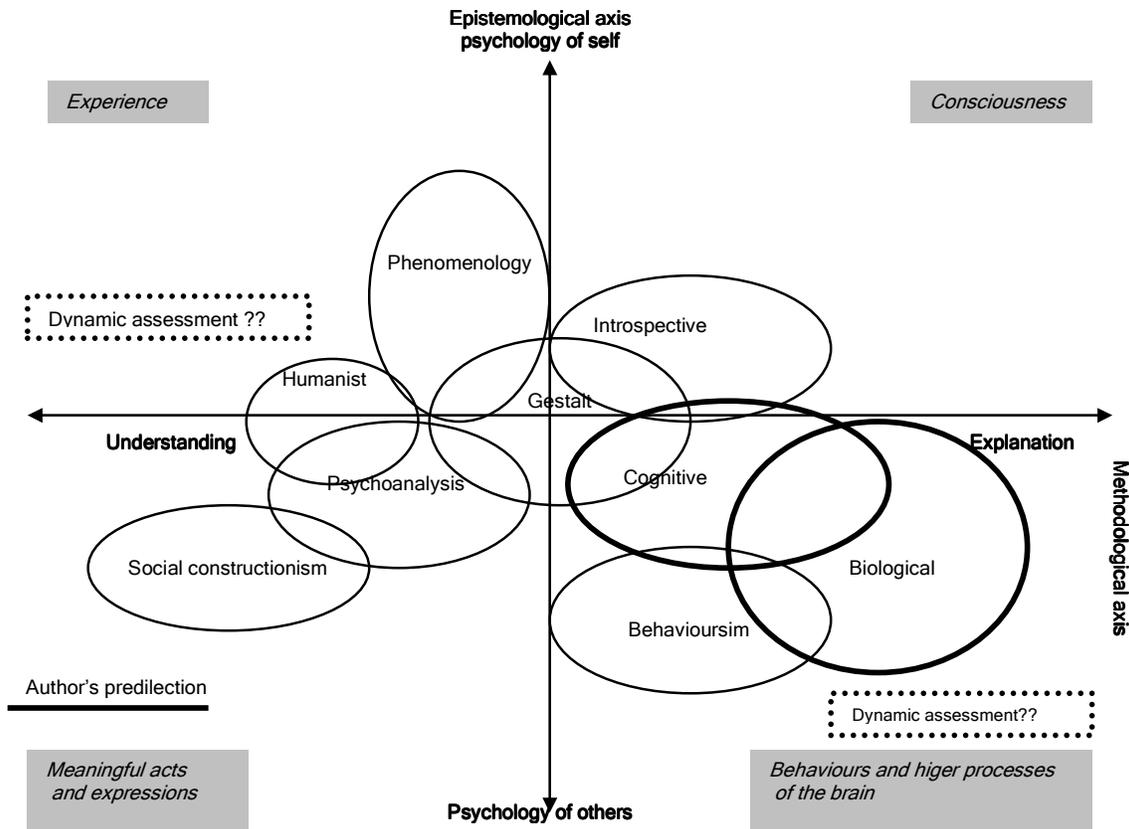


Figure 27 (ii) Trends in psychology with placement of dynamic assessment along axes of epistemology and methodology (Watanabe, 2003, p. 131)



Mere conjecture hints at a future with myriad research areas in which the current discipline of psychology will have dissolved into subsumed sections of other disciplines. No value statement can be added to this sentiment; it is akin to the question of whether humans are born good or bad, they are neither - they simply are. This state of affairs should neither shock nor please anyone as such, for it is a statement of the continuing development and progress (also not to be value-construed) of this particular branch (or branches as the case may be) within the social sciences.

Formal recognition of science as a rigorous approach towards the understanding of how things are cannot be exactly dated⁷ but can be divided into specific periods spanning time since Greek scholars, flowing along through the Renaissance, Newtonian revolution, the Enlightenment, Darwinian era and in onto the modern atomic era (Gribbin, 1998, 2003; Schwartz & Bishop, 1959a). Lest this list seem somewhat too aligned to physics, astronomy and biology it cannot be underestimated that these explorations and discoveries are due primarily to the luminaries working in these areas throughout these various stages (Asimov, 1987; Ergang, 1967; Hawking, 2003; Schwartz & Bishop, 1959a, 1959b; Uglow, 2003). The trajectories on which they placed ensuing science⁸ and the preceding as well as subsequent technological innovations intermingled within the history of science (Cossons, Nahum & Turvey, 1997; Segré, 1980; Snow, 1981; Weber, 1980) and also their impact on the path followed by psychology (Robinson, 2000) can also not be underemphasised. Recall also the number of luminaries within psychology who were originally trained within the natural sciences; and that the period from the 1930's till the 1940's was deemed psychology's immature physics period (White & Pillemer, 2005).

3.2.1 Explanatory mechanisms in science

Explanation in science can roughly be divided into three approaches, namely the deductive-nomological (DN model), causal and pragmatic approaches (Leahey, 2000). Also known as the covering-law model, the deductive-nomological model can be traced

⁷ Gribbin (2003), although conscious of the arbitrary dating of the start of the scientific method utilises 1543 as his starting point.

⁸ Of course for a more balanced and equal-assignment-of-credit view, Murray (2004) offers a tantalising romp through history and focuses on the luminaries within global philosophy as well the arts.



back as far as Aristotle (Bechtel, Mandik & Mundale, 2001; Robinson, 1985) and has developed and been formulated into its modern form (1948) by Carl Hempel⁹ and Paul Oppenheim. It remains the model adhered to in current scientific and social science explanation (Bechtel, 1988; D'Andrade, 1986; Mouton, 1993a; Nagel, 1979;) as it seeks to not only logically deduce arguments and observables (via premises that state sufficient conditions for the truth of the explanation) but also to predict.¹⁰ Although utilised for theory testing it is not as useful for theory construction (Smedslund, 2000) but has proven its worth in areas such as physics in which logically-deductive arguments are predicated on sound laws¹¹ from which a deductive argument can reach a conclusion that is also sound and perhaps more importantly, valid (Woodward, 2002).¹² Nagel (1979) maintains that probabilistic explanations can be construed as temporary halfway stations to the ideal of deduction in which explanatory premises contain statistical assumptions about elements while the resultant explanation contains just a singular statement about an element. A well-known theory of this DN type, considered also as a "strong"¹³ theory is that of Newtonian gravitation (Eysenck, 1985). It ties in well with positivist thinking whilst offering explanatory functions and its results can said to be useful. This type of explanatory model is satisfied with predictions and law-like control of nature (considered as the main aims of theories in general; Reynolds, 1971) yet does not have as a goal explanations into how and why things work the way they do. The tension arising from the need to predict as opposed to explain has permeated through to the foundations of test theory and measurement within psychometrics (see chapter 4) (Das, Naglieri & Kirby, 1994; Meier, 1994). However, lest we should go astray and erroneously think that empiricism has been the goal always striven for it should be noted that Newton's law of gravitation as well as Mendeleev's periodic table are conceptualisations based partially on presuppositions and hypotheses (hence *a priori* ideas) (Emsley, 2003; Niaz, 2005) mingled with experimental data (the quantitative imperative; see chapter 4).¹⁴

The sharp distinction between pure empiricist natural science and fuzzy social science is a myth. This attribute is emphasised in causal models of explanations which seek something beyond utility; that of truth.¹⁵ Early crude empiricists maintained that the truth can be obtained via an objective study of phenomena, that this can be concluded with certainty and due to lack of error inherent in the deductive method the conclusions would be valid and objective and thus not open to interpretation (Doyal & Harris, 1986). Of course it is these very four issues which became rather problematic for some natural science disciplines and more so for the social sciences. As Secord (1986) argues, Bhaskar maintains that egregious errors emanating from just such a model creates more problems than it solves. Firstly, the world is not a closed system but open and thus extremely difficult to express in terms of law-like governance; secondly, observable regularities are confused with the abstract entities of science; thirdly, concomitant events are often misconstrued as cause and effect thus resulting in the incorrect inference that laws govern systems whereas they in fact restrain the system and lastly, the task of the social sciences is not to determine regularities in behaviour but to discover their nature and effects. Clearly psychology was following the dictates of a severely limiting and possibly altogether incorrect framework for conducting its own research.

Clark Hull's reliance on the inductive hypothetico-deductive method of theory development was strongly positivist influenced and thus influenced many psychological experimentalists. Taking "primitive" theoretical statements (although supported by physical fact) as his starting point he was able to manoeuvre through a system of protocol sentences and arrive at conclusions based on induction and hypotheses testing (of which the limitations will be discussed in chapter 4) (Reber & Reber, 2001; Sanders & Rappard, 1985; Turner, 1967). Testing in this model is absolutely contrary to Popper's view on testing (Sanders et al., 1985) but

⁹ Who later adapted the DN model to allow for inductive-statistical explanations in which events could be highly probable (Bechtel, 1988). And so the pieces of the puzzle are being put together, where in 2006 psychological assessment has as its historical origins logical positivism, hypothetico-deductive, inductive and nomological-deductive as well as statistical inferential foundations. It has a lot to contend with especially if it now seeks to re-establish itself within a broader emancipatory and critical framework. If this is where it belongs then so be it. If it belongs to the stratum of reductive psychology then so be it. The issue here is that it has to sort itself out before it can continue on a trajectory of stagnation. See chapter 4 in which Hempel's paradox is viewed from a mathematical point of view thus linking certain aspects of chapter 3 to those in chapter 4.

¹⁰ Turner (1967), however, questions the similarity of logical structure between scientific explanation and scientific prediction. Prediction is not only the hall-mark of the scientific quest but reflects an almost too human character trait as humans regularly employ methods of prediction - whether they be accurate or not (Ravenscroft, 1993). Scientific prediction then, is really only an extension of this intuitive behaviour. Thankfully, our intuitive appeal for prediction has matured into a matter of scientific prediction overlaying our inaccuracies in predictions which in everyday life do not consider prior probabilities inherent in the logic of statistical prediction (Kahneman & Tversky, 1973).

¹¹ Science is peppered with laws and include among others electromagnetic field theory of light, Richter's law of reciprocal proportions, Charles' law of gases, the law of combining volumes, the Zeroth law, the valence concept and the laws of mass actions (Bryson, 2004, p.153).

¹² It must be stressed that although this model appears logical and straightforward it does distinguish between the context of discovery and the context of justification in which the former need not be conducted in any logical manner at all (much of dynamic assessment's roots are not so-called logical discoveries) but in which the latter is placed under scrutiny and it this context of justification with which the positivist model of deductive-nomology concerns itself (Bechtel, 1988).

¹³ It is considered "strong" due to the large number of independent observations made over time which have concluded similar findings (Eysenck, 1985).

¹⁴ Niaz (2005) draws a further comparison between the strategies deployed by Galileo and Piaget; both utilised idealisations based on presuppositions characteristic of constructivist-rationalist theories as opposed to the naïve falsificationist strategy employed by the quantitative imperative - an issue to which we shall return in chapter 4.

¹⁵ Something which is hardly ever attained nor attainable in the social sciences (Faust & Meehl, 1992). Is this really the case though? The supposedly "purest" of scientific disciplines, namely mathematics, is itself not consistent within its own tenets of truth or indeed can even determine what a truth might be; in this regard see chapter 4's dealings with the mathematical realm.



as Meehl (1990) fervently states, Hull's now dead theory did highlight the lengths to which research psychologists needed to go if indeed their chains of logical deduction were to in any way be considered rigorous. If this monumental effort was necessitated by a "simple" theory of rote learning what hope for the more complicated and auxiliary-filled "soft" theories permeating psychological discourse? Along with this psychological experimentalist trend, inferential statistics became institutionalised in American experimental psychology during the 1930-1950's paving the way for novel cognitive concepts to arise out of various statistical inferences from experiments (Gigerenzer, 1991; Porter, 1997; Schmidt & Hunter, 1997). A reason given for the use of such generalising statistics is due to the inherent complexity of the social sciences (Lykken, 1991) in deriving exact causes and consequences of various behaviours, thus relying on statistical aggregates (Nagel, 1979). This so-called monumental confluence of hypothetico-deductive, nomological-deductive and inductive method of hypotheses confirmation or disconfirmation along with the overwhelming influence of inferential statistics¹⁶ and the need for quantification arising from the empiricist programme¹⁷ (Gergen, 1987a) as utilised within psychology, led this discipline down a path of scientific technique enabling (or granting) it a status alongside that of natural science disciplines (Rozeboom, 1995). Surely this is hard-core science? So it would seem. Throw into this mix strides made in cognitive psychology and cognitive neuroscience (with all its attendant problems) and one can immediately understand the attraction of reducing psychology to its more physical counterparts.

Causal explanatory models can be considered as traversing a step further along the path of understanding, as it applies positivist rigour with contemplation of how truth can be attained and does not merely focus on the attainment of law-like structures purporting to control and predict (Leahey, 2000). In other words, correlations obtained within the explanatory system of nomological models can be turned into causal models within the latter approach and this model attempts a finer-grained understanding of the varied causal features in a system (Woodward, 2002). Criticism is again at this latter view, as it cannot objectively state that a system in its entirety can be known (akin to the relativist qualm that we can never truly understand ourselves if we are to be the investigators within the system; see chapter 2 in which this view is severely tempered by its own limitations and constraints). It is virtually impossible to trace all possible causal mechanisms working in and on the system. Pragmatic models of explanations maintain that explanations are neither correct nor incorrect but should be considered as temporary satisfactory answers depending on the nature of the context. In sum, deductive-nomological accounts of science are often construed as anti-realist in which the observation of a system is enough for its theories to be useful, whereas the casual-realist accounts are inferred from the observations of the system and hint at the truth of the system (Leahey, 2000).

3.2.2 Scientific theories

The nature of natural science theories can be broadly divided into reductionist, realist and instrumentalist accounts of explanations (Brody & Capaldi, 1968) and these demarcations of science method are also applicable to models within test theory in psychological measurement (Borsboom, 2005). Reductionist trends, as exemplified by physicists such as Ernst Mach¹⁸ (1838-1916), (although Mach could also be classified as an instrumentalist) (Campbell, 1989; Mautner, 2000) and William John Rankine (1820-1872) seek to reduce explanations to more basic theories of explanations and when applied in very specific contexts afford a workable manner of explanation (Meehl, 1986). When taken over into the social realm however, problems arise. In its most extreme (and at times absurd) form, reductionism strives to explain phenomena in ever-decreasing scales of explication (Gell-Mann, 1998; Seager, 2001; Wimsatt, 1986) and in so doing explains aspects at one level of description in terms of another level (Dupré, 2001; Hempel, 2000; Quine, 2000; Ryan, 1970) hoping to have as a result the unification of science (Cronbach, 1986; Eysenck, 1987; Hooker, 2001). This not only occurs across disciplines but within disciplines too such as the ever finer grained or reduced scaled version of the particle nature of reality (in which tinier particles are explained within even briefer intervals of time; Gleick, 1991¹⁹). Human behaviour can be traced to neuronal firing patterns evidencing sequences of such-and-such a nature (Wilczek, 2002), in turn these firing neurons are controlled by neurochemicals which are themselves constituted from molecules (Kalat, 2004) which are in turn made up of atoms which are reduced still further to quarks which can be reduced into vibrating strings,²⁰ and as with many a slippery-slope argument, the initial statement is bounded to the last statement thus concluding that vibrating strings are related to human behaviour. This is denoted as being rather absurd but similar attempts to simplistically reduce human behaviour are the hall-marks of extreme reductionism. The interactions between microlevel patterns of neuronal functioning and macrolevel behavioural responses has been posited as a neuronal theory of personal identity (akin to the reductionist trend to identity consciousness with brain occurrences) and has been employed as explanatory mechanism to describe neural Darwinism (Sacks, 1995).

¹⁶ Is it true that "numbers, massaged with statistics, become facts" (Gazzaniga, 1995, p.165)? Is this psychology's legacy? Should it be?

¹⁷ Leading psychology "to an inverse relationship between the growing precision of its analysis and the shrinking significance" (Shames, 1987, p.26). Psychology also took its lead from the logical positivists and paved the way towards operationism as exemplified by S.S. Stevens (see chapter 4) in psychology (Grace & Ferreras, 1998).

¹⁸ To offer a rather tentative but direct link with one of psychology's founding scientists, William James in fact encountered Mach in 1882 and was quite impressed with his expansive knowledge of most things (Bernstein, 1993).

¹⁹ How far can physics take us on a journey through lower-level descriptions? What lies beyond string theory and its latest incantations such as M-theory (Kaku, 2006)? These are indeed exiting questions. Sterile neutrinos are the latest of a long line of particle offings (Muir, 2006).

²⁰ String theory (Greene, 2004; Kaku, 1999); a proposal for uniting "quantum and relativistic physical theories that have resisted, since Einstein's day, all efforts to knit them into a unified theory" (Hébert, 2005, p.488).

A tempered view consists of taking the behavioural aspect and treating it as an already reduced phenomenon akin to a physics-envy view of sorts (Midgley, 1999) although this too runs into its own problems. The boundary between ontological and epistemological reduction needs to be drawn as the case may be with physical entities such as real-world items (ontological realism might posit that reality is at its most minimalist a real thing and is explicable in terms of fundamental physics) as opposed to representational items such as theories or frameworks (epistemological, in which theories and laws in one science can be explainable in terms of laws from another science) which are reduced (Silberstein, 2002; Tipler, 1996). Are we seeking to reduce behaviour to brain or a theory of behaviour to a theory of brain? If so, according to methodological realism, this would be a good path to follow (Tipler, 1996). The two are distinct and this study focuses almost exclusively on epistemological reduction, in other words studying theories/models as they pertain to dynamic assessment within intelligence. This is the case with some evolutionary psychologists positing just such a reduction from behaviour to biology (mind-brain to adaptive functioning),²¹ invoking lawfulness where there perhaps is none, after all biology is a case in point in which few laws are in fact known (Looren De Jong, 2003). It was thought prudent at the outset (chapter 2), however, to argue the author's predilection for ontological reducibility.

Realist trends within physics are exemplified by the theories of Arthur S. Eddington (1881-1944), Werner Heisenberg (1901-1976) and the philosopher of science, Emile Meyerson (1859-1933) for instance. Realism purports to verify, develop and prove for and within theories actual real world entities that can either be directly or indirectly observed or at least inferred (Erwin, 1985). It seeks to describe reality (Preston, 2001) in as truthful a manner possible seeking this truth from phenomena accessible to experience (direct or inferred) (Fetzer, 1993) and to make manifest (at least theoretically) that which is unobservable (Achinstein, 2001; Trout, 2001a). Some philosophers (Van Fraassen for instance) temper this strict realist notion by stating that science should aim for empirical adequacy only and that there need not be a complete correspondence between theory and empirical models (Giere, 2001b; Sklar, 2001). This account of reality is undoubtedly successful, as technological advancement has evidenced (Leplin, 2001) which cannot be denied; what is in question is its application within the social sciences. Broadly, realism can be sub-divided into among others (Mautner, 2000)²²

- category independence (subsumed within this are realist trends such as ontological, conceptual, scientific modal and moral realism) which posits that reality as such exists independently of our knowing
- semantic realism in which every declarative statement has a truth value even though we might not know this value
- epistemological realism in which mind (read brain) and world co-exist in explanation

As progress in technological advancement has shown, what was once considered a representationalist view of a realist account of nature (in other words certain subatomic particles were thought to exist as hypothetical constructs only) has now been reified into manifest entities (these said subatomic particles actually exist), in other words, some realist accounts of science which were previously representationalist are now manifest reality (Hacking, 2000). This increasing empirical content within science, which sought to advance rationally, was at odds with the notion of science advancing as a psychological or social enterprise (Giere, 2001a). Realist accounts of psychological constructs such as intelligence propound that, as a construct, it exists in some manner which can be indirectly accessed via scores on intelligence tests. One might not be able to see or touch or otherwise sensorily perceive "it" but it exists as a relevant concept. It exists via its causal relation to instruments that find it (Borsboom, 2005).

Instrumentalist accounts within physics are exemplified by, among others, Nicolas Copernicus (1473-1543), Isaac Newton (1642-1712) and Pierre Duhem (1861-1916) (Brody & Capaldi, 1968). Philosophers included in this category include the pragmatists (and in some instances positivists, the movement from which it sprang) William James, Ernst Mach, Moritz Schlick, C.S. Peirce and John Dewey (Leplin, 2001; Mautner, 2000). Instrumentalist accounts of science view theories as neither right nor wrong but as tools towards better understandings of how things are (Borsboom, Mellenbergh & Van Heerden, 2004; McErlean, 2000; Misak, 2001). Subsequent technology arising out of instrumentalist theories are considered merely as artefacts and devices which themselves embody no value as such but are valued for their utility only (Tiles, 2001). Theories are thus denied truth-values (Musgrave, 1985) (according to strict anti-realists though as there are tempered views concerning this issue). They assist in prediction and are considered more for their utility value than for any supposed truth although in some instances their utility value is considered as truth (Lakatos, 1993; Mautner, 2000). Their lack of falsity or truth implies that they cannot be held up to refutation as they cannot be tested. They are instruments describing ideal situations not accounts of how things actually are hence they are not empirically observable (Nagel, 2000; Popper, 1993). In its strictest form, what matters most to the instrumentalists is the predictive value of their theories (if it works what does it matter whether the theory is a true account of what is actually happening?) (Healey, 2001; Kincaid, 2002; Michell, 2002; Newton-Smith, 2001b) and not the truth

²¹ Although there is much to attract one to evolutionary psychology (as a researcher who in fact leans towards a reductionist agenda, various arguments posited by the evolutionary psychology discipline are compelling), a case can be made in favour of a less restrictive account of human functioning in present-day society.

²² Other sub-divisions include legal, political and literary realism with which this study does not concern itself (Mautner, 2000).

nor falsity of the theory per se.²³ Such a view, seems to pervade the literature on intelligence assessment and research in which predictive validity is paramount to substantive theory progress, moreover, the particular semantics chosen in any particular theory of mental test measurement is likely to change depending on the chosen theory (Borsboom, Van Heerden & Mellenbergh, 2003). Instrumental accounts of intelligence for instance are not as concerned with the existence of the construct as are realist accounts of intelligence. Instead, the instruments utilised to interpret what manifests as intelligence are part of the theoretical workings of such accounts of intelligence. Truth is not of concern but the usefulness of the theory is of prime concern for instrumentalist accounts. Instrumentalist accounts “licence inference to as yet unobserved data” (Borsboom, 2005, p.103).

For instance, models of how forces may work in nature typically describe a system which is not a true system as found in nature. A rock hurtling through the air will slow down due to friction but in a pure model (not an ideal body) this is not the case which may illustrate all the forces acting upon the rock without that of friction. Likewise, in social psychology theories, “laws” of behaviour prescribe certain actions that will occur when certain conditions are met (Kelly’s causal attribution theory,²⁴ in which behaviour of self and others can be determined via causal mechanisms which can be precisely identified) (Baron & Byrne, 1997) but do not result under different conditions and so is not truly a law²⁵ as commonly understood in the deductive nomological account. Just as in the descriptions of ideal bodies in ideal conditions, theories in psychology too are ideal systems (Converse, 1986; Niaz, 2005). In order to make more law-like the psychological phenomena warranting explanation, the collection of large-scale statistical information was originally viewed as a step in the right direction (Danziger, 1990), securing for psychology a place on the stage alongside science. This is an issue with which researchers are currently saddled, namely statistical aggregation derived from an individual difference model of human functioning (itself being predicated on Galtonian psychology as opposed to other initial branchings within psychology as evidenced by Wundt and Binet) (Danziger, 1990). As opposed to realist accounts of occurrences in nature which maintain a correspondence of terminology and actual physical entity (there truly does exist such a thing as an electromagnetic field), instrumentalist accounts will acknowledge that any explanatory model will merely serve to elucidate the functioning of a system. In other words depending on the nature of the context, liquids can be treated as compounds of discrete particles but in order for other hypotheses to be granted explanatory status this same liquid can be treated as continuous (McErlean, 2000). Symbolic representation²⁶ of instrumentalist theories provide heuristic value of how to calculate and infer from and predict certain outcomes (Nagel, 2000) but reject semantic, epistemological and ontological realism in the conventional sense (Tuomela, 1973). Although there are innumerable laws in physics for instance, the same cannot be said of psychology which only evidences a few general laws (Reber & Reber, 2001). Not all instrumentalists (like with any movement or philosophical affiliation) can of course be blanketed under the same term and the world can be viewed simultaneously as sets of signs as well as real-world things (Hughes, 2001).

3.3 Social science

How natural and social sciences are bound or not is a contentious and heated debate. The arguments are even considered stereotypical (Gould, 2001a) and McErlean (2000) offers an illustration of how the two may be related in various manners. As (i) a continuum of knowledge gathering, (ii) building blocks, (iii) developmental branches and (iv) non-hierarchical overlapping schemes. Each of these illustrations have their own merits but do not consider the entire relation in its totality. It seems that an

²³ For instance, nature’s constants are very precisely determined but physicists are unable to determine why they yield the values they do; the fact that these constants cannot be explained yet form the edifice upon which science is based could be crudely considered an instrumentalist account of such constants (Barrow & Webb, 2005). To add fuel to the fire it is a perfectly plausible notion entertained in the scientific literature that so-called “laws of nature” may only be applicable to this particular universe that we happen to inhabit at this time. So, laws of nature too are a relativist notion, of course chapter 2 summed up the argument for a realist and objectivist stance regarding just such issues (Tegmark, 2003). If taken to its logical conclusion one might be forgiven for thinking that if even the bastion of all natural sciences is an inconstant and variable field of study what of the rest?

²⁴ Interestingly enough, Gigerenzer (1991) states that Kelly’s causal attribution theory and theoretical development proceeding it in fact followed a naïve manner of experimental design, a sort of folk Bayesian statistics in the brain (as opposed to Fisherian). In a way this makes sense, seeing as causal attribution mechanisms do indeed work, as theorised, in everyday life. Are we, as evolved beings, more attuned to Bayesian statistics as opposed to other types of statistics?

²⁵ Perhaps the closest approach to lawfulness of sorts achieved within psychology is the phi-gamma law pertaining to stimulus-response behaviour (McDonald, 1999). This law’s underlying function in addition to Spearman’s common factor model resulted in item factor analysis or commonly known today as item response theory (see chapter 4 for more on IRT).

²⁶ Ramsey sentences (a technique devised by Cambridge philosopher Frank Ramsey) are just such an example in which theories are “deflated” by virtue of eliminating concepts and circular definitions from the theories and replacing them with symbols or variables instead and in philosophy of mind mental concepts can be done away with entirely (Maslin, 2001; Mautner, 2000; Meehl, 1992; Nagel, 2000). Such semantic instrumentalism is thus without meaning (Tuomela, 1973). Theories are “axiomatised in first-order predicate calculus using proper axioms and hav[e] distinct observational and theoretical vocabularies which are related to each other via a dictionary of correspondence rules” (Suppe, 2001, p.390). Similar to this is Craig’s theorem which seeks to vindicate instrumentalist accounts of science by proving that theoretical terms are not essential in specifying a theory’s observational variables (Erwin, 1985; Hempel, 1970; Suppe, 2001b). Such predicate calculus along with syllogistic reasoning and inductive logic (exemplified by Mill’s principles of experimental reasoning and Bayesian statistical inference) all make for the hall-mark characteristics of positivist tools of science-practice (Shweder, 1986) which of course has plied its way through psychology and on into assessment which is where this sub-discipline now finds itself located. Ramsey was a modern exponent of the logicist movement within mathematics (Eves & Newsom, 1965) and first put forward “the subjective interpretation of probability theory” (from Bayes’ theorem) in 1931 (http://en.wikipedia.org/wiki/Bayesian_probability). See chapter 4.



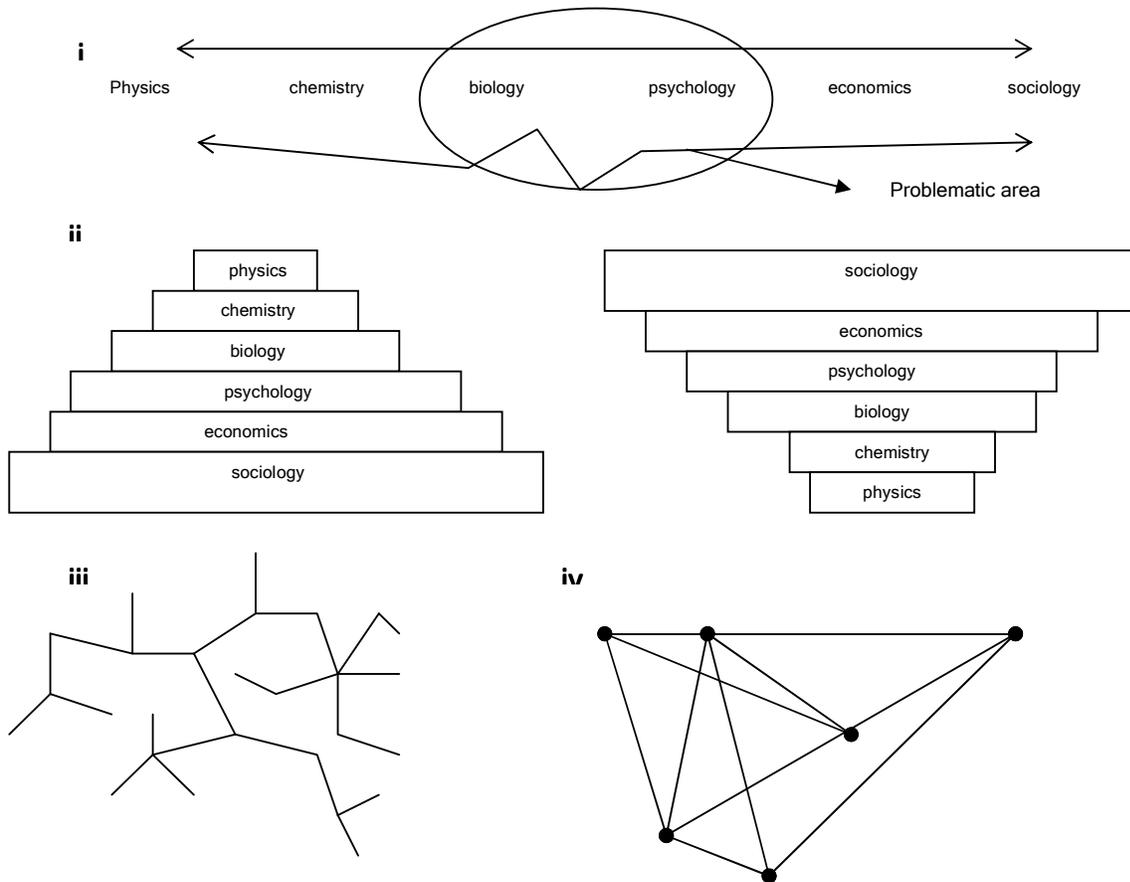
amalgam of these descriptions would best capture the link between the types of sciences. Both natural and social sciences can be layered in parallel as the following table depicts:

Table 5 Conventional hierarchical arrangement of disciplines, in which physics is considered the ultimate as reductionist explanatory mechanism (McErlean 2000)

Entities	Disciplinary parallels
Social groups and multicellular living things	Sociology, psychology, economics
Cells	Biology
Molecules	Chemistry
Atoms	Macro physics
Elementary particles	Micro physics

Economics is reduced to the psychological; psychology is reduced to the neurobiological; biology is reduced to the chemical; chemistry is reduced to the physical until such time that economic behaviour is treated as a result of cascading cause and effect. Why are the natural and social science so divided? It cannot be that the subject matter is in any measurable way to “blame” as such for the extent to which knowledge acquisition is scientific is not determined by the problem's origin (in this regard Wilson; 1999 firmly maintains that the social and natural sciences are indeed compatible). What can be apportioned “blame” is perhaps the manner of investigation (Sanders & Rappard, 1985) notwithstanding psychology's long utilised model of the natural sciences as *the* model (Van Strien, 1987). Therein lies the issue at hand. Are the social sciences really predicated on axioms of determinism (prediction and control), atomism (reduction and parsimony) and universalism (replication) (Kristensen, Slife & Yanchar, 2000)? Are new axioms warranted to the extent that the social sciences are no longer classifiable as science? Eysenck (1987) states that lack of theoretical foundation is in fact an “unverbalised axiom” of social sciences and underlies much of psychologising (p.49). Figure 28 illustrates the relations between varying disciplines; a manifold of knowledge acquisition seemingly disparate and at odds but also conciliatory if given the right conditions (Wilson, 1999). Gould's (2001a) sentiment wraps up the case for a stereotypical rendering of the polarity of both the hard and soft sciences; “we have, on the one hand, the “hard”, or physical sciences that deal in *numerical precision*, prediction, and experimentation. On the other hand, “soft” sciences that treat the complex objects of history in all their richness [which] must trade these virtues for “mere” description *without firm numbers* in a confusing world, where as best, we can hope to explain what we cannot predict” (own emphasis) (p.496).

Figure 28 Relations among scientific disciplines, (McErlean, 2000)



It can be stated that psychology should or should not be classified as a science but it can also conceivably be stated that as human beings undertake the study of physics, chemistry, geology and so on that science itself is a psychology of sorts which is why it becomes difficult to determine which “layer” of a discipline supersedes any other layer as is commonly understood to be the case (Stankov, 2005a) in illustration number ii above in figure 28. Essentially all scientific activity is undertaken by human beings and hence is “contaminated” by human flaws in whatever manner speaking. The retort to this might well be that despite all human failings and subjectivities, the colour blue can still be measured as 480 millimicrons on the light spectrum no matter what. The big question then is thus: can the social sciences be referred to as such? Firstly one needs to establish what it means to refer to something as a science. Depending on the definition one assumes this question can be answered in many different ways. Newell (1990) highlights the difference between theorists and theories stating that the latter are approximations yielding answers which the former do not do; “theory gives the answers, not the theorist” (Newell, 1990, p.13) but this can be disputed within the social sciences where knowledge is said to be constructed by the theorist. It is precisely for this reason then that chapter 2 was thought integral to this study, as the author’s leanings in many regards will undoubtedly influence the outcome of just such a study. Wilson’s (1999) treatise on consilience, although not free of critique conveys a strong message of the necessity to reconcile disparate and seemingly disunifying subjects of study and where natural scientists have made strides social scientists have lagged behind, “social scientists by and large spurn the idea of the hierarchical ordering of knowledge that unites and drives the natural sciences” (p.201). Gould (2000b) maintains that just this hierarchical arrangement or “the status ordering of the sciences has become so familiar a theme that the ranking from adamantine physics at the pinnacle down to such squishy and subjective subjects as psychology and sociology at the bottom has become stereotypical in itself” (p.278). Evolutionary psychology and cognitive science has been proffered as unifying or integrative disciplines for cross disciplinary social sciences (Kenrick, 2001) including serving as unifying theory within psychology (Newell, 1990) but others state that these subjects are themselves in need of taxonomic models (Bowman, Markham & Roberts, 2002)! These domains are not, however, without their detractors but when their subject domain is considered it is understandable at certain levels as to why this has been suggested. It is worth citing Lehman, Laird and Rosenbloom (2006, p.2) and their appraisal of Newell’s concerns for cognitive unification as it resonates so well with this thesis’ concern for similar unification or at least the building of a unificationist framework in which to serve as possible meta-theory for dynamic assessment



If we think about cognition as a big picture, then a microtheory is a way of cutting a portion of that picture into a jigsaw puzzle. Each theory may cut up its own portion, even a portion that overlaps with another theory's, in a different way. So, when each discipline throws its set of pieces out on the table, how do we know that there is any set of pieces that will allow us to recover the big picture? The only way, Newell argues, is to go ahead and try to put the whole picture together, to try to build unified theories of cognition (UTCs) (Newell, 1990).

Strict behaviourists assumed that laws had in fact stated generalities which could be replicated in other similar contexts such as the law of effects, with Thorndike later producing variants on this law (Reber & Reber, 2001; Thorndike, 1981). However, even this seemingly scientific law in the sense of strict correspondence between language and entity (extensional context) is not what it appears to be as the behaviour of many an organism is not a straightforward law-like one. Moreover this law is tinged with slight teleology (or "functionality", where the role an action plays brings about its goal, Nagel 1979; Van Hezewijk, 2003) and it represents a law more to do with the intention within the organism than the behaviour as such (Lillegard, 2004). Despite the law's hypothetical-deductive veneer it still is not strictly a law as understood in the natural sciences.

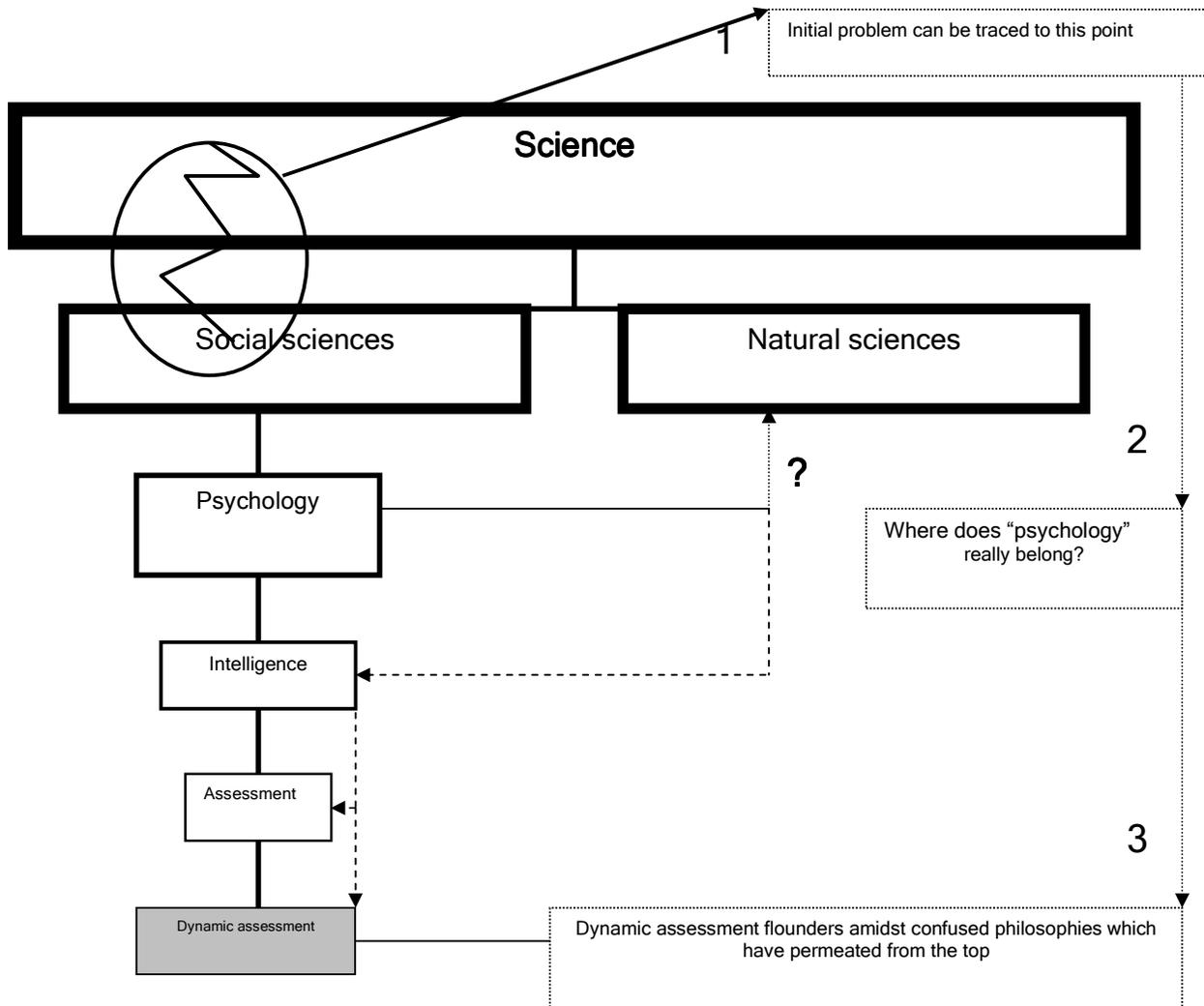
Once again it would seem that a dichotomous continuum looms ahead on which must be placed the practice of social research and the methodology utilised to do so. Need there be such a dichotomy? Social science in its currently accepted form will never truly be scientific in the sense that it pertains to the natural sciences, this much is clear; and yet, it need not conform to a strict natural science methodology in the first place. But then it surely cannot refer to itself as a science. Thus we have a divided camp of social studies researchers who profess to work in one and the same area but make use of at times opposing methodologies. Society demands of a discipline that it avail itself of predictive cause and effect mechanisms, as most often is the case, research is conducted in order to clarify how things happen the way they do and how better we can predict things in general. This is in addition to the pragmatic need to aid humankind to do its work in everyday life (Dilthey, 2000). Much of social science research does not even attain minimal levels of predictive success however much it admits of explanatory and understanding value (Kellert, 2000). Philosophical debates ensue however whenever it is purported that explanation inevitably leads to prediction, yet if much of social science research can stand up and be counted as explanatory mechanisms surely they fail somewhat miserably as predictive vehicles? This brings into question the deductive-nomological model of explanation and its supposed explanation-prediction symmetry (Kellert, 2000). Models of learning and the tests that are subsequently devised to test for learning and potential to learn are all predicated on some sort of model or theory which states its explanatory power, yet many of these devised tests are unable to predict results in any meaningful way. They explain but cannot predict and this, it seems, is odd. The question of natural science explanatory model fit within the social sciences becomes suspect; not the model per se but its application.

Rosenberg (2001) asks the simple question of why the methods employed by the natural sciences are so unsuccessfully employed by the social sciences and offers four tentative answers:

- i. the social sciences have not sufficiently progressed due to not having applied the methods well enough
- ii. the methods which the positivists employed for the natural sciences were in fact flawed from the start and these flawed methods were simply carried over into the social science arena
- iii. the social sciences are in fact on the right track, are using the correct methods and are progressing well enough but due to the inherent complexity of the subject matter progress is slow ("hypercomplex"; Wilson, 1999, p.202)
- iv. the social sciences have indeed made strides but different standards need to be utilised when judging the social science enterprise

In tracing back the philosophical dilemma in which it is hypothesised dynamic assessment finds itself, it is not unreasonable to suggest that the method of assessment is not entirely to blame for its predicament, if as has been stated above, the larger context in which it is placed is itself muddled and confused. The inability to distinguish methods and the difficulties which the social sciences in general face merely compounds the problems that lie ahead for the various sub-disciplines which are cloaked thereunder. This methodological and philosophical muddle seeps down through to the various disciplines. Exploring the meta-theoretical background and underpinnings of not only dynamic assessment within intelligence but the spectrum of the social sciences is warranted when it is understood that there is a basic tension underlying the social sciences. Psychology is merely one of the disciplines which finds itself on a tight rope betwixt and between various methodologies. Intelligence assessment is a subsidiary of psychology and dynamic assessment is itself a manner of assessment and so right down to the core of the approach this basic tension is found. How should this tension be addressed? From the bottom up or from the top down? Possibly both avenues can be pursued but it is contended here that a bottom-up approach would most likely succeed as individual disciplines will have to figure out a way of solving their own dilemmas and so individually manoeuvre themselves without being given a mandate to do so from above. Dynamic assessment in this regard can be viewed as being situated at the bottom of the pile and this is just as good a place to start sorting out fundamental issues as anywhere else. Figure 29 illustrates the level at which dynamic assessment is placed and how initial problems become magnified later on down through the sub-levels.

Figure 29 Permeating tension: how an initial higher-level problem becomes manifest at lower levels



Machlup (2000) offers a framework²⁷ from which comparisons can be drawn between the natural and social sciences and concludes with a score card enumerating the results achieved from this comparison. The grounds of comparison include the following:

- invariability of observations
- objectivity of observations and explanations
- verifiability of hypotheses
- exactness of findings
- measurability of phenomena
- constancy of numerical relationships
- predictability of future events
- distance from everyday experience
- standards of admission and requirements²⁸

²⁷ The astute reader might well note that the list itself is very “natural-science biased”; in other words conclusions can already be formulated long before any comparisons are made by mere fact that the grounds for comparison are criteria that are routinely utilised within the natural sciences as a given. The list could look radically different if viewed from an hermeneutic vantage point for instance. But the point within this chapter is to analyse just this issue: the social sciences as compared to the natural sciences as well as the practice of the social sciences within a milieu of natural science biased views and opinions.

²⁸ A slightly more pragmatic concern which, in a manner, stands apart from the previous criteria which are more general and abstract in nature.



Invariability of observations

Are natural science observations as homogeneous as conventionally accepted? Is invariance not present to a greater degree than is commonly accepted? Likewise, are observations within the social sciences as heterogeneous as conventionally accepted? These are the questions asked by Machlup (2000) in defence of social science when accused of inferiority. The fact that words are coined which have generally accepted meaning ascribed to them attests to some social science endeavours attempts at explanation. Nomothetic and idiographic explanations of phenomena evidence the degree to which explanations are generalisable or individualising; the former being aligned to the natural sciences and the latter to the social sciences. Both spheres of science employ nomothetic and idiographic explanatory powers. Surely biology is clearly far more idiographic in explanation than physics for instance and social science explanations often extrapolate across times, for instance in terms of the predictive formations of human cultures across the globe (Coertze, 1977; Haviland, 1991; Howard, 1989) and so are more nomothetic than previously thought. Due to the larger number of intervening variables within the social sciences then, there is thus less chance for recurrence of precise sequences. This is an inherent and intrinsic part of the social system and something that cannot be done away with.

Objectivity of observations and explanations

Considering value-neutrality and the stance taken by the social sciences, Machlup (2000) delineates four different meanings of the word “value” and how they impinge on the objectivist/subjectivist debate within the social sciences, namely;

- proper scientific detachment is interfered with due to possible pecuniary interests that the researcher might have or his view on the “good” that his research may do
- without biasing the results certain issues can cloud the investigative process, such as ethical considerations with which the researcher has to contend (an issue not problematic for the geologist or molecular chemist for instance)²⁹
- the social scientist often confronts research because of the belief in the possibility of having problems solved and the value that these results might have for human beings in some or other way
- often results are only explicable with reference to subjects’ values and thus the entire enterprise is value-ridden

All this leads to the inevitable “subjective difficulties of maintaining scientific detachment in the study of human affairs” (M. Cohen, 1931, p.348 in Machlup, 2000). Although naïve in its sentiment, is it not possible to objectify our subjectivities? Pie in the sky perhaps but Machlup (2000) maintains that “scientific objectivity” is not affected by subjectivities. In agreement with Maree (2003), it is maintained that objectivity is still possible even within a post-modernist stance on what it means to do science. Here one does not necessarily absolutise objectivity; and phenomena can be viewed from multiple angles in an objective manner. Radnitzky (1970) admits as much when he cautions against neglecting the objective and objectifiable processes inherent in natural sciences which has its place surely but simultaneously states that implying that all such processes are equally objectifiable is unwise.

Verifiability of hypotheses

High-level hypotheses reign in the natural sciences in comparison to low-level generalisations which are more common in the social sciences resulting in much testing and appraising of theories within the social science domain where it is often possible to delimit an area of investigation and study it. The nature of some natural science investigations may prohibit any attempt at confirming or disconfirming various postulates due to many unverified and possibly unverifiable hypotheses.³⁰ The issue of verifiability within the social sciences in terms of controlled experiments for instance, is not reflective of the methodology itself but more so a reflection of the subject matter of the discipline, which brings one back to the original issue of the need for another methodology for the social sciences or a resultant subsuming of disciplines into other disciplines. In sum though, the social sciences and specifically hypotheses resulting from research questions are indeed less verifiable than the natural sciences and their hypotheses.

Exactness of findings

Exactness can mean many things; accuracy of prediction, measurability or reduction to mathematical language (Machlup, 2000). Is it precisely these characteristics which currently define assessment within intelligence which conveys its attempts at a nomological-deductive framework within which to work, a striving towards a natural science model of investigation; yet this is at various junctures within psychology, condemned. Dynamic assessment in its most clinical of forms is precisely opposed or

²⁹ At least not in the sense generally referred to in the social sciences.

³⁰ Of course current unverifiable hypotheses may become, over time, verifiable or not as our understanding and technology allow us the luxury of pursuing investigations hitherto unreachable.



counterintuitive to this type of model and so attempts are made to objectify and make more robust its techniques in the hope that it will be more acceptable within mainstream psychological assessment. Once this is achieved, critics decry its statistical and natural-science biased approach. If interpretative models are utilised instead to gainsay this type of model, accusations of subjectivity, non-normativity and lack of predictability are hurled at the approach. Society demands objectivity, norm-based measures and predictive success in many areas of social science research, none more so than in assessment and evaluation. Researchers and practitioners are pulled and yanked from one side to the other. Is there a solution?

Measurability of phenomena

Espousing the quantification of data is indeed laudable if indeed it works, which of course it does in many instances, but in areas such as societal “measures” one is dealing with data of another type. Machlup (2000) states that natural science quantification of data is conducted in two steps; obtaining the data and numericising the data concluding with a result. Achievements in science are most often evidenced (and hence restricted) to areas where quantification is paramount (Osborne, 2000). However social science cannot avail itself of such a luxury as it necessitates at least two steps (with intervening variables); observation, data, numericisation, interpretation and result. Economics is a science which has as its raw data experience a numerical form as this field concerns itself primarily with numerical data.³¹ Physicists are able to measure various phenomena via equipment in which there really is a one-to-one correspondence with real-world data. Psychometrics too utilises various instruments which purport to measure various psychological phenomena; but unlike physics there is no one-to-one correspondence between reality and the measure as there is a process through which a theory yields concepts which are operationalised (via intervening variables) for the purposes of the specific theory (note that these purposes will change from situation to situation and thus are not uniform between studies³²). This is precisely the route followed by representational measurement theory concerning mental test scores (Borsboom, 2005). Measurement and quantification within psychology is discussed at length in chapter 4.

Constancy of numerical relationships

As alluded to above, even the constants of nature are subject to change over space and time, but for the present purposes, it is assumed in general that constants in nature remain unchanging. Many such constants exist such as the velocity of light, Planck’s constant of angular momentum, the gravitation constant and the fine-structure constant among many others, yet no single constant is known to the social sciences even though economics regularly computes ratios (Case & Fair, 1996a; 1996b) and sociology can account for certain ratios such as accident rates and epidemiological forecasts which simulate spread of diseases (Cockerham, 1998) and so on. These do, nevertheless, vary across time and place as well. There are no constants in the social sciences. Moreover, numerical assignments to constructs within psychology is a contentious area (see chapter 4).

Predictability of future events

Machlup (2000) states that it is unfair and unreasonable to even compare predictive abilities between the natural and social sciences. The main reason is that controlled laboratory experiments differ in quite radical ways in both sciences. Unorganised and unmanipulated reality yields to natural science predictive models quite a fair number of misses in comparison to the number of times predictions are accurate. As tools and instruments become more finely attuned towards those aspects that are to be measured so too do predictions become more accurate. Many of natural science’s efforts do not work very well such as weather forecasts, stock market fluctuations, lunar exploration and the history of science which is indeed filled with blunders and futile predictions (Youngson, 1998) yet “social scientists, for some strange reason, are expected to foretell the future and they feel badly if they fail” (Machlup, 2000, p.271).

Distance from everyday experience

Although not entirely in agreement with what Machlup (2000) has to say about natural and social science jargon the author does point out the nature of jargon and its distancing effect. Talk of electromagnetic fields, quarks and genotype-phenotype interaction is bound to impress on the layperson a view of science as evidentially correct. Since the social sciences deal almost exclusively with humankind as its data, many conclusions may seem false when compared to the layperson’s folk psychology knowledge of these same accounts resulting in discordant accounts and hence leading to a rejection of unscientific claims; a non-science posturing as science.

³¹ This is actually an ironic if not somewhat amusing statement, seeing as the one and only driving force behind stock exchanges is greed; a wholly and most human trait.

³² Of course it has recently come to light that even the so-called constants of nature are not constant or uniform across time and space (Barrow & Webb, 2005); just another progressive step towards greater enlightenment within science in general.

Standards of admission and requirements

Machlup (2000) is blunt and to the point when he states that regarding average IQ scores, natural science students on average outperform their social science counterparts, although generalisations such as these have to be interpreted with caution. Depending on college and university graduation requirements various disciplines expect varying standards. Prestige is linked to those disciplines which attract the higher performing IQ individuals and it is a common phenomenon that natural science students on average perform better than social science students; “many sociology departments have been notorious for their role as refuge for mentally underprivileged undergraduates” (Machlup, 2000, p.272). This is a social phenomenon dictated and mediated by the society in which students function. A counterargument might well be that high school and college examinations are attuned to a natural science curricula and that skills within these disciplines are what are needed in society in general. One may choose to look at this situation from a relativist stance and comment on the relative notion of what is considered required skills, but the truth of the matter is that fundamental skills within technological and natural science areas of study are most often the most demanded by society. One often hears about the lack of skills in natural science areas in South Africa for instance, and hardly much mention is made of the lack of skills within sociology or psychology. Could it be that more people take to social sciences because they are simply easier to study? It would seem that the answer to this is affirmative and Machlup (2000) certainly makes a compelling case for this conclusion.

It would seem from the above that the social sciences are considered inferior to the natural sciences but is this truly the case? Machlup (2000) maintains that it all depends on what questions are asked and the kind of answers one seeks. In stating that “sandpaper is inferior to velvet in smoothness” or that “biology is inferior to physics in internal consistency” one is not really stating that any one field of inquiry is less robust than the other, after all, it is desired that sandpaper be inferior to velvet, seeing as that is what is needed to sand wood for instance. One could quite easily turn the situation around and state that physics is inferior to psychology in determining the likelihood of success of cognitive behavioural therapy on certain patients. To state this is absurd. Yet it is often the case that two methodologies are compared in this manner; perhaps the comparison is unfair. Faust and Meehl (1992) offer six general working assumptions, the use of which seeks to enable the resolution of questions in science (and are here applied to the social sciences as well):

- science is the best available method for acquiring knowledge³³
- best available and optimal should not be conflated
- the scientific game is one of probability
- the human brain has a limited capacity in terms of integrating data
- description can help inform prescription
- small increments can have large pay-offs
- meta-theory may be difficult but is not impossible

Naïve or folk psychology, however palatable and accessible it is to the lay person does not have a proven track record of veracity and sound provable claims. However, in contrast to this supposition, Jordaan (1989) maintains that formal psychological theory originally emanates from common sense and so distinguishes between formal and informal psychology. This somewhat flies in the face of all that has preceded this discussion in terms of objectivity of concepts and terminology. Relativist understandings of concepts, it seems, will never be entirely eliminated, one can however control the degree to which they impinge on understandings within theories. Considering all that has gone before, the scientific method as such is the most consistently applied tool in the knowledge-acquisition game to date. However apt the scientific method is in ordering and logically deducing theories from hypotheses, it cannot be maintained that it can test the veracity of scientific theories itself. Although meta-analytic studies seek to cumulate findings across independent studies it is not necessarily a tool for theory verification. Faust and Meehl (1992) state that although the scientific method often utilises logical and verifiable steps in its procedures, science is itself often a game of chance in which the best strategies are employed to minimise such chance effects. Succeeding only some of the time is perhaps the best that can be achieved in a stochastic system such as the real world. Limited human capacity precludes exact detailing of synthesised undertakings, the more information to be cumulated the less accurate are the final renditions of what is supposedly reality. Science after all is a very human undertaking, but there are systems and limits in place which seek to control those aspects. Yet the psychology of science cannot turn away from the fact that this system too is faulty by mere fact that it is a human system; there is “mind in society” (Meehl, 1992, p.340). Just because the past has yielded instances of descriptions which can be utilised as predictive scenarios for the future does not necessarily mean that “is” now becomes “ought”. Science has not always proceeded along a course of optimal functioning and surely never will. In some areas of science, progress is slow, but what might be considered a small step forward could in turn be described as fairly large gain in total. Psychology might be considered just such a discipline in which small progressive steps are construed as very unlikely and the probability of success is pinned at a very low base. Even small steps are thus large in an area with very

³³ This is contentious; however if an argument is to conceivably get anywhere, a starting point is necessitated.



small probabilities of success. In determining the validity of scientific progress, a database of successful and unsuccessful studies and theories can be drawn up in which ratings of these theories can be conducted. Whether psychology has available such a large database is debatable, seeing that as a formal discipline it has not been in existence for all that long in comparison to some disciplines within the natural sciences. Physics had developed to a point necessitating controlled experiments in order to answer increasingly more empirical questions during the seventeenth century; in biology this occurred around 1770 but psychology is of course only 126 years old (Toulmin & Leary, 1985), so perhaps it is unfair to always compare disciplines.

3.3.1 Explanatory mechanisms in social science

If the social sciences continue to tie in with natural science methodology problems are bound to occur as there are numerous inherent differences between the two areas of study. Rosenberg (2001) cites seven main dimensions according to which an analysis can be made of the chances of explanatory successes within the social science if indeed they continue to follow the natural science agenda on methodology. These dimensions will be briefly discussed.

- i. Causation, law and intentionality
- ii. Regularities vs. rules
- iii. Ontology in social science and biology
- iv. Teleology and function
- v. Reflexive knowledge in social science
- vi. From interpretation to historicism
- vii. Dangerous knowledge, ideology and value freedom

The main issue dominant in the first of these dimensions is one of intentionality as deductive models of explanation simply do not hold if premises are intentionally bounded (Nagel, 1979, refers to intentionality as humans' volition). In other words cause and effect are difficult to ascertain in a system which precludes viewing it as cause leading to effect. The single-most cited problem evidenced in the literature regarding intentionality is how initial physical states can originate if they are intentionally based (Doyal & Harris, 1986; Rosenberg, 2001). As this issue is paramount to the study and to the understanding of consciousness (Plotkin, 2003) much philosophical work in this area is carried out by the phenomenologists (Maree, 1995; Mautner, 2000; Rosenberg, 1985). Intentionality "ruins" linear cause and effect experiments and is almost impossible to eliminate altogether. With such a predicament even the basics of nomothetic-deductive explanations runs into problems - the perceived cornerstone of scientific methodology as understood by the author.

Secondly, the seeking of lawful relationships within natural science investigations led social scientists to attempt similar endeavours in establishing law-like statements about occurrences on the behavioural front. Such naturalist attempts fostered an outlook which was based on natural science models of science. However no such law-like postulates could be found even though behaviour was often rule-bound and fixed in certain contexts, the boundaries of what constituted rules and regularities were often blurred and adapted to suit the ever-changing human science landscape. No predictions could be based on a system of rules or regularities unlike those predictions within the natural sciences which regularly yielded laws with a proven track-record of predictability. Anti-naturalists maintain that the aim of the social sciences is not to predict behaviour but to understand it and because of its inherent lack of fit with traditional nomothetic-deductive ideals and unlaw-like status it is maintained that it will never be a predictive science as understood in the natural sciences (Rosenberg, 2001). Eliminativism goes one step further and seeks to do away with intentional states undergirding behaviour and aims to study behaviour from neuroscientific non-intentional perspectives.

Thirdly, in keeping with McErlean's (2000) illustration of the hierarchical arrangement of disciplines, Rosenberg (2001) compares biology and psychology in terms of how they manage to successively explain away individuals (one cell or one human being); smaller functioning groups (groups of cells or groups of human beings); larger groups (molecules or tribes of human beings) right through to larger and more expansive groupings (biological species or the human race). Similar mechanisms are utilised to analyse various sizes of groups. Due to the similarities evidenced between the two disciplines an argument of further similarity ensues which states that if biology can explain away larger group functioning (evolutionary mechanisms) in terms of holistic smaller group functioning (family groupings, breeding populations) which are themselves independent of the smaller entities within them then so too can the social sciences explain in similar fashion.

Fourthly, causal explanations become confusing if teleological arguments are brought into the fray of debate. Rosenberg (2001) states that within biology where certain organs are explicable in terms of their functioning (kidney's exist in order to remove soluble wastes) so too can certain societal institutions be explained away by stating the functions they perform. Such purposive or teleological explanations are dubious at best according to the more empiricist philosophers primarily because it reverses the order of cause and effect, i.e. "something exists because it has a function to complete" is often the not so subtle error made when considering the workings of evolution; there simply is no teleological end-point in sight, it is a "blind mechanism" which does not favour one outcome over the other (Fearn, 2001) and is often referred to as the teleological fallacy (Feldman, 1997).

The white coat of an arctic rabbit is not white *because*³⁴ of the snowy background, it is white purely due to random genetic mutations which occurred and through which enabled the successful adaptation and breeding of these white-coated rabbits. It is an ability, for want of a better expression, that ensures its successful replication in the gene pool (Fearn, 2001). Likewise, human beings did not evolve a large brain in order to facilitate our endeavours into abstract mathematics (Gould, 1998). However, recent studies do suggest that evolution does tend to favour those with higher levels of intelligence over time within the larger population³⁵ (Van Schaik, 2006). As time progressed, fewer darker coloured rabbits managed to breed successfully due to predation and thus left a gene pool of active breeding white-coated rabbits.³⁶ Cause and effect can be muddled quite easily, although Nagel (1979) maintains that by giving teleological explanations one is not necessarily implying that the “future is an agent in its own realization” (p.25). At times the white rabbit is still eaten despite its adaptive advantage, but it did not sprout a white coat to avoid predation in the first place; that “event” was purely pot luck and as Rosenberg (2001) states citing an effect to explain its own cause is hardly much of an explanation. Social sciences tend to want to explain things in holistic fashion citing the tendency in biology to explain phenomena at group levels (eco-systems for instance) but often the case is such that the argument for the existence of phenomena and their properties is the explanatory role of these phenomena and their properties. There are views however which contest the role that science has to play in explaining away what is often perceived to be purposeful mechanisms at work in nature.

Fifthly, social science arguments, theories and predictions are often reflexive and by this is meant that by nature social science's subject matter can be influential to itself due to the ability to self-fulfil or disconfirm findings while in the process of being developed. Rosenberg (2001) uses the example of the broadcasting of voting results thus far attained during an election campaign, but in doing so potential voters can change their vote depending on the nature of the count to date. This is a reflexive process and a process not often seen in the natural sciences. There lies a special kind of responsibility within the social sciences when formulating and building theories and models; theories and models of a different kind are thus necessitated by the social sciences. Critical theorists for instance maintain that the aim of social science is not to predict nor explain but to emancipate the very people it seeks to study. Surely intelligence assessment and included in this dynamic assessment, is hardly emancipatory? It subscribes to rigorous empirical nomological-deductive explanations of functioning, claims predictive success and is a field rife with statistical manipulations and significant differences - so much for the reflexive emancipatory outlook envisaged by the above stance on the social sciences.

Sixthly, fully determined cause and effect reign supreme in the natural sciences in which laws are considered fundamental across time and space, antecedent causes result in future effects and any retrodictive causes can be found because any intervening variable is linked via a causal chain of cause and effect. Not so with the social sciences in which historicity often casts a different colouring of interpretation on events occurring during certain periods in history, the source of which is the above mentioned intentionality of human beings. Had humans not intended that feudal systems take root in civilisation then socialist political movements would not have resulted for instance. Historically, conditioned actions result because of intentional stances and social sciences do not work according to strict causal laws but according to historicised explanations; each historical period thus harbouring its own unique explanatory mechanism. Psychology, as with many other social science disciplines, infuses theory with observation language and is often unable to disentangle the two aspects due to the very historicised nature of the subject matter. In other words the language used to describe theory is itself a product of yet more theory, unlike physics for instance in which observational terms are clearly separated from theory (Sanders & Rappard, 1985).

Lastly, a most vexing issue is that of the value-ladenness of value-neutrality of the social sciences and the answers it seeks to certain questions. Certain subject areas, no matter how well conducted the studies are, are simply considered too inflammatory for results to be disseminated. What then is the role of the social scientist? Gatherer of information, objective scribe of societal functions and phenomena, emancipator and liberator or knowledge broker? Intelligence testing is one of a few issues discussed by Rosenberg (2001) when detailing the agenda of the social scientist. Despite years of research and literally thousands of research papers documenting and debating the heritability of IQ, there is as yet no conclusive answer or solution to this question, mainly because of its incendiary nature. Dynamic assessment could be considered an emancipator of sorts, seeking to cut through what is considered unfair and discriminatory assessment practices, yet also having to fulfil the need of rigorously

³⁴ This “because of” notion is often cited as clear evidence of a “weak relation of partial causal contribution” (Woodfield, 2001, p.493).

³⁵ Note that the dinosaurs reigned supreme for well over 160 million years without ever notably increasing in intelligence, at least not what can be inferred from the fossil record. Once mammals got a toe-hold into the environmental niche created by the mass extinction of the dinosaurs there seems to have taken place an ever-increasing propensity for mammalian brains to yield to environmental press. The fact that humans exist in their current form attests to this. However, had the dinosaurs continued unabated (however unlikely this scenario would have proven) mammals might never have taken hold. There is no teleological argument here: it just so happens that individuals with “higher intelligence” (as we perceive it to be) have, on average, tended to remain in the gene pool.

³⁶ Of course random drift as mechanism of selection can also be argued; such that despite any advantage conferred upon white rabbits, it may be that predators are in any event unable to tell the difference between white or brown rabbits and the choice of predation is not causally linked to the colour of their coats but by random causes (Millstein, 2002).



tested, veracious and empirically verifiable method of assessment; in other words emancipate in as scientific a manner as possible! A contradictory and ambiguous instruction at best.

3.4 Psychology

Psychometrics, in its attempts to measure certain aspects of behaviour, necessitated that these aspects be operationalised which some might view as the first step down a slippery slope for the very act of operationalising is subjectively and interpretatively approached (Govier, 2000). The terms aggression, intelligence, patience and so on can all be radically operationalised in any amount of ways depending on the criteria utilised and stipulated; unlike natural science counterpart terms such as pressure, velocity and wave length. Moreover, the price incurred by operationalising concepts such as intelligence within every test labelled as such, is that the concept or construct is defined differently for each test. The end result is a concept which is clearly defined but is defined in limitless ways which of course is not progressive (Borsboom, Van Heerden & Mellenbergh, 2003). One need only contemplate the chaos that would ensue if “the speed of light” was defined differently (yet accurately within each test built to measure it). Wundt’s disciplinary split between the natural and mental sciences resulted in psychology straddling a middle-way between the two branches of knowledge acquisition (Miller, 1985; Robinson, 1985; Tweney, 1989). This necessitated that psychology choose for itself, from natural science methodology, the requisite tools warranted by certain types of experimental programmes as well as tools chosen from the social sciences enabling the study of particular social phenomena dependent on mediatory experiences on the part of the researcher (co-creating the experience along with the “subject”). Wundt distinguished this difference in 1896 and the discipline still straddles the demarcation point 110 years later (Giorgi, 1985).

Defining terms and how they are operationally put to use contextualises the terms within any particular study and thus its explanatory powers are more obvious. Yet stability of terminology between studies is not an easy task to accomplish and hence the lack of predictive power within such a discipline as psychology, a discipline without appreciable movement and lack of cumulation in the last 400 years in comparison to that of physics (Faust & Meehl, 1992; Rosenberg, 1985; Rossi, 1997; Shadish, 1989). Govier (2000) maintains that adding statistical “talk” to vague operationalised terminology merely adds to studies’ pseudoscientific³⁷ status upon which much attention is lavished due to the figures employed in these studies’ findings. They are hence accorded more “scientific” credibility. All these concerns would undoubtedly lead to social scientists feeling inferior in terms of the veracity of their claims. An inferiority complex (Machlup, 2000) has often been associated with social science endeavours when held up alongside their natural science counterparts in terms of funding and prestige for instance. Why the need to write about the social sciences in this way then? If, as it has come to be expected, the social sciences do not resemble the natural sciences in a methodological manner (Cohen, 2000), what does it matter whether the social sciences do not proceed along a course equivalent to that of the natural sciences? The only answer to this question is another question: why the need to proceed along lines similar to that of the natural sciences within psychology? Why the need for statistical inferential studies? Why the need for predictive validity at all? Why the need for a call for figures in place of narrative? Why the preference for empirical research as opposed to theoretical research? It seems that the discipline is its own worst enemy and those practicing within it are predisposed to think of natural science models when conducting studies yet seek to criticise the models to which they adhere (Campbell, 1986; Eysenck, 1985; Secord, 1986).

Kukla (2001) envisages the roles of theory (ratiocination) and empiricism (observation) within psychology as paralleling the roles these endeavours play in physics. Theory bequeathing an infrastructure for empiricists to follow which in turn mediates theory so as to better reflect empirical reality and both strands are necessary for “the business of science” (Pascual-Leone & Sparkman, 1980, p.89). That the two methods have received unequal attention from psychological practitioners and researchers alike goes uncontested mainly due to the perception of theorising as minor ingredient in the larger recipe of a functioning psychological discipline. The perception of theorising within the area of physics is one of almost adulatory awe at the, at times, astonishing findings which result from “armchair theorising”. Theoretical physics encompasses a vast array of tantalising areas, among others, the role of mathematics within electrodynamics, the electromagnetic theory of light, electromagnetic waves, general relativity, wave mechanics and microphysics much of which has since been empirically validated and has often shed light on matters that neither observation nor empirical studies could. These areas have at times lead to unexpected discoveries and technological advancement (D’Abro, 1952; D’Agostino, 2000; Einstein & Infeld, 1961; Harré, 2002; Pavitt, 1997; Petersen, 1969; Slife & Williams, 1997). Perhaps it is just this: the very medium through which theoretical and applied physicists are able to communicate with each other is clearly understood by all parties concerned, that of mathematics for instance; yet psychology is

³⁷ Akin to Popper’s demarcation dispute as to what constitutes science and non-science (Chalmers, 1999; Eysenck, 1985, 1987; Faure & Venter, 1993; Radnitzky, 1970). It is more often the case that treatises on pseudoscience are aimed at the natural sciences in which false accounts of theories are dispelled (for instance see Dawkins, 1999a; Gardner, 2003; Sagan, 1997; Shermer, 2001 and although not levelled as an affront on pseudoscience, Dawkins, 2004 goes to considerable lengths to rectify certain pseudoscience notions). There are fewer such works dealing with pseudo-social science an area that would prove ripe for scepticism in its most productive form. Gould’s (1997a) now notorious attack launched at intelligence measurement is one such book. Meehl (2002) refers to incomplete theories as “apseudo” if they do not evidence false postulates.

lacking such a common mode of expression resulting in miscommunication and talk at cross purposes (Weems, 1998). The form that such a consensus mode of communication would take is another issue.

The ever-swinging pendulum³⁸ still sways the way of rationalism only to reverse full thrust back to empiricism and is currently mid-swing between the two poles. The consensus view today reflects a compromise between objectifying subjectivities and allowing due credit to the role of theory as category within the greater method of science; empiricism being yet another such category (Kukla, 1989, 2001). To ignore or underestimate the theoretical contributions that have been and are being made within the social sciences, with emphasis on intelligence research would be tantamount to ignoring the mammoth contributions made by theoretical physics to the study of physics in general. In receiving recognition for the scientific validity of theoretical physics, a fundamental change occurred in the latter half of nineteenth century physics (D'Agostino, 2000).³⁹ Gustav Fechner's experimental psychology (considered the founder of modern quantitative psychology; Michell, 2003) and Wilhelm Wundt's laboratory on focused psychologising paved the way for a more scientific study of the mind (David, Miclea & Opre, 2004; Sahakian, 1981). This in turn allowed for the study of cognition within controlled environs in the late nineteenth century making, along the way, many presumptions about measurement not yet verified or debated (Michell, 1999). Having thus burgeoned from a scientific platform, it is hardly surprising that investigations into human behaviour have since relied on methodology borrowed from the natural sciences (Michell, 1997). Cognition and the field of cognitive science has been lauded throughout the expansive literature on cognition as *the* field with which much of the social sciences should align (if many sub-disciplines are indeed to survive) (Turner, 2001). Theoretical psychology as a discipline within psychology is not the thriving sub-discipline (Looren De Jong, Bem & Schouten, 2004) that theoretical physics is to physics and the most cited reason for this is the immaturity of psychology as a science (Dawson, 1995).

What is the model of good psychology espoused by social science? Does it differ and if so how and what are the implications for theories and models within the social sciences? Robinson (2000) succinctly sketches the development of psychology at the advent of the twentieth century and captures the essence of this new psychology by describing it as an amalgam of experimental science, Darwinian theory and positivism. At first, evolutionary theory played a predominant role in securing standard notions and criteria for theory building and served to tie together the underlying rationale for both natural science as well as the newly forming "moral science" (psychology) whilst simultaneously offering credibility to this new endeavour.⁴⁰ Proponents of a human psychology cautioned against the insufficient framework offered by evolutionary theory stating that it was both incomplete and misleading.⁴¹ The fact that behaviourism was to become a dominant force in psychology having been informed by functionalist biology attests to the strength of the scientific enterprise and its all-pervasive influence on the direction that psychology was to take (Robinson, 2000) even though behaviourism's "downfall" in the 1960's and 1970's (mainly due to the rise of a competing information theory paradigm) lead to a decreased emphasis on many of its tenets (Green, 2001; Harnish, 2002). Regarding learning disability research, Swanson (1988) finds fault with the basic framework upon which this science rests, relegating it to the status of normal science as opposed to mature science status. A mature science, he states, is one in which there is an agreed-upon framework, in which re-thinking of assumptions takes place as well as a recognition of the change in reality which it has undergone since its inception. Dynamic assessment could perhaps be classed as applied research as it seeks to solve immediate real-world problems and attempts to answer the "how" questions within the field, i.e. how testees respond to assessment, whereas basic research seeks to contribute knowledge to a field without necessarily having any immediate relevance, i.e. asking "why" questions such as why respondents answer the way they do (Swanson, 1988). However, both inform the other at some time. Basic research may have applied value later on, as has been the case with the development and later deployment of item response theory, a model whose time has now come regarding the available computational power (Embretson, 1997b) (see chapter 4). Applied research can inform the knowledge base in a more general manner. Dynamic assessment's status as model and/or theory attests to its undefined status as basic or applied science as the two have seemingly divergent agendas.

³⁸ This over-used analogy does seem nevertheless to capture the essence of fluctuations between rival views on many a matter within knowledge gathering.

³⁹ It has been argued that the mathematical sciences are perhaps the most detached and value-free of all the scientific enterprises (Livio, 2003) although the discipline of ethnomathematics will most likely advocate an alternative point of view. Mathematics is a field as diverse as any other (Fuchs, 1967) but invariably mathematical concepts have predicted manifest reality such as Maxwell's "purely mathematical description of electromagnetic fields" (Hey and Walters, 1997; Millar, Millar, Millar and Millar, 1996, p.221) with scientists happily disregarding currently available empirical evidence to the contrary but ultimately being vindicated in their original mathematical stances. Quantum research is just such a field of enquiry (Al-Khalili, 2003). Radical as such an approach might be, it "can be deduced from Maxwell's, Hertz's and Boltzmann's convictions that *experiment was not a crucial test for theory's validation*" (own emphasis) (D'Agostino, 2000, p.xiii), as was the case with quantum electrodynamics which was rendered theoretically inconsistent prior to Feynman's contribution (Kukla, 1995a).

⁴⁰ "New endeavour" does not necessarily indicate that psychology as a field of interest suddenly became apparent. Up to the time of Darwin, philosophers had long debated the need to study human beings, but it was only during the latter half of the nineteenth century that psychology established for itself a separate recognised domain of enquiry. Helmholtz had as far back as 1862 argued that science and philosophy had already become divorced (Robinson, 2000).

⁴¹ In this vein it is ironic then to think that evolutionary psychology has grown to such proportions.

3.4.1 Explanatory mechanisms in psychology

The traditional formal presentation of how a scientific investigation proceeds is illustrated in figure 30 below. Following this figure is an overlay of central tenets pertinent to the study which can be viewed in figure 52. From this can be seen how closely the enterprise of psychological research mimics the methodology followed in the natural sciences, an area upon which initial psychological studies were largely predicated and still are today. It comes as no surprise then that early psychophysics in the tradition of Fechner (who was a retired physicist)⁴²would be so natural-science orientated (Leahey, 2000). It is interesting that notable committees instituted throughout psychology's early formal history were partly chaired by natural scientists. The 1940 Ferguson committee instituted to question measurement within psychology was headed by A. Ferguson who was a physicist (Michell, 1999).⁴³ In fact the history of psychological measurement is peppered with individuals, now chiefly remembered for their psychological contributions, but who were originally natural scientists. The origins of psychological methodology are almost entirely presupposed by natural science technique in order for it to be accepted and be proven more relevant (Meier, 1994). The reigning methodology has perhaps had its greatest lasting influence within the area of individual differences, notably intelligence assessment. Physiological psychology and behavioural psychology have what is seen to be mixed agendas with cross-pollination occurring throughout the varied discipline's history. Pertinent to this study is the influence from natural science methodology regarding measurement. Measurement is a manner of numericising constructs which are supposedly quantifiable (Wille, 1994) and it is quantification which is considered the hall-mark of scientific progress. Few areas within psychology are so riddled with a strict positivist outlook as testing under which is housed dynamic assessment. In essence the main aims of a psychological science in its nascent days was to infer, predict, decide, understand probability through the utilisation of precise statistical techniques as well as to rely on models and connect such models to the requisite techniques (De Leeuw, 1994) so as to ensure a robust scientific discipline (Schönemann, 1994). The entire research process is paralleled to much of natural science methodology beginning with hypotheses which are formally set out, followed by experimentation in one form or another resulting in theory-consistency or divergence depending on the statistical or measurement techniques deployed (Mouton & Marais, 1988). Aspects and notions such as rationality, pragmatism, law-like concerns for an apt description of behaviour, early mechanistic interpretations, nomological approaches, empiricism, verifiability, reliance and valid accounts of measurable phenomena, reduction, realism and universal ascent of such findings make for a science-oriented psychology (Leahey, 2000).

Is psychology's explanatory concern more in keeping with theories, models or paradigms? Looking more specifically at cognitive psychology, David, Miclea, and Opre (2004) state that the terms paradigm, framework, theory and model are often used interchangeably by cognitive psychologists which does not make for an easy time when attempting to allot, categorise, label and effectively 'place' theories and models. Theory and model as concepts too are utilised interchangeably within educational research (Haertel, Walberg & Weinstein, 1983). This statement is to be considered over-and-above the fact that the construct "model" is itself problematic and open to various interpretations (Craver, 2002). Within the realm of cognitive psychology then, paradigms may be construed as overarching collections of theories themselves consisting of explanatory models. A paradigm is judged according to the number of successful theories it generates and cannot itself be assessed in the same manner that its theories are assessed. Several models may emanate from a theory, a single theory forming part of many theories generated from within a paradigm. Theoretical stances may be confused with paradigmatic approaches towards phenomena thus furthering confusion as to a framework's approach towards its area of interest. A theory can be defined as a set of formal expressions which enables the development of axioms which dictate allowable theories. Truth and validity are ensconced within these formal systems which allow for a certain degree of prediction and control over phenomena they purport to study (Mautner, 2000). However, the social sciences and indeed psychology cannot with surety state that this is indeed the path it traverses when forming its theories. These theories are of a more generalised type of guiding principle.

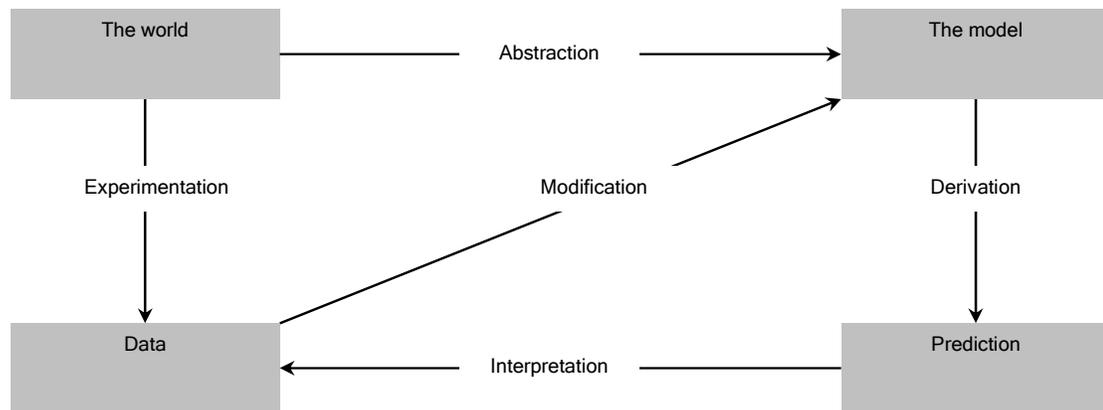
Psychology and indeed the purview of intelligence research including dynamic assessment research must be mindful of theory-begging, which is inconsistent with good science practice of first establishing the fact and then the theory as opposed to conferring fact upon a theory (Mautner, 2000). Is psychology largely a collection of naïve theories derived from hypotheses which are themselves based on propositions which consist of various constructs which are in turn visualised in the form of schemata and models? Or are models, schemata and constructs derived from theory? Observation leads to formulation of ideas which in turn drive tentative constructs and hence hypotheses and theory. But observation is itself theory-laden which results in a theory-laden view. This theory-laden observation colours the theory, hypotheses, propositions, constructs, schemata and models. A recursive feedback loop ensues in which designation of a beginning and an end is blurred and vague. The contentious debate between theory-laden observation and observation neutrality has passed down through the ages without being adequately resolved. On the one hand any and all observations are tinged with subjective accounts of how any

⁴² The first thing that comes to mind when considering direct importation from physics into psychology of terminology is the term "vector" utilised specifically in intelligence assessment. Such vectors provide information on magnitude and direction, as is commonly understood in fundamental physics (velocity is a vector but speed is not). Response vectors within IRT models (see chapter 4) include magnitude as well as direction of responses.

⁴³ The famous Louis Thurstone of primary mental abilities fame was an electrical engineer as well as a psychologist. Is it any wonder that early quantification in psychology traversed the path it did? (Wright, 1997b).

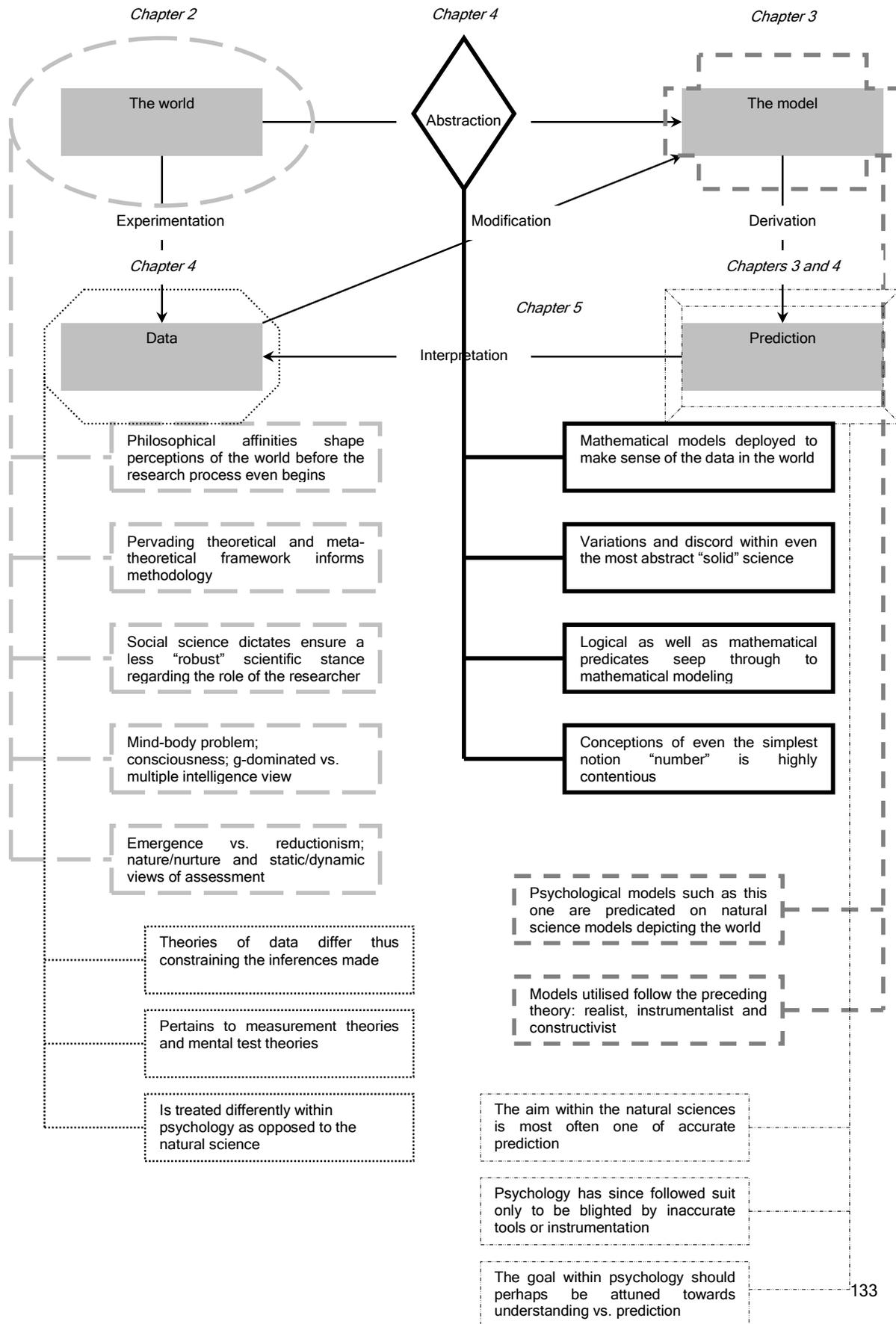
information is transmitted and interpreted. It is nigh impossible to discard innate prejudices as these are bound up with our perceptions. Opposing this view is one of being able to muster objectivity even though background interpretations and “influences from the perceptual system” may play a role (Estany, 2001)⁴⁴ (see chapter 2). It is upon such a foundation that theories are generated within the social sciences and permeate the field of intelligence research and thus in its turn subsumes dynamic assessment research. Are most explanatory mechanisms within dynamic assessment theoretical in nature or do they reflect models, as there is a difference between the two? Theories offer depth explanations, heuristic value and classificatory functions whereas models provide these aspects bar explanatory functions. Are “laws” and “rules” akin to theories? Is dynamic assessment a theory at all? Theories and models are not necessarily synonymous terms as they may engage different levels of conceptualisation and ultimate manifestations of their ontology and epistemology (the nature of what is and what can be known). Firstly, the philosophy of science has more often than not engaged the subject matter of the natural sciences in determining the role played by theories and models. As such, it is not always possible to draw direct inferences from natural science philosophy and apply analogies in an appropriate and workable fashion to the social sciences (Outhwaite, 1987).

Figure 30 The nature of scientific investigation (Coombs, Dawes & Tversky, 1970, p.3)



⁴⁴ In this long-ranging debate philosophers of mind and science have been at odds concerning how perception is influenced by theory commitment and how the two are inseparable (Churchland, Hanson, Kuhn and Feyerband for instance) or how they are in fact two systems which co-exist but do not necessarily influence the other (Hempel and Fodor’s modularity of mind for instance and Gardner’s vertical representation of mind; Anderson, 2005; Raab & Gigerenzer, 2005).

Figure 31 Utilising Coombs, Dawes and Tversky (1970, p.3) and overlaying central tenets pertinent to the study

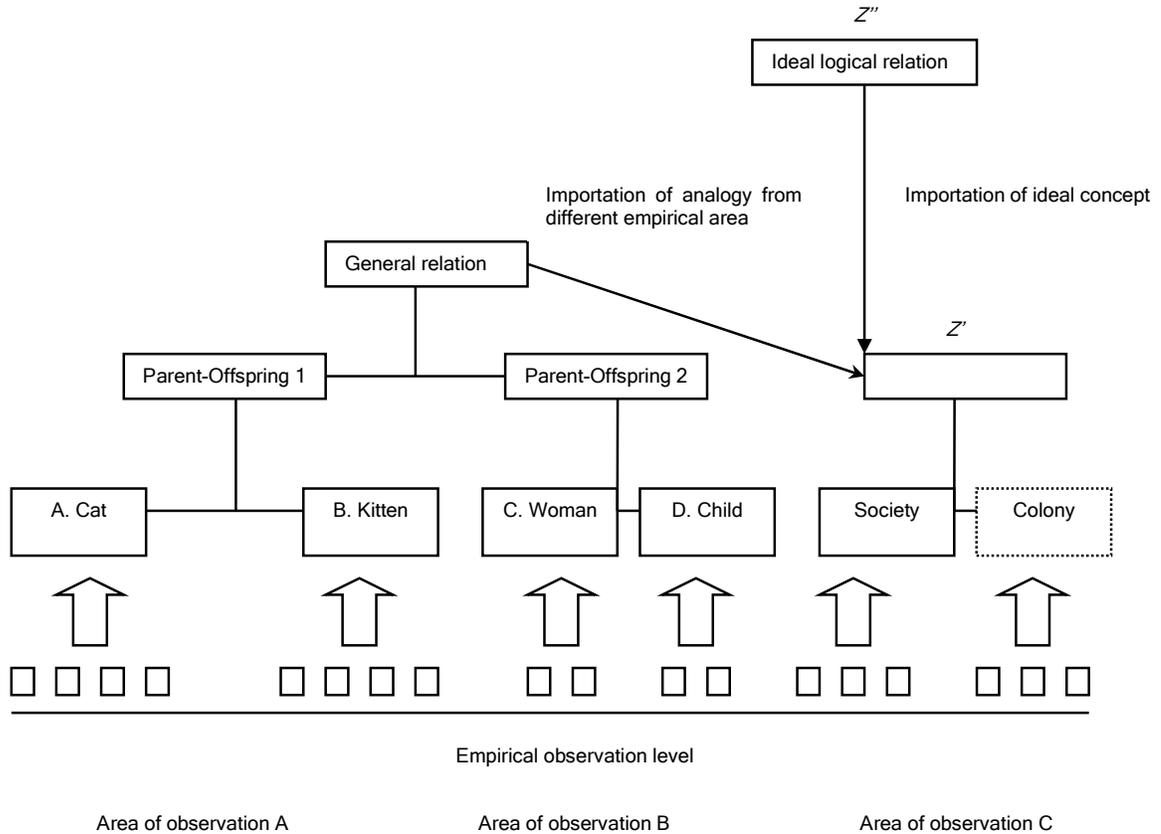


Notwithstanding, there are key differences between the roles played by theory and models. For instance, Atkinson & Shiffrin (1968) emphasise the need to distinguish their framework and model from theory. Among the myriad information processing models, an example from stage-theory models is given. Exponents of stage-theory type of models include Atkinson and Shiffrin (1968) as well as Craik and Lockhart (1972) who have built upon the theories and frameworks of others (Broadbent, 1987; Miller, 1956) and concern themselves with the workings of memory, as memory is thought to be “tied to levels of perceptual processing” (Craik & Lockhart, 1972, p.676). Of note is that both sets of authors make explicit their stances towards a deployment of frameworks and models as opposed to prescribing theory in their papers. The “stage” referred to in stage models of information processing assumes that information is initially processed at a level discriminating between lines, angles or pitch for instance and later processed at a level more discriminating in terms of depth of processing i.e. matching input against information already stored, hence information processing occurs at various levels (Craik & Lockhart, 1972). Eventually information is compared with prior experiences and may thus be ‘coloured’ to a degree through interpretation. In their oft-quoted paper, Atkinson and Shiffrin (1968) set forth a general theoretical framework for understanding memory and proceed to garner support for their various models within this framework by describing a number of experimental results. Of note is their categorisation of the memory system into two main divisions, namely; permanent structural built-in physical features in which impermanent control processes play an important role in allowing for flexibility within any given situation and secondly, memory which is divided into three structural components including the sensory register, short-term store and long-term store. The authors typically explain their model of memory by referring to the at-the-time dominating computer analogy.⁴⁵ Working within the confines of any operating system, the programmer writes code instructing the hardware as to its functioning, which is reflected back in their model as the ever-changing control processes which adapt to differing situations and in so doing instructs the permanent features to carry out tasks. In essence, their proposed model illustrates the route of information processing, initially traversing the sensory register following the path to short-term storage and eventually long-term storage. Decay and loss of information occurs on route but the process is largely controlled by the individual’s control processes. One well known form of short-term memory control process includes the chunking or grouping of information which facilitates recall of information that is grouped together. Atkinson and Shiffrin (1968) emphasise the fact that their framework encompasses models which are derived from a general framework and is thus not a theoretical model of memory functioning.

Theoretical models can be construed as models of the presupposing theory which serve in the capacity of demonstrable prototypes of theories. In other words models are workable prototypes of theories which may or may not always concede to reality. In attempting to construct a model which will be based on theory (a natural out-flow of theory construction), the model can later be tested, enhanced, changed and adapted based on the empirical findings of reality. Theories are based on loose model conceptions or observations (which themselves are “permeated by theory”; Chalmers, 1976, p.33). ‘Theory’ and ‘model’ are intertwined as models are changed in order to better reflect reality and hence alter the underlying theory. Thus a cycle ensues: ensuring ultimate fit between theory - model - reality. Models classify and suggest new relationships in an heuristic manner whereas theories classify, define relationships as well as explain and interpret but the defining borders between models and theories are often vague (Mouton & Marais, 1988). Models serve to reify theoretical concepts (Wimsatt in Fiske & Shweder, 1986) and facilitate the understanding of theory (Hempel, 1970) which is in turn informed by models’ engagement with reality and hopefully elaborates on the real construct. This crucial difference between empirical construct and theoretical concept is paramount to the discussion on test theory and measurement (as well as its mathematical underpinnings which will be discussed in chapter 4). To prime for this discussion, Cattell’s (1988b) illustration is itself a good analogy of what is referred to here (an analogy depicted within an analogy!). See figure 32 below where the ideal logical relation would retain theory-reality isomorphism but due to reality as is does not. The use of analogy importation from across disciplines also serves as psychology’s foundation for much of its ideas and methodologies. There are however two problems with just such a scheme; firstly the ideal relation is not manifest as such and secondly the importation of “other” methodology is a luxury not necessarily appropriate. Nevertheless this is the situation with which psychologists are currently saddled. These ideas are commonly interrelated and utilised along with “maps”, “analogies” and “metaphors”. Analogies perhaps being the more oft-repeated term when comparing the theory/model to something else. Ryan (1970) equates theories with attempts at making claims, whereas models make no such attempt. In other words, models highlight the way something may/may not work whereas theories purport to explain why the mechanism works and how in fact it does work.

⁴⁵ Of note is their use of the term “buffer” (although not coined by them) which refers to the amount of information that can be stored at any particular time in short-term memory, although not indefinite, maintenance of a set amount of information is stored, which allows for the loss of items only to be replaced by new incoming information. This is similar to what occurs in printers for instance, where printing jobs are spooled to the printer buffer and stored in memory.

Figure 32 Cattell's (1988b) depiction of the basis of distinction between theoretical concepts and empirical constructs



Within and between disciplines and how the notion in the above figure permeates this study in a number of areas

The bridging within and between disciplines is problematic but not insoluble. Isomorphic renderings within psychology as when neural speed (empirical construct) is related to intelligence (theoretical concept) poses just as much a problem as when the bridging principles involved relate natural science methodology to social science methodology. The isomorphic bridge in this instance happens to be constructed on an analogy or on the importation of such analogies. One could in all reasonableness refer to the latter as a meta-bridge in keeping with the theme of this treatise. See figures 2 section 2.4.5.3.2; 24 and 25 section 2.8.6; 28 section 3.3; 33 and 34 section 3.4.2.1 for more on this and how it relates to other areas within this study. There is a particular need in this study to be keenly aware of the arguments, figures and issues discussed in each chapter as they play forth on the larger issue at hand

3.4.2 Psychological theories

Empiricism and rationalism as basic philosophies within science always co-occur and attempts to discredit either will forever result in a stale-mate of sorts. The crux of the matter is that both are necessary contributors to the development and eventual progress of formal science as understood to be the case in mainstream Western science.⁴⁶ To imply that observation be the only means towards understanding would be to contradict the history of science, especially the history of physics from the late nineteenth century. Without guidance from formal theory many “discoveries” would either not have been made or would have lingered in their eventual manifestation. Theory of course without accurate observation and inferential observation⁴⁷ as the case most often presents stands on its own neither to be proved nor disproved, verified nor falsified without the requisite tools to test its credibility. Science 101: theory informs observation which informs theory. However a less futile discussion focuses on the manner in which theory and observation can work to the advantage of scientists and social scientists (specifically psychologists). Cattell (1988a) delineates a division of psychological progress with the Wundt-Pavlov tradition advocating laboratory-like progress on the one hand and the Galton-Spearman tradition on the other hand traversing the terrain in more exploratory fashion. The former heeded natural science methodology and so imported its scientific tenets whereas the latter adapted to the needs of the discipline without fear of recrimination (Cattell, 1988a). It is from the Galton-Spearman tradition that psychometric measurement proceeded and which rendered itself amenable to mathematico-statistical treatments of analyses (Cattell, 1988b).

Theories and models are manifest constructions of happenings which are difficult to observe and ascertain in terms of the mechanics of how things function. Their heuristic value is most evident as simplified versions of reality, versions which are more amenable to understanding and by reduction (their reason for existence) are able to facilitate the transition between humans’ ideas of reality and reality as is.⁴⁸ Realist and anti-realist accounts of models yield varying explanatory power depending on how they are deployed (Green, 2001). Mentalistic models of human cognition and human behaviour could not truly subscribe to realist accounts of explanations as there is no purported one-to-one relationship between what is being explained and what in fact is trying to be explained. If a theory designates the existence of electrons then such entities exist and are not inferred models of what might constitute an electron per se. However its existence is inferred from more indirect means than pure observation and as Carnap (1993) states, there is no line or demarcation as to what constitutes a direct observation and an indirect observation, for it often happens that existence of entities is so evidently obvious but only so through indirect means that it becomes increasingly difficult to state its existence via direct or indirect means. This is much the same as claiming the existence of neurons from which a neuronal model can be built unlike the mentalistic model of how short term and long term memory works which is clearly not an isomorphic representation of what constitutes memory (Worrall, 2002). The greater the move away from material physical constructs the less realistic the model accounts become and vice versa. Of course this has a bearing on the level of description one has taken of the system. For instance, modelling the cognitive system through means of diagrammatic illustrations consisting of flow-charts is much less a real account of how the brain works than a model in which actual neurons are discussed.

Much has been written and debated about theory construction as it pertains to sociology (Abercrombie, Hill & Turner, 2000; Agassi, 1981; Berger, Wagner & Zelditch, 1989; John, 1980; Hughes, 1981; Riggs, 1979; Ritzer, 1992; Ryan, 1970; Skinner, 1986; Turner, 1993; Weinstein & Weinstein, 1992; Zhao, 1991). The waxing and waning of attempts to devise for it a more robust and scientific theory having passed through successive theory construction phases emphasising either mere speculation or theories emphasising data collection. Sociology like psychology travelled through qualitative and quantitative phases throughout its development acquiring for the discipline a quantitative mantle shirking the more qualitative aspects of theory development and growth. A number of researchers within science have lamented the lack of overarching frameworks within which to unify or at least to attempt to bring to their respective fields some sort of unification process⁴⁹ (Swanson, 1988; Smedslund, 2000; Vosniadou, 1996a). Runyan (2005) refers to unifying notions between theories within psychology as “cooperative empiricisms” in which there is no substantial progress or development towards a greater similarity to natural science models but a wealth of theories emanating from various strands within the larger discipline of psychology. Theory construction was at a time synonymous with the quantification of the social sciences (an issue which will be looked at in chapter 4) operating from the bottom-up as opposed to the more recent additions to this repertoire of top-down theorising characteristic of meta-theorising (Madsen, 1988, Radnitzky, 1970). John (1980) traces the development of theory construction and cites three main competing approaches within sociology, namely

⁴⁶ The author feels the need to qualify this sentence in this manner in order to avoid any relativist attacks on how science is practised and to avoid such confrontation which is deemed unnecessary and inhibitive of how science is in fact practised.

⁴⁷ The fact that the presence of electrons, mass of photons and even black holes for instance has to be inferred and cannot be directly observed (MacKay, 1995; Rossi, 1997).

⁴⁸ See chapter 2 in which I state my predilection for a realist view although being very cognisant of the relativist claim.

⁴⁹ It is a somewhat misguided notion that the natural sciences are unified. The social sciences too are not unified, what hope for a unification of all knowledge procurement activities?

- propositional theory (advanced by Blalock), in which the core propeller is the generalisability of empirical findings where the system is predicated on axioms, also referred to as primitives (Hempel, 1970; Reynolds, 1971). These are of course not testable and are taken as a given within a mathematical system for instance and are synonymous with “initial conditions”⁵⁰ in astrophysics (Barrow, 1992; Smedslund, 2000). They can perhaps be empirically identified by the circular definitions which start to occur when attempting to define them (De Grolier, 1990) and theories, which when taken as a whole can be logically deduced within the system. Hypotheses are testable as they are operationalised as are the eventual theories. Proceeding on a basis of verificationism will flounder however as there is no real progress towards suitable theory selection but a mere trimming down of those theories already available. This type of theory construction also belongs to the falsificationist school in which all theories are merely temporary explanations proceeding along a continuum of ever-increasing explanatory prowess (Bohm, 1997) but never really achieving all-knowing status. Generality and empirical support are the main drivers of this type of theory construction and due its confluence of positivism with advances in logic and mathematics this was typically referred to as “the received” or “standard” view (Capaldi & Proctor, 2000; Craver, 2002; Leahey, 2000; Lohman, 1997b; McErlean, 2000; Outhwaite, 1987; Shadish, Houts, Gholson & Neimeyer, 1989). This view greatly influenced psychology from the 1930’s - 1960’s, although there were burgeoning movements away from the “restrictive” account of science in the 1950’s (Outhwaite, 1987). Propositional and predicate logic (inductive and confirmatory; Salmon, 2001) enabled the clarification of explicated systems and as such logico-mathematical language, due to its precision, became the preferred choice among early philosophers of science (Machamer, 2002). As philosopher Dennett (1999) states, for good science, logic needs to be assessed at the most fundamental level, “down in the trenches, where the data are gathered, where the details matter, where relatively small-scale hypotheses about fundamental phenomena can be actually tested” (p.266). Logic is of two types, deductive and inductive, the latter emphasising the classification of arguments as valid or not whilst the former measuring the inductive strength of arguments (Skyrms, 1993) which is a logic most often associated with the social sciences, yet very problematic (Shames, 1987a). McMullin (1978) cautions against the unrestrained use of logic in all matters scientific by referring to philosophers’ hubris in attempting to dictate theory of science solely from the logical standpoint and advises that logic should only be employed when and if scientific practice becomes side-tracked or “corrupted” and “degenerated” (p.222). A science, any science for that matter cannot only be predicated on logic alone, for reality surely does not conform to logical dictates at every turn, however it is a very useful tool for counteracting illogicalities within explanations and as logic is the “study of the canons of principles of correct reasoning” (Harré, 1988) it has much to offer theory development and science as a whole. Logic was once thought to undergird mathematical foundations, an issue discussed in chapter 4. Logical predicates seem not only to pervade science as method but mathematical deduction too; areas which both play important roles in how psychological assessment has proceeded throughout the twentieth century. The logic of null hypothesis significance testing utilised to extremes within psychological research echoes the perceived utility of logical deduction and thought process and this too is discussed in chapter 4
- grounded theory (advanced by Glaser and Strauss), which also seeks to generalise findings from the empirical data but does so in a more phenomenological manner. Although touted as emergent and grounded, the approach is quite objective with *a priori* assumptions pervading this approach, making it at times, less phenomenological than initially described
- exact theory (advanced by Willer), advocates that, unlike propositional theory construction in which phenomena are considered as ordering the experiences of the observers, the theorising orders the phenomena and is akin to much of natural science theorising. The theory guides the understanding of the phenomena (Quinn, 1998). Scope and isomorphism (a one-to-one mapping of levels, see chapter 2) are imperative to this type of theory construction in which theoretic calculus for instance, is used to map and predict the model which in turn is compared to the empirical data. Empirical and rational orientations are somehow bridged in Willer’s exact theory approach but cannot be equated with either of the afore-mentioned two approaches

This list is similar in nature to the three forms that social science enquiry takes; namely, empirical-analytic, historical-hermeneutical and critically orientated views (Shames, 1987b). Simplified models have been construed as theories serving to explain the functioning of a system in its most idealised way ignoring factors which would merely upset the model. Yet it’s main aim is to provide adequate explanations at a level more understandable than the level at which the observations in fact take place (Leahey, 2000). Intricately involved in theory and model development are a number of terms which are loosely defined within these larger systems and include among others, terms such as hypothesis, proposition, schemata and concept. Gerstlé (1989) questions the origin of concept formation as emanating from theory or as becoming theory itself. In other words are they independent terms or are they dependent on the theory itself?

⁵⁰ The Big Bang.

3.4.2.1 Concepts, definitions and propositions

If concepts are considered as elements of theory is it necessitated of them that they be clear, concise and must strive to delimit extraneous meaning which will only serve to confuse any meaning inherent in the term. "Ambiguity, homonymy, synonymy, existential vagueness, opacity and contradiction must be done away with by means of conceptual and terminological analysis and, in particular, through the process of formation and reconstruction of concepts" (Gerstlé, 1989, p.608). This take on what a concept should be is by extension a positivist one which highlights the need for closure and stability of concept terminology. In stark contrast to this definition of what constitutes a concept, the more essentialist and interpretative approach towards concept definition and understanding is less confined and constricted by the framework in which it operates, here it is more open to change and not stifled by the inherent limitations in language.

If concepts are considered dependent on theory, theory is foremost as any observation will be inherently theory-laden and hence theory is prime followed by the enclosed concept which is then utilised by theory. Gerstlé (1989) maintains that from this viewpoint words and meanings should not take away from the ideal of truth within theory. If this be the case, concepts are constrained by the level of theorisation and even by the general conception of scientific work. Gerstlé (1989) compares the two origins of concept formation and is illustrated with slight attenuation in Table 6 below.

Table 6 The relation of concept to theory (Gerstlé, 1989)

Concepts (as words)	Resulting in	Theory
Designation or terms or concepts		statements or propositions or theories
Words	may be formulated in	assertions
Meaningful concepts	which may be	true
Meaning	and their	truth
Definitions	may be reduced by way of	derivations
Undefined concepts	to that of	primitive propositions
Meaning	is	truth

Riggs (1979) designates for the area of concept study a term devised by linguists, known as "terminology" and is a necessity if researchers between and within disciplines are to understand jargon-filled research papers often pertaining to very similar areas of concern. Dissenters are against such a view of course, denouncing the lack of metaphorical writing within the social sciences, stating that this aspect is what makes the sciences social in the first place. Once again, there is a continual tension between what is construed as science and what is not. Three modes within the terminological approach each emphasis a different strategy towards concept formation and understanding, namely the analytic mode which investigates the lexicographic characteristics of terms, the normative mode which attests to the one-word-one-meaning idea and the synthetic mode which, according to Riggs (1979), overcomes the limitations imposed within the former two modes by advocating a stance evidencing a mix of both modes whilst allowing for greater range than either of them. Semantic and conceptual problems which plague the social sciences (see chapter 2) are thought of as severe limitations in developing some sort of consistent and consensus-forming dictionary for the social sciences (De Grolier, 1990). That various strategies have had to be abandoned can be considered as neither "good" nor "bad" for a science of social enquiry. Yet the point here is not only that there is frequently a lack of consensus towards terminology and concepts across the social sciences but that this is a hallmark feature arising from one single discipline as well, such as psychology.

If concepts are dependent on theory and as such yield arrays of meanings what is to become of the theories? Following on from this argument, then what is to become of the hypotheses and underling implicitly stated meta-theories? Vague and ill-defined concepts where authors do not explicitly state what exactly they mean by these words (De Grolier, 1990) does not auger well for future inter-disciplinary collaboration which will of course become necessary as time progresses (this might be akin to the "levels of description" problem across which scientists will have to forge conversations with one another). With regards to this terminology, "what matters is not the words we use but the distinctions which they mark, and it is important that there should be agreement on which distinctions we have to mark" (Ryan, 1970, p.94). Fawcett and Downs (1992) state that once concepts have been highlighted they need to be classified on the basis of their variability (uni-dimensional versus multidimensional concepts such as gender and degree of aggression for instance), extent of their observability (concepts which are directly or indirectly observable, inferred constructs and theoretical terms) and measurement characteristics (enumerative, associative, relational, statistical and summative units of measures).

Harré (1988) distinguishes two broad categories into which scientific concepts can be placed, namely, material concepts (or concrete as described by Reynolds, 1971) which include terms and concepts such as mass, force, length, charge and are utilised as descriptions of things, materials and processes and formal concepts (organizational or structural concepts) which include terms such as causation, existence and identity as well as spatial and temporal concepts. These terms are viewed as



relations of things to other things, in other words, something is termed as existing now or before and nothing new is learned from this term other than its placement within space and time (referred to as abstract concepts by Reynolds, 1971). Can this distinction be drawn in psychology for instance? Intelligence is a reified construct and can be determined via material concepts such as scores in tests of intelligence⁵¹ and its reified existence as “something” measurable in time and space co-exists as formal concept. Yet, Harré (1988) probably did not have this in mind when he classified concepts into these categories. So much has been written on the philosophy of the natural sciences, that when employing arguments in favour or in contradistinction, very few examples arise when discussing similar issues within the social sciences.⁵² Fawcett and Downs (1992) are quite explicit about how carefully concepts and propositions should be constructed and analysed in theory, as ill-defined concepts with poor operationalisation connecting them to empirical data merely serve to blur the understanding of what it is that is trying to be derived within the theory. As discussed in chapter 1, Sternberg and Grigorenko (2001a) maintain that current conceptualisations of dynamic assessment often results in ill-defined conceptualisation, operationalisation and the subsequent analysis of the construct. Regarding specification of propositions in formalised theory, table 7 illustrates the typology of propositions.

Table 7 Typology of propositions (Fawcett & Downs, 1992)

Types of propositions		
Nonrelational	Relational	
Existence propositions	Existence of relationship	Direction of relationship
Definitional propositions	Shape of relationship	Strength of relationship
Empirical indicators	Symmetry of relationship	Concurrent and sequential relationships
	Deterministic and probabilistic relationships	Necessary and substitutable relationships
	Sufficient and contingent relationships	

The following discussion on propositions follows Fawcett and Downs (1992) discourse very closely. Propositions are declarative statements incorporating concepts and seek to assert statements. Statements linking or describing concepts are either of the nonrelational or relational type. Nonrelational propositions state something about a concept and either state that the concept exists or defines the concept and consists of the following sub-types:

- i. existence propositions state that a concept such as learning potential exists and existence propositions can also explicate the level at which the concept functions, such as low or high learning potential;
- ii. definitional propositions actually involve descriptions and characterisations of the concept. Convention often drives the acceptance of definitional statements and is thus not verifiable but accepted as such by the community of researchers for instance. In order for the theory to be testable however, such definitional concepts will need to be constitutive (a definition is clarified by means of other concepts, i.e. learning potential is the degree to which improvement is evidenced from mediation) and operational (a definition is clarified by means of empirical or observational data which now involves rules of correspondence, i.e. learning potential is derived from subtracting the pretest score from the posttest score). Operational definitions are either derived from measures or experiments. Correspondence rules are necessary to link concepts to their real world counterparts and is illustrated in figure 33. The often tenuous link between theoretical entity and empirical construct manifests too within dynamic assessment’s conceptualisation of the concept of learning potential which is yet another term without clarity and can also be defined in very much the same way that intelligence is defined by the manner in which it is assessed (Elliott & Lauchlan, 1997). As intelligence is varied in its multitudinous manifestations, psychometric measures will thus only “pick” up on measures amenable to such testing whereas other forms of intelligence manifestation necessitate other mechanisms of elicitation (Fuster, 2005). In chapter 2, it was mentioned that bridge principles could be equated with correspondence rules and this is diagrammatised in figure 34 below
- iii. empirical indicators are the instruments used to derive measures from concepts which are in this manner operationalised. Although not a part of the theory as such, these are part and parcel of the research design process (Fawcett & Downs, 1992; Kerlinger, 1981).

Relational propositions link two or more concepts and as with nonrelational propositions also declare a statement but include various associations between the concepts. (See section 4.4.1.3 in which measurement of constructs is based on rules of representation via relations “among objects which are reflected by the corresponding relations among the numbers assigned to

⁵¹ The author realises that Boring’s circular and not very helpful definition is being employed here.

⁵² Once again, a retort to this would be that social sciences should not predicate themselves upon natural science foundations of explanatory mechanisms. Well then, what is one to do? Devise and invent a new framework of explanatory criteria? Perhaps.

them” (Coombs, Dawes & Tversky, 1970, p.29). Patterns of co-variation between concepts occur and various relations can be identified. Types of relations include:

- i. the existence of a relationship merely states that some relation exists which is usually a recurring relation and hence not a once-off phenomena. For instance, mediation co-occurs with increased posttest scores (note that at this stage no directionality is given to the relation, only that it exists)
- ii. the directionality of the relation asserts the degree to which two or more phenomena co-occur and how they manifest in this co-occurrence. For instance, mediation resulted in increased posttest scores
- iii. the shape of the relation indicates the nature of the mathematical function of the relation; be it linear, quadratic or cubic. For instance, graduated prompts as mediation strategy might rely on an index of hints given and can perhaps be said to link to posttest score results, however with less strict definitions of mediation (more clinical approaches), this becomes increasingly difficult and at times impossible to calculate
- iv. the strength of relation concerns the degree to which two or more concepts are linked and how they are linked via a scale delineating the strength of these relations. Effect sizes are often utilised to illustrate the strength of these relations; appendix 1 evidences meta-analytic effect size results for a number of studies utilising dynamic assessment as strategy as opposed to normal static based assessment approaches and yields information regarding the strength of the cumulated relation
- v. the symmetry of the relation refers to the manner in which two or more concepts (which are at times referred to as variables) are related. There may be a strict asymmetrical relation between two concepts resulting in a one directional interaction. For instance, after mediation a difference in posttest scores will result but if there is no mediation there might still be a difference in posttest scores; i.e. nothing conclusive can be stated about the posttest scores in relation to the mediation (it could be that practice effects result in changes in the posttest scores, or other extraneous factors such as fear, anxiety and other non-cognitive impingements). Symmetrical relations evidence a circularity of sorts, with two or more concepts falling back on each other; i.e. reduction in anxiety leads to better posttest scores and better posttest scores lead to reduced anxiety
- vi. concurrent and sequential relations evidence when one or more concepts is followed by another concept either simultaneously or at a later time. Increases in posttest scores result after mediatory interventions, thus the one follows sequentially from the other. If both concepts appear at the same time, the two are said to co-occur; for instance, anxiety and pretest scores might co-occur if respondents have not been told about the testing procedure, i.e. writing a test results in immediate anxiety
- vii. deterministic and probabilistic relations refer to the degree of certainty that a relation will occur; for instance deterministic relations are often referred to as laws (Cattell, 1988b) or universal truths with one concept always preceding another or vice versa. For instance, lack of parental and/or cultural mediation can result in poor performance on various psychological assessments and can even result in maladaptation to any one specific culture (of course the more social-psychological the arguments become the less likely they are to ascribe to such lawfulness). Probabilistic relations state that there is a chance of a co-occurrence of variables given certain conditions. It is possible that there is a greater likelihood of increased posttest score if mediation precedes posttesting. This is not always the case, but it is more probable if no mediation is given. Rasch (1980) points out the trend away from strong determinism within physics towards more probabilistic models as encompassing more relativistic and accommodating models. Such probable models are better able to explain certain anomalies far better than models previously relied on (for instance Rasch cites radioactive emission). Rasch (1980) maintains that, similarly, probabilistic models can be successfully utilised within psychology in an attempt to better explain certain findings on tests of achievement for instance as human behaviour is far more complex than even models of radioactive emissions (see chapter 4). Rasch, who contributed his probabilistic model to psychometrics, was first and foremost a mathematician (Wright, 1980). Moreover, Cattell (1988b) maintains that models are by their nature theories reduced to some mathematical, physical or symbolic essentials
- viii. necessary and substitutable relations refer to the need for certain or specific co-occurring variables within a relation. For necessary relations specific variables need to occur for the relation to manifest; i.e. if and only if mediation takes place will an increase in posttest scores result (this is not an instance of what truly does occur in reality, but this example is employed merely to illustrate the meaning of this type of relation). Substitutable relations occur when either of two variables can result in a relation manifesting the same thing, i.e. either practice effects or mediation effects can result in increased posttest scores
- ix. sufficient and contingent relations refer to the conditional nature of a concept on the relation; i.e. reduction in anxiety before writing a posttest is sufficient to aid in the increase of posttest scores but not necessary. Pretesting can result in increased posttest scores, but an intervening third variable, namely mediation, can contingently alter the relation in a positive or negative manner



Figure 33 Concept-empirical indicator link via operational definition with example (Fawcett & Downs, 1992, p.28)

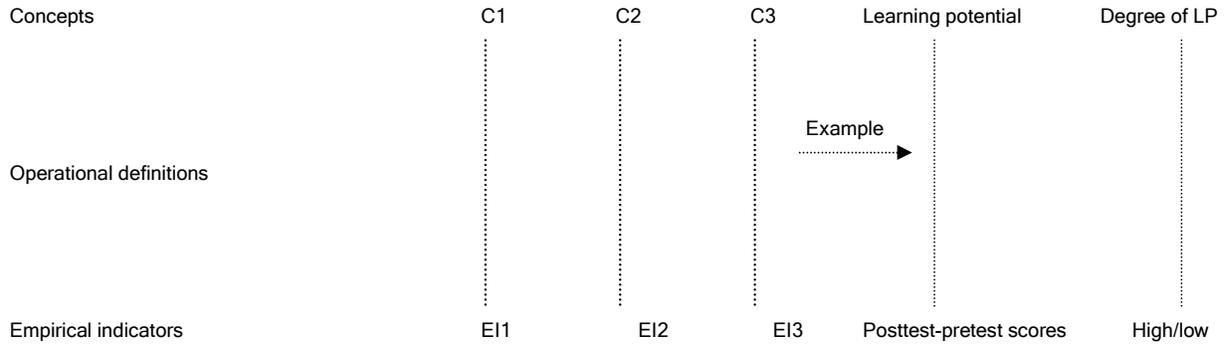
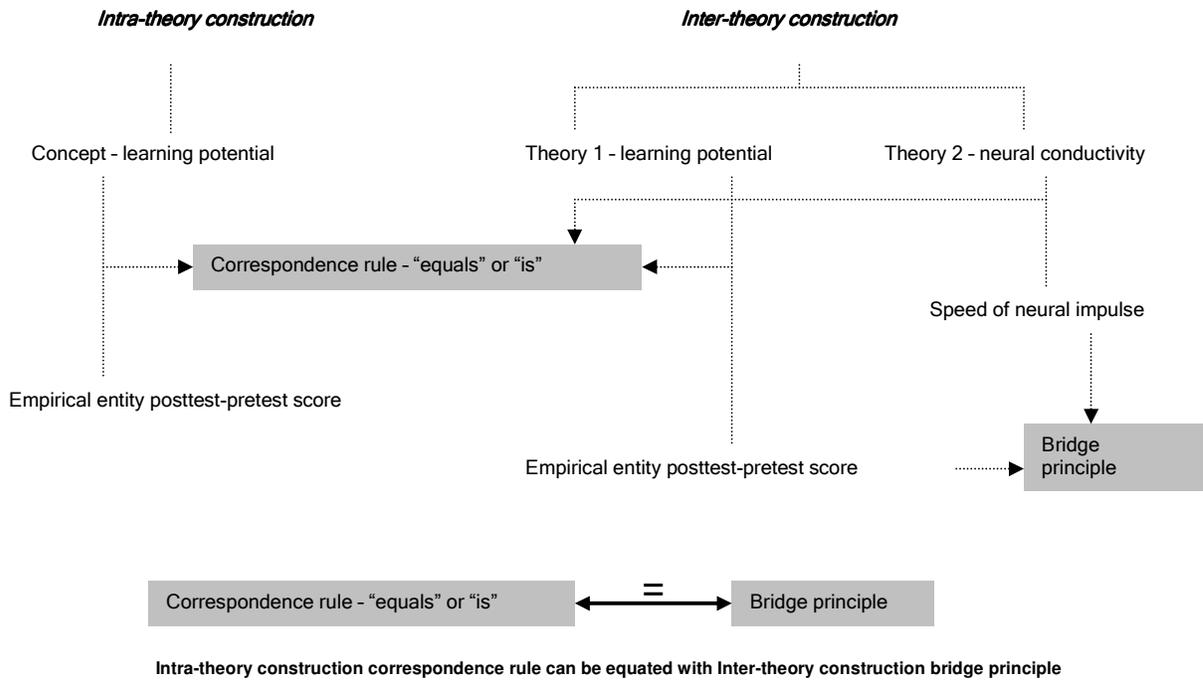


Figure 34 Relation between intra and inter-theory correspondence rules and bridge principles





Propositions can also be classified according to the degree of abstraction involved and range from restricted concretised propositions which are very specific in nature to unrestricted abstract propositions which are more general in nature (Gibson, 1960 in Fawcett & Downs, 1992). An illustration will suffice:

- Abstract proposition - learning potential is evidenced via mediatory interventions
- Concrete proposition - Increases in posttest scores often result from two-way communication processes during which various mediatory interventions are employed in assisting the learner with problematic areas of cognitive processing
- More concretised - posttest raw scores are higher than pretest raw scores after an intervention strategy aimed at improving certain cognitive functions (such as helping the learner to become less impulsive when answering or working out problems)

3.5 Theory appraisal

Methods of knowledge acquisition are only as good as their theoretical guides and theories are only as good as what precisely? How does one determine whether or not a theory is considered “good” or “bad”? (Eysenck, 1987). Unfortunately there simply is no universally agreed upon list of definitional criteria against which to score or “grade” various theories (Michalos, 1971). Any attempt is itself predicated on a theory of growth! (Serlin & Lapsley, 1990). This is neither good nor bad, for it is not a value-loaded statement. In fact critics of theory appraisal have referred to theory appraisal criteria as banal (criterion of parsimony) and misleading (criterion of empirical validity) (Gergen, 1987b). Such extreme views should, however, be tempered for due regard for how theories are in fact rated in reality. That the list of such criteria should be malleable is perhaps closer to what can be expected from theories within varying disciplines. The underlying assumption is, for the moment, that such a list is in fact warranted. Such lists’ usual reliance is upon logical positivist approaches towards theory appraisal and due to the rejection of strict logical positivism within psychology (having gone out of favour). Eysenck (1987), taking his lead from Lakatos states that a theory is good or bad if the former advances progressively and the latter degenerates progressively and so puts an end to the need for any list at all. Even naïve views concerning theory would advocate (perhaps simplistically but not altogether incorrectly) that natural science theories are somehow more robust than their social science counterparts. Psychological theories are more often than not a bit wishy-washy and ill-defined, or at the very least, not very generalisable beyond very narrowly specified contexts. Von Bertalanffy (1970) states in this regard that psychological theories which are mathematically formulated and which permit quantitative prediction are more viable as opposed to the loosely constructed verbal models and theories. Positing point estimates (as in null hypothesis significance testing) is often viewed as more scientific than conjectures utilising ranges (Meehl, 1997). It is this very issue which is criticised in psychological research in which point estimates are unreliable due to the complexity of the subject matter which prefer to make use of composites (see chapter 4).⁵³ There will of course be opposition to this sentiment from many corners of the discipline. So then, how to define what a makes a theory worth its status as theory? The literature on this topic frequently cites a number of guiding principles, at times applicable only to natural science disciplines, at times attenuated for the social sciences and yet in some instances it remains ominously silent on how to adjust the list for the social sciences. Broadly speaking, McMullin (1978) defines four main logical modes of theory appraisal that have made their presence in the history of science, namely the

- i. intuitive-deductive model - deductive arguments are drawn from plausible axiomatised sets of intuitively derived assumptions and theory results, of which these axiomatised truths are beliefs in phenomena (based on the historical record of their occurrence or not; i.e. the sun rises every morning and sets each evening,⁵⁴ therefore stating that the sun will rise tomorrow is merely playing to past occurrences). These *a priori* principles become established as truths (Hempel, 1970). This Aristotelian ideal was focused upon by Descartes and found its way as ideal method in some parts of science (Moring, 2002). However no generally secure principles exist from which to deduce, unless one can prove the axioms, which one cannot as they are intuitive! (akin to transcendental argumentation which relies on *a priori* conditions of knowledge; Mautner, 2000; something which Hume argued vehemently against; Delius, Gatzemeier, Sertcan & Wünsch, 2000; Royce, Coward, Egan, Kessel & Mos, 1978; Shames, 1987b and who also referred to as unconscious statistical inference). This ties up with the discussion on null hypothesis significance testing which, as a tool of inductive inference, cannot be logically defended but can stand up to practical scrutiny even though effects are due to probable and possible causes (Krueger, 2001; Mislevy, 1995)
- ii. inductive strategy - commonly referred to as the “Baconian method” in which there are no precisely defined formal rules dictating this method and generalisation can be plausible at best. Modern renderings of Baconianism include exploratory descriptive statistics which can be traced back to K. Pearson, Galton, Quetelet and Laplace (Mulaik, 1988). Bacon coined

⁵³ Although not discussed in this thesis, the works of Dawes (1971, 1989, 1995) offers further insights into the nature of clinical and statistical decision making, which has pertinence to the domain of dynamic assessment due to its reliance on both qualitative and quantitative methodology. See chapter 5 section 5.2.11 where this issue is brought up in relation to one dynamic assessment model in particular.

⁵⁴ Of course this does not even happen, but it becomes cumbersome to state the case by referring to “the earth rotates every 24 hours and will do so in the next 24 hours”!



the term in 1620 and although he⁵⁵ advocated other methods in addition to the inductive (Harré, 1988; Reynolds, 1971) he ironically put forward his appeals against such a method (Shames, 1987b). Bacon's appeal that the tempered nature of Greek contemplation to be infused with the nature of experience (Bacon, 1959) is perhaps such an old notion today that it is difficult to appreciate the leap that was made at the time of his writing. The line of thought proceeds from observation-statements to some sort of generalisation without producing concepts that were not stipulated in the original observation-statements (Bacon, Mill, Herschel and Ockham). There is the danger of ignoring possibly relevant phenomena during an experiment and thus excluding possible variables which might enhance the theory. Also this type of reasoning is nondemonstrative since it will always remain a possibility that an hypothesis is false even though the data are correct. This is so because the truth of the data does not necessarily entail the truth of the hypothesis (Lipton, 2001) and inductive conclusions go beyond the evidence yielded by observations (Brown, 2001). The data elicit inductive implications and simultaneously endow the data with evidential support (Rozeboom, 1970). Inference to the best explanation then becomes a partial solution to the problem of inductivism (otherwise referred to as abduction or retroduction; Lipton, 2001; Mautner, 2000). How does one go about choosing this so-called best explanation? Surely this very much depends on the nature of the criteria at hand, which has already been shown to be a matter of taste and difference. Rozeboom (1970) states that this approach, although probably nearest to a scientific method per se, is still "naïve and hopeless" (p.91)⁵⁶

- iii. hypothetico-deductive model - also referred to as the retroductive strategy in which the hypothesis is warranted on the basis of logical deductions made from it (Carnap). All else besides verification and refutation (itself based on testability criteria) are unnecessary and superfluous (Rozeboom, 1970). Doubt of course will remain as one can never be entirely sure of the veracity of the claim but this strategy does not rest on secure axioms yet inferences from this model are suspect. Even though it is saddled with fallacious issues, it is considered as more versatile than the remaining three models or strategies
- iv. negative hypothetico-deductive model - which is a falsificationist take on the hypothetico-deductive model, permitting the negation of a deductively derived assertion, in other words it does not allow for positive assertions to be made (Lakatos and Popper; Nickles, 2001; Stewart, 1990). The more falsifiable a theory the better it is and the greater the challenge to falsify the theory. It may well be that a genuinely "true" theory will not be falsified at all, but the point is that it should leave itself open to falsifiability (Chalmers, 1999). There are cases in the history of science where, had theories been rejected on the basis of false findings rejection would have been inevitable but they remained to be vindicated at a later stage (Royce, 1978). This is similar to what Eysenck (1987) refers to as weak theories which are unable to explain fully all phenomena but remain good theories for decades. This brings into question of course the counter claims or competitor theory claims (see discussion on competitor theories). Naïve falsificationism tends to treat singular accounts of theories as opposed to more sophisticated falsificationism which investigates competing theories and research programmes and as such concerns itself with the growth of science as a whole. However, predicating the growth of science on the falsificationist criterion alone is somewhat misleading. Little is learned from the falsification of a bold conjecture and likewise little is learned from a confirmation of a cautious conjecture (Chalmers, 1999)

The first three can be classified as verificationist models and the fourth as falsificationist in strategy. Necessary knowledge is obtained in the first mentioned model but this strategy relies on intuitively forming first-order assumptions or axioms which are open to debate. However, the falsificationist strategy is not a strategy often followed in working science. As much has been said and aimed at decrying the need to falsify when in fact the business of science is to verify something, not falsify it. Some regard the inductive and hypothetico-deductive strategies as two aspects of the same underlying reasoning procedure or as two fundamentally different ways of explanation (Harré, 1988). Rozeboom (1970) adds another model, namely, the omnitheoretic model in which a Feyerbandian theory-free base of observation is assumed. Figure 35 illustrates the difference between inductive and deductive explanatory mechanisms in science in which theory-generating or theory-testing propositions take the form of inductive and deductive reasoning strategies respectively (Fawcett & Downs, 1992). The scientific method necessitates both inductive as well as deductive processes (Kline, 1998) as exemplified within the inductive-hypothetico-deductive method of science as discussed above. The spiralling effect which is both engulfed by and is perpetuated by the scientific method is illustrated in figure 36 below.

⁵⁵ Bacon is in fact referred to as a "metascientist per excellence" as he contributed to knowledge about science more so than in science (Oldroyd, 1986, p.59)

⁵⁶ This is viewed as the closest model to a science method per se and simultaneously viewed as hopeless. Where does this leave us? Science proceeds with noticeable effects yet philosophers and scientists are unable to agree as to the method of its momentum; this all seems rather odd to say the least.

Figure 35 The basic difference between inductively and deductively derived theory (Oldroyd, 1986, p.62)

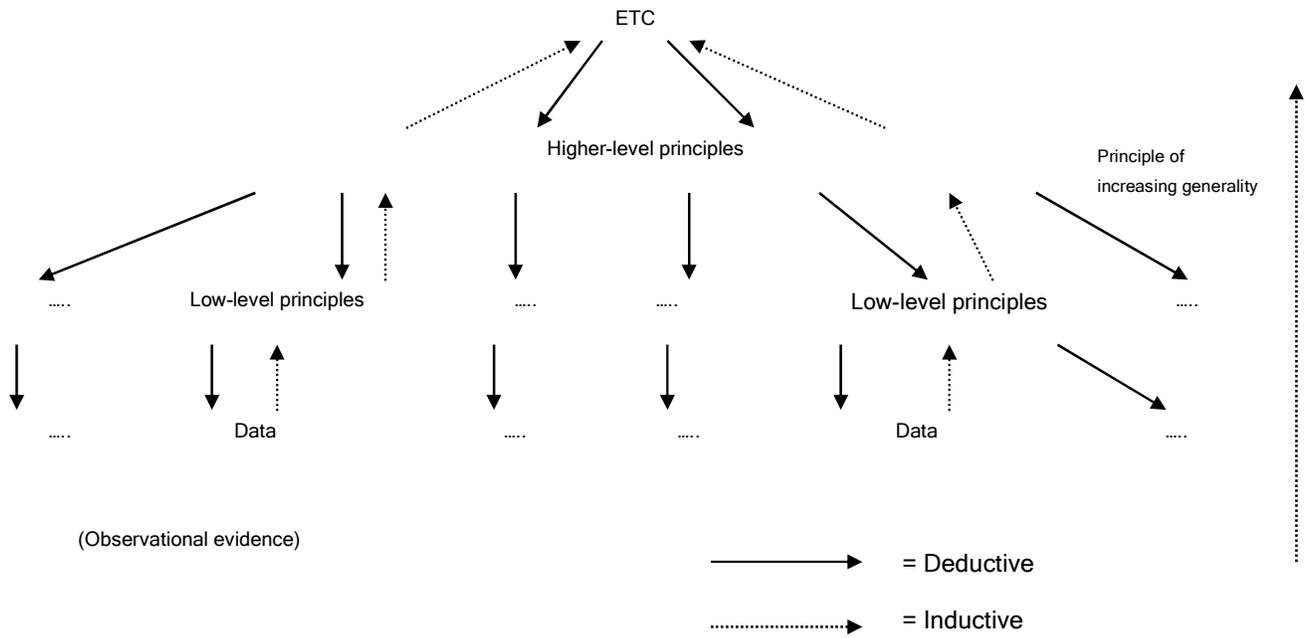
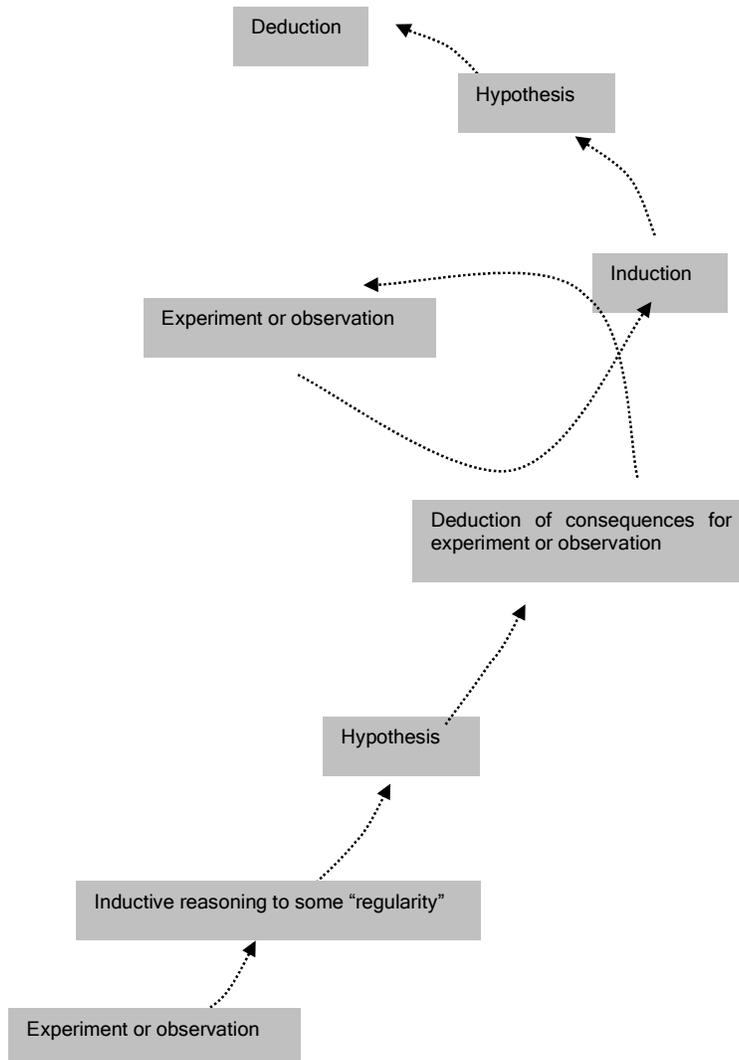


Figure 36 Cattell's (1988a, p.17) inductive-hypothetico-deductive spiral





There are of course various problematic issues surrounding inductively and deductively derived knowledge. Logically deducing a valid explanation from valid premises does not mean that the conclusion reached is true, no matter how valid the structure of the argument is. Inductive “deduction” is likewise saddled with problems, namely that in order for a conclusion to be reached an argument necessitates premises that are wide ranging, numerous in scope and applicable in as many contexts as possible and of course, there is no logic involved as such (not valid; Hájek & Hall, 2002) in inductive arguments as with deductive arguments (Chalmers, 1999). Inductive arguments can never be “proven” as such, as the method employed to do so is itself an inductive one and such circularity is doubtless erroneous, as originally formulated by Hume (Mautner, 2000; Shames, 1987a). Popper alternatively stated that science is inevitably carried on in an hypothetico-deductive manner and not in an inductive manner (Mautner, 2000). Errors in science abound, flops are universal yet there is progress, slow at times, but technological advances and numerous examples of success is everywhere evident and can attest to the limited veracity of the general claim of science. What is going on? Do the methods work or don't they? Can we successfully apply these methods to the social sciences? Negative hypothetico-deductive or falsificationist accounts of science are not impressed with yet more empirical evidence in favour of a theory. This is so because it simply does not leave itself open to falsification, which is the whole point with this method and as a result inductivists are at odds with falsificationists as they regard an ever increasing platform of empirical results as evidencing greater probability that their theory is correct (although not logically deducible) (Chalmers, 1999; Hájek & Hall, 2002; Trout, 2001b). Yet can a theory ever be completely falsified if, on account of some extraneous variables within the test situation, the result becomes coloured due to the impediment of the experiment? So it is not one theory per se which is tested but rather clusters of similar theories (Chalmers, 1999; Mautner, 2001),⁵⁷ attending to the holistic version of research programmes (Hookway, 2001).

Not only are there intractable and at times insoluble problems within the practice of science itself, but these problems then cascade down through to the level of empirical data collection and observation; which, when mixed in with the often intractable problems facing dynamic assessment become almost overwhelming. The theory is problematic; the guiding meta-theory is problematic, the omnipresent epistemological and ontological issues are problematic, the nature of this type of knowledge acquisition is problematic and so on. Is there a way out?⁵⁸ Matters pertaining to natural science philosophy and what is considered “science” have not yet been agreed upon, much less have any agreements within the social science as to what “science” is been reached. If the latter is to assume the cloak of the former, it too will be burdened with just the same baggage (Turner & Roth, 2003). Lakatos' methodology of scientific research programmes finds a midway mark between Popper's negativist account of falsificationism and Kuhn's view of scientific change which advocates that if a research programme working through auxiliary assumptions is able to generate new theories or changes to existing theory and so evidencing novel predictions then this is construed as progressive. Yet modifications which offer no new predictions and close off the theory to falsification are degenerate. Research programmes whose theories are able to make novel predictions and that also outrun empirical growth are construed as progressive. Yet lags in theory growth and development whilst accompanied by strides in empirical development are seen as stagnant (Nickles, 2001). The latter is reminiscent of the current state of dynamic assessment.

Philosophers of science busy themselves with what constitutes theory, be it good or bad, useful or not. More often than not theories are credited with no more veracity than the intuitive appraisals which scientists invoke for them. A theory has to have a point of origin somewhere and this demarcated line comes into existence in a rather clinical fashion (Kukla, 1990a) and subsequently analysed as such but in an attempt to order and objectify a process of theory appraisal, Meehl (1992) develops a statistical framework for such an endeavour, a quantitative framework which aims to appraise theories in terms of verisimilitude (McMullin, 1990). The plight of theory is similar to what constitutes fashion, at times certain looks are in and during others they are out. Certain theories perennially make their appearance only to disappear again. It is perhaps a “good” thing that no one specific theory remain in existence for too long, as this would seem to indicate almost perfect knowledge on the topic (which is of course not unreasonable per se, but highly unlikely in the social sciences as it has thus far evidenced) and as history has illustrated theories grow and develop (Whitt, 1988). This reminds one again of the falsificationist/verificationist account of science attributable to Popper. Strict positivistic-like theory criteria simply will not do for a social science endeavour, and this much has been acknowledged. However (and here a big however is warranted), this point is seemingly either missed or ignored by a very considerable majority of social scientists, especially those within the psychological assessment sub-discipline, who stubbornly resist a revised theory appraisal framework and insist on continuing on a trajectory planned and followed by natural science models. Who or what is the cause of this?⁵⁹ Nevertheless without neglecting the object of this study and veering off on a tangent let the situation rest there as we move on to what constitutes a statement becoming classifiable as theory.

⁵⁷ Known as the Duhem/Quine thesis (Chalmers, 1999; Hookway, 2001; Mautner, 2001).

⁵⁸ This study has often had the nasty side-effect of producing more questions than answers.

⁵⁹ An area that is ripe for debate and open to recriminations and firm rebuttals. Science (natural and social) is a sociological and psychological endeavour after all and open to all manner of abuse emanating from various hidden agendas, already alluded to in this study.



Kordig (1971) posits six guiding principles, themselves invariant across time, which seek to determine the characteristics of a good theory:

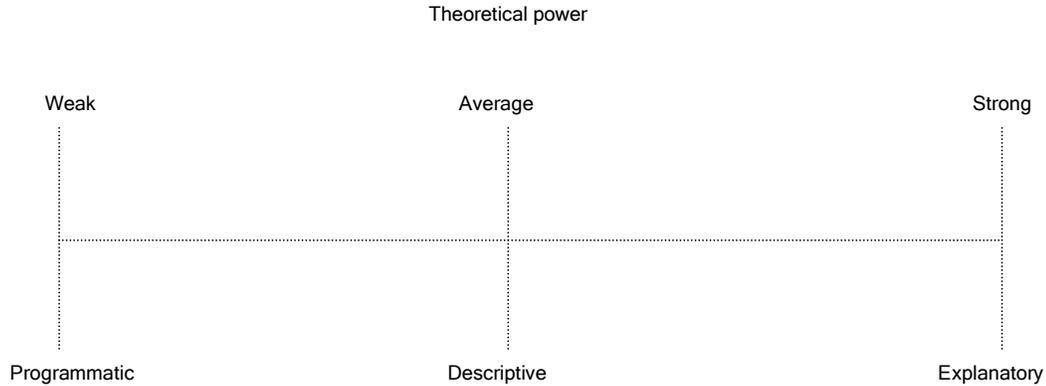
- empirical confirmation
- logical fertility
- extensibility
- multiple connection
- simplicity
- causality

A theory is deemed plausible and valid if the consequences are, via correspondence rules, in agreement with the protocol observations and facts; that is, if the theory is empirically testable (Gergen, 1986; Mouton & Marais, 1988). Correspondence rules are techniques for reifying the constructs and assigning partial observational status to theoretical terms (Suppe, 1993) depicted via the perceptual process and are equivalent to “operational definitions” utilised in the social sciences (Mouton, 1993a). In other words, intelligence as a construct might for the sake of the argument be reified as speed of neural conductivity and hence be physically measurable on a numerical scale which evidences a one-to-one correspondence between speed and accuracy on the one hand and intelligence on the other. Or it can be reified via a score on a test and is regarded as an empirical indicator (Mouton, 1993a). This is indeed how nascent views of intelligence evolved especially within the works of Galton, Cattell and Spearman (Vernon, 1986). The term intelligence is meaning variant as discussed above and thus each definition (an area with many attendant problematic issues of its own; Hempel, 1993) and its associated observation will in turn cast a shadow over the received view of theory development). Nevertheless subsequent confirmations bode well for the veracity of the theory which also ties in with a theory’s past track record, in which past instances of success bodes well for the continued use the theory in future (Faust & Meehl, 2002). Meehl (1992) refers to this theoretical criterion as consisting of the “number of corroborating facts derived and the number of dis corroborating facts derived” (p.346). Closely linked to this is the criterion of historicity or the historical criterion, which McMullin (1978) highlights as an often utilised criterion in science and theory development. If it has worked in the past then it has been validated by scientists and accepted as such. This is of course a dangerous route to follow as blind acceptance of this criterion might well lead theorists astray in what they presume to be a faultless scientific past. More errors are made than is usually stated in scientific texts.

Logical consistency is regarded by many scientists as pre-requisite for a sound theory, yet the extent to which this can be utilised as criterion for sound theory within psychology for instance is questionable. If logical coherence is an aim of regulative science (Kordig, 1971) then social science theory might well flounder in this area, although psychometrics has indeed made a valiant effort in this regard. The dynamic assessment movement may not perform well in this regard due to its context-boundedness and critical stance on normative assessment. Logic may well aid in studying the soundness and veracity of inferential claims, but as Goldman (1985) states, logic is silent on cognitive states, and can say nothing about psychological states. Secondly, formal logic itself does not consist of rules. McMullin (1978) asks whether logical rules are accepted as *a priori* principles carrying total conviction in their own right simply because they have worked in the past or do they themselves necessitate some form of practical sanction? Logic is also not equivalent to epistemic rules. It bestows on theory development a framework which aids in proper and valid rules of deduction and inference but cannot rule on inner psychological states as such. This domain is better looked at from a theory’s heuristic value which lies in the affordance of empirical results being utilised in order to refute or uphold such a theory (Archer, 1996).

Perhaps the most glaring difference between natural and social science theory is the degree of generalisation that theories evidence (but models are usually more generalisable due to their more imprecise nature; Mouton & Marais, 1988). The greater the area (space and time) covered by any one theory the better the theory (Wilson, 1999). However most tentative law-like theories emanating from psychology are not generalisable at all over and above the very specific situation to which the theory pertains. Does psychology need a list of different criteria for the determination of what is considered a good theory? Or at the very least a good model? Is a list necessitated at all? This all very much depends on where the discipline is allocated within the grander scheme of science. Newton-Smith (1981) maintains that theories ought to preserve prior observational successes and incorporate these instances into the prevailing dominant theory which is obviously more robust in terms of predictive success and even better for theories is if they can attempt solutions confronting rival theories (Lamal, 1988). In connection with this attribute of theory appraisal, theories should provide scope for future development and findings and should serve to guide, control and describe future research (Swanson, 1988). Newton-Smith (1981) refers to this as a theory’s fertility and Meehl (1992) refers to this as a theory’s deductive fertility or fruitfulness. Royce’s (1978) concurrence with the notion that psychology as a discipline “should drop the scientific charade” (p.262) and enter into discourse seeking psychological explanations is maintained when it is highlighted that psychology’s greatest contribution to date has been statistical, observational and methodological (Shames, 1987b) but it has been held back due to its unsophistication in theory development. Royce (1978; 1987) presents on a continuum, a scale illustrating the range of theoretical power that theories in general possess and is illustrated in figure 37a below.

Figure 37a Continuum of theoretical power (Royce, 1978, p.262)



Likewise, Royce (1978) offers a continuum on which metatheoretical reification (“getting close to the way things are” p.264) can be placed and takes his lead from Nagel (1961)⁶⁰ who delineates a four-fold categorised taxonomy of theories; analogical, descriptive, instrumentalist and realist. In contemplating this metatheoretical continuum, analogical theories are heuristic (and not yet “the thing itself”; Royce, 1973, p.14). Descriptive theories gain credence from their empirical data reality, and the realist and instrumentalist accounts are massive theoretical undertakings and seek to maintain logical consistency as the “nomological net expands and [seeks] to maintain contact with an expanding reality” (Royce, 1978, p.265). Less developed theoretical explanations lie at the left hand side of the continuum and merely serve as aids, to “metamorphosing” into reified constructs of actual things being evidenced by realist or instrumentalist forms. This is illustrated in figure 37b below. The most obvious inference states Royce (1978) in the comparison of the two continua is that greater theoretical power leads to greater meta-theoretical reification. One can, in a manner then, compare inter-theoretical bridging principle accounts between two theories at different levels of description⁶¹ to that of a move from analogical and descriptive theories. For instance learning potential in which theoretical constructs are, via correspondence rules, identified as empirical entities; towards theories which seek to reify certain theoretical constructs via their own correspondence rules such as learning potential being expressed as neural speed of conductivity for instance (Reed & Jensen, 1992) or as a measure of utilising the brain more efficiently (Neubauer & Fink, 2005). Such construal of theoretical constructs (or statistical constructs such as fluid g) via operationalisation (as evidenced from the links between Gf and prefrontal cortex structures) can indeed be fuzzy at best within the domain of intelligence research (Kane, 2005; Wilhelm & Engle, 2005). It is at least possible to speculate about the changes in brain functioning before and after mediation in which the brain can account for more effective processing of information after having received mediation of some sorts. Learning potential as described in various dynamic assessment models is at one end of the metatheoretical continuum considered as analogical or descriptive entities moving along towards more reified concepts of learning potential as expressed by physiological programmes. Royce (1973; 1978) attenuates this model with his own terminology and is illustrated in figure 38 below. In addition to this, the levels of theoretical power are examined as ogive curves along an axis of conceptual linguistic precision and theoretical power, see figure 39. This entire account is a thesis put forward in this study, and is a thesis very much in contestation.

⁶⁰ The author has utilised Nagel’s 1979 text (chapter 6: The cognitive status of theories) - merely a later addition. Reference will be made to the 1979 text when Royce makes reference to the 1961 text.

⁶¹ In chapter 2 it was mentioned that theories at different levels of explanation need not seek unification due to the fact that they seek to explain phenomena at these different levels and can do so quite without the need to bridge to another level of description. Cognisance is taken of this, but this discussion merely illustrates how such inter-theoretic reduction takes place.

Figure 37b Continuum of metatheoretical reification (Royce, 1978, p.262)

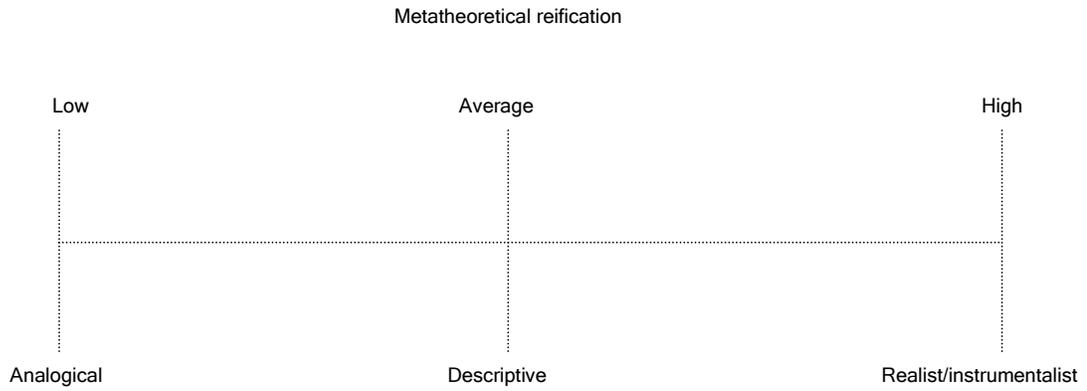


Figure 38 Continuum of metatheoretical reification with additions by Royce (1978)

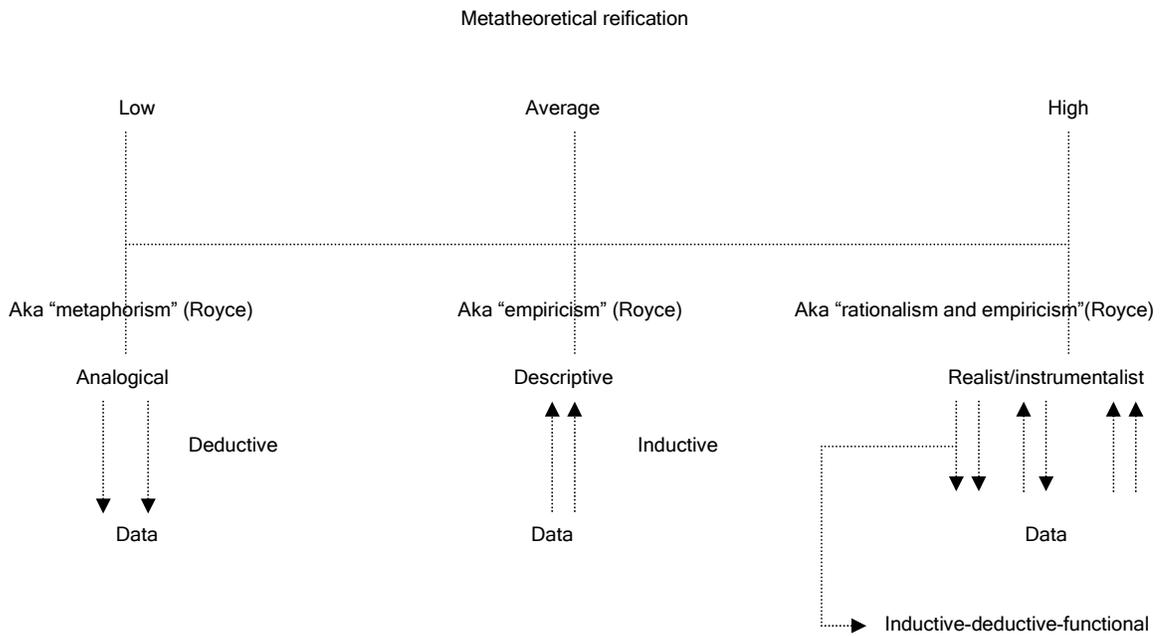
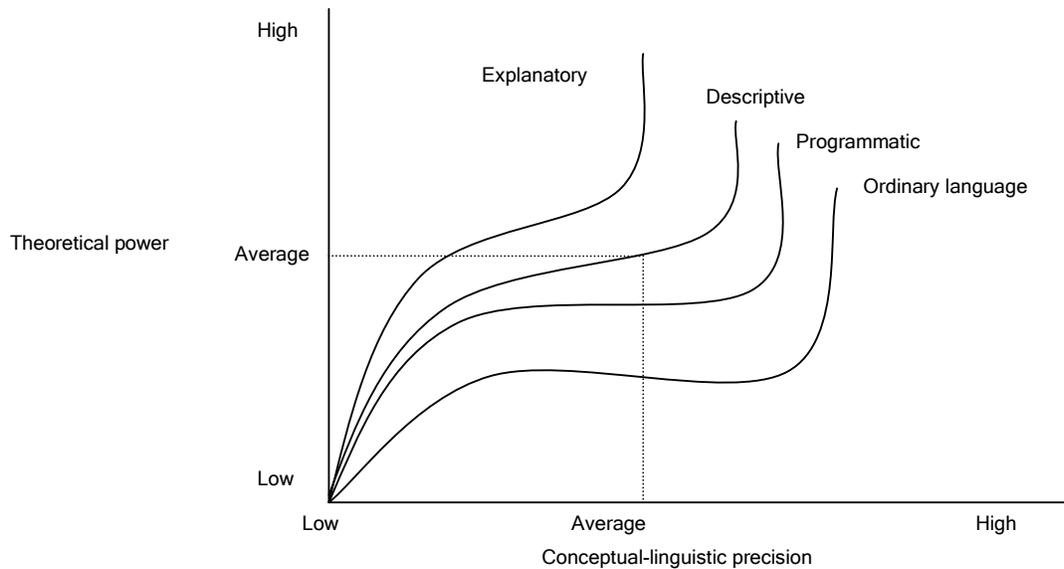


Figure 39 Dependency of theoretical power upon conceptual-linguistic precision (Royce, 1978, p.268)





Considering figures 38b and 39, the following can be stated regarding the level of understanding in a theory depending on the nature of its maturity (its maturity status being effected mainly by the complexity of the subject matter; Royce, 1987). Explanatory theories need not be as precise in linguistic formulation as programmatic theories need to be, due to the fact that the former evidences higher levels of theoretical power as it becomes more reified. In comparison, programmatic theories due to their analogical nature remain purely descriptive and do not express much theoretical power at the same level of conceptual-linguistic precision that explanatory theories manifest for instance. Ordinary language formulations of theories need to attain extremely high levels of conceptual-linguistic precision before they can even reach the low level conceptual-linguistic precision evidenced by explanatory theories. These ogive curves are worked out and conceived as such by Royce, (1978). Royce (1978) maintains that Madsen's classic on comparative motivational theories is levelled at the programmatic stage and as such Madsen is able to illustrate the "conceptual chaos" at which these theories operate (p.269) (see discussion on meta-theory). Are dynamic assessment theories manifest at programmatic levels too reflecting the broader vista of psychology as discipline in which it is still at the empirical and experimental stage?

Simplicity, parsimony or verisimilitude has often been construed as a hallmark of theory credibility (Archer, 1996; Boyer, 1995; Chaitin, 2000, 2006; Fawcett & Downs, 1992; Horwich, 1978; Kuhn, 1977; Meehl, 2002; Newton-Smith, 1981; Pascual-Leone & Sparkman, 1980; Royce, 1978; Wilson, 1999). It would seem on the face of it, that the fewer elements contained in the theory the better it is (Cohen & Nagel, 1949). Why the need for parsimony in the first place⁶² (Lohman, 1997b)? Surely two theories advocating the same correct principles are similar in terms of their being correct despite the explanatory variable utilised within the explanation? Many theories of nature posit complicated renditions of explanations and as Meehl (1992) states, are hardly ever simple either conceptually or formally. Parsimony cannot always be logically justified but defended perhaps, from a psychological position (Brandt, 1984). As Fearn (2001) states; "one must take care to use Ockham's Razor only to prune the expendable element of an explanatory theory - that is, those which do not do any useful work in explaining phenomena" (p.59). This Razor advocates a method of time saving only, and is not necessarily an insight into the world. Deployment of Ockham's parsimonious suggestion is blamed for the unrestrained use of null hypothesis significance testing (Rindskopf, 1997). Eliminating non-effectual hypothesis maintains the status quo of simplicity, it being the most admirable level at which to operate within the natural and social sciences. However, this is problematic and will be duly discussed in chapter 4. The case might well be that two theories account for the same phenomena but the one fails in terms of empirical computability from results (Horwich, 1978). Is parsimony truly a litmus test for the survival of a theory (Meehl, 2002)? Philosophers of science have not in general agreed on reasons as to why this characteristic is so important or what in fact this ideal entails (Musgrave, 1985). Conceptual confusion might arise when there are too many simultaneous variables to consider when fewer will do the job just as effectively and there have been tentative frameworks proposed for certain areas such as cognitive science for which simplicity could serve as unifying principle (Chater & Vitányi, 2003). It makes the system altogether easier to understand. Meehl (2002) states that many criteria scientists utilise to appraise theory can be correlated with verisimilitude and offers an actuarial basis for theory appraisal entitled "cliometric metatheory". Verisimilitude is important for yet another criterion in theory appraisal and that is internal consistency, for, if a theory is to be internally consistent, it would need to be remodelled if it, for instance, contained negations. This aspect of internal consistency is an *a priori* concern and needs to be ironed out before the theory is complete (Newton-Smith, 1981; Pascual-Leone & Sparkman, 1980). The value in parsimony lies in the economical use of terms and sparse or no use of theoretical leaps when presenting such as theory (Archer, 1996) an echo of Robinson (2000, p.1020) who states that "the search for the most economical description is the goal of science that, when reached, subsumes the greater number of observations under the fewest and simplest concepts". Moreover Meehl (2002) delineates four types of parsimony, namely,

- parsimony 1: simplest curves (allowing the data observed to fit the most likely statistically derived curve possible in attempting to move closer to the truth as exemplified by how the data do in fact statistically fit the theory). When various curves fit the data it is impossible to determine the correct one (after all they were inductively determined) and hence the reliance on parsimony becomes a logical next step (Harré, 1988; Sober, 2001)
- parsimony 2: economy of postulates (the one most often referred to when discussing parsimony in general)
- parsimony 3: economy of theoretical concepts
- parsimony 4: Ockham's razor (don't invent a theory to explain a new fact explainable by ensconced theory)⁶³

In fact, Chaitin (2000) offers a unique perspective on the issue of parsimony although he does so from a metamathematical position encompassing the area of computer programme structure (Casti, 2000). Chaitin (2000) maintains that the size of a

⁶² Newton-Smith (1981) poses this very question to the reader. "Relative simplicity to a large extent lies in the eyes of the theoretician and not in the theory" (p.230).

⁶³ Contrary to Feyerband's proclivity to sustain as many theoretical attempts to explain the same phenomena as possible (Meehl, 1992) but the author is agreement with Meehl (1992) on this issue, maintaining that it is more economical to stick with the theory already in place that adequately does the job. There are sufficient other theories to work on without ceasing to work on a problem which already employs a governable theory.



computer programme is equivalent to the concept of entropy and that the smaller and more elegant a programme to describe a system the better. Writing a programme for the properties of a gas, say, would be far more complex than a programme written for a crystal for instance and he states that the idea of programme size complexity is connected with the philosophy of scientific method. The smaller the programme for describing a system the better it is at explaining the phenomena⁶⁴ (Chaitin, 2006; Coveney & Highfield, 1995). However once a point of maximal complexity is reached (infinite entropy) there is no parsimonious method of describing the system and hence the system becomes its own description which nullifies the whole idea of having a theory in the first place! Is this perhaps not akin to the area of intelligence research in which the human being in interaction with the environment behaves as a maximally complex system; one which cannot be reduced to theoretical principles? And hence, the only tool useful in describing the system is the system itself? Perhaps the study of intelligence then, should end? Perhaps. Or perhaps different questions should be asked. Another allied issue to this one highlighted by Chaitin (2000) is that one can never be sure of how simplified a theory can be. When a programme is written to describe a phenomenon, a simple programme might be in the waiting; a smaller more elegant programme; how can one be sure that the current programme is the smallest one available? The answer is that one can never know; an issue which is beyond mathematical reasoning.⁶⁵ McMullin (1978) concurs with other philosophers of science when discussing theory appraisal. Issues such as simplicity, coherence, consistency, elegance and the elimination of inexplicable coincidence cannot be classified as a criterion of appraisal as such. They are therefore categorised as a systemic criterion as they are often assumed to be criteria due to their historical role in science and theory appraisal and simply because it has become conventional to do so. Hence, these criteria are classified as quasi-formal properties of theory. In fact verisimilitude, simplicity and parsimony have been viewed as aesthetic criteria, somewhat less important in comparison to other criteria which have an impact on the scientific community and society in general as well as the heuristic value of a theory (Royce, 1978). In conclusion Royce (1978) sums up various theory appraisal criteria into three broad categories which reflect different aspects of theoretical power, namely

- its cognitive and knowledge generating capacity - empirical testability, degree of empirical-formal fit, comprehensiveness of scope, parsimony, depth of penetration, degree of formalisation, degree of cohesiveness and explicitness of conceptualisation
- the social or communicational impact of the theory - impact on both the scientific community and the culture in general as well as the heuristic value which leads to more research
- the theory's aesthetic value - simplicity and parsimony (akin to the notion of mathematics' aesthetic qualities, often utilised for its heuristic power; Farnelo, 2002b; Miller, 2002; Wilczek, 2002)

Kordig (1971) maintains that causality really is a property of physical laws and is not merely regulative. In other words causality is not merely a clear stable path of cause and effect which follows in deterministic fashion but is part of the law itself. Contextual and universal causality should be also be captured within a sound theory (Mouton & Marais, 1988). Theories within the social sciences are perhaps bereft of linear causality thus making the job of theorists increasingly difficult (Wilson, 1999). The web of interacting variables with which psychology has to contend, makes the case for causality all the more complicated. To this list Newton-Smith (1981) adds that a theory's track record is important as a continuing corroboration of observations and facts bodes well for future success. Inter-theoretical support is another criterion which adds credence to a theory, yet the obvious example within physics (the incommensurable theories of General Relativity and Quantum Mechanics) seems to be at odds with this criterion. The one theory being unable to explain phenomena in terms of the other. Likewise a similar situation arises in intelligence assessment in which *g* loaded tests become less *g* loaded through mediation. Can the one theory of *g* loadedness explain the observational facts encountered in test-mediate-retest situations? This leads on to another theory appraisal criterion put forward by Newton-Smith (1981), namely that of theory smoothness which amounts to a theory's ability to adjust when it encounters failure. The greater the ease with which this is conducted, the greater the veracity and soundness of the theory.

In sum, when viewing the preceding six criteria it seems almost hopeless that psychology and its subsidiary disciplines will ever attain a state in which these criteria will apply. Either the list needs drastic revision and attenuation or the discipline needs to be overhauled in such a manner as to allow for the creation of a new list of criteria. Meehl's (1992) list is more comprehensive in that it includes a greater variety of criteria than those advocated above and includes criteria such as the amount and quality of facts derived from the theory, the novelty of these facts, their numerical precision, the theory's passive or active reducibility (theory as reduced or theory as reducer); initial plausibility (which might not be easy to detect without sufficient scrutiny of

⁶⁴ Complexity, K_x of an object, x is described by the length of the shortest programme that can be written for that object in any standard universal computing programme syntax (Chater & Vitányi, 2003).

⁶⁵ One can clearly see the progression of ideas from Gödel's incompleteness theorem, through to Turing's halting problem and on into Chaitin's theory of randomness (explicated here by computer programme problems) (Coveney & Highfield, 1995; Stewart & Golubitsky, 1993). But this does tie up with the idea of the need for parsimony, especially in the fields of computational programming, where the need is perhaps greater than in the social sciences. However, this is not to say that the same principles cannot be utilised within intelligence assessment research for instance and the need for parsimonious theories. At least this may be one tentative answer to the question of why there is a need for parsimony in the first place.



course); computational ease and somewhat less rigorous, the beauty, depth and elegance of a theory⁶⁶ (similar to Royce's, 1978 aesthetic criterion).

There is also the issue of theory as derived from data and theory as driving data which has not yet been mentioned. The necessity of initiating theoretical research with a collection of data which refutes or confirms prior renditions of theory is not in fact a necessary requirement; as a plethora or mass of data may often be quite useless without the accompanying theory behind it (Kukla, 1995a, 1995b; Strube, 2000; Swanson, 1988). This "autonomy from the data" (Kukla, 1995a, p.201) might seem an unscientific enterprise but if theory is able to grow from logical suppositions and predications which allow for the fostering of new ideas (which can be tested), then as an enterprise it should receive accommodation from practitioners of both basic (which is theory-driven) and applied research, especially within the field of intelligence research (and its varied and allied sub-disciplines) which abounds with theoretical models.

A similar set of criteria for judging the worth of hypotheses in general is delineated by Copi (1972) and includes:

- relevance
- testability
- compatibility with previously well-established hypotheses
- predictive or explanatory power

An hypothesis is only relevant to the fact that it supposedly helps explain. These facts need to be deduced from the hypothesis either directly or from a combination of causal laws. Hypotheses which are not amenable to testing (whether it be falsifiable or confirmatory testing) in either direct or indirect inferred manners cannot be considered a scientific hypothesis. It is preferable that the current hypothesis is in keeping with previously acceptable hypotheses and also that it be self-consistent. Radical departures from accepted hypotheses are suspect, at least within the natural sciences, however the same cannot be said about the social sciences, especially in an area in which there are few proven theories, hardly any laws and overwhelmingly numerous hypotheses. The number of deducible observable facts from an hypothesis credits the hypothesis' explanatory power base. The more facts that can be deduced from the hypothesis the more powerful the hypothesis. Added to this are a further four criteria advocated by Kuhn (1977) and Lamal, (1988), namely, accuracy in that consequences of the theory should be in agreement with experimental findings and observations; consistency, both internally and with other related theories; should evidence broad scope in extending beyond the strict delimited confines it originally sought to explain (referred to as progressive theories by Pascual-Leone & Sparkman, 1980) and lastly theories should prove fruitful in endeavouring to explain as many new and known variables and relations as possible. Among various theses concerning theory appraisal lurk a number of salient issues which may well render the criteria suspect. Who, after all, is to say that criteria are indeed laudable in all circumstances (Lamal, 1988)? Of course this leads down a relativist path in which criteria will have to be stipulated for each and every research programme thus nullifying their impartial and objective characteristics.

Before science can be conducted a list of criteria is necessary to determine the relative strength or weakness of theories and hypotheses, but this cannot be done if the list is not a true reflection of the prevailing conditions inherent in each programme. Are there simply no objective and all-encompassing theory appraisal criteria? Eysenck's (1987) list is perhaps one of the shortest lists and evidences only five basic and broad criteria for appraising theories, namely that a good theory should be progressive, it should explain what previously was considered as anomalies, it should act in the capacity as criterion between computing interpretations, it should attempt unification of disparate disciplines and lastly it should evidence practical value. In agreement with the fourth criterion employed by Eysenck (1987), Wilson (1999) emphasises the disparate paths that both major disciplines are presently travelling along in terms of unaccommodating theoretical links. Wilson's (1999) four qualities for any good theory encompasses parsimony, generality, consilience ("units and processes of a discipline that conform with solidly verified knowledge in other disciplines have proven consistently superior in theory and practice to units and processes that do not conform" (p.219)) and lastly productiveness. In sum we cannot judge the fruitfulness of dynamic assessment theories and models within intelligence if we do not as yet possess a criteria list; for at the outset then, we will have to admit that any such list will not suffice for all theories at all times. Once again such studies are relegated to time and place and are specifically attuned to the area under investigation. How is one to proceed? Perhaps this is the feeling with which positivist-minded social scientists are confronted. These issues are not solved within this thesis; a framework is indeed chosen and attenuated and from there the procedure follows but as to a resolution of the above-mentioned issues, no solution is yet in sight.⁶⁷

⁶⁶ The author will refrain from discussing this particular criteria as the argument can become quite lengthy and convoluted.

⁶⁷ Another topic worth pursuing then.

3.6 Towards a meta-theory

Various authors posit varying views as to what they perceive meta-theory to be. Some advocate that it is indeed a theory of theories or the study of such theories and scientific tools (Faust & Meehl, 2002; Weinstein & Weinstein, 1992). Others maintain that it is more of an overarching framework and serves guiding parameters from which to view and study other theories in a reflexive manner (Abrams & Hogg, 2004; Bjorklund, 1997; Cervone, 2000; Ritzer, 1996). It is viewed as a set of epistemological and ontological assumptions which orient and direct theorising as well as limiting the nature of psychological accounts (Gergen, 1987a; Lawler & Ford, 1993) but not necessarily prescribing new theory (Stam, 1987). Theoretical psychology as subdiscipline has even been described as a meta-discipline (Slife & Williams, 1997). One can study all theories within a discipline,⁶⁸ or theories specific to a sub-discipline⁶⁹ (Madsen, 1988). Ritzer (1996) maintains a distinction between two courses according to which meta-theorising can proceed; meta-theorising in an attempt to re-visit extant theory in order to gain a greater understanding and meta-theorising as a function of theory development. The main point of concern is of course the legitimacy of such a meta-theory, in other words is it itself an empirical exercise? Meehl (1992) claims that it is. Berger and Zelditch (1993) are of the opinion that meta-theories are nonempirical structures as opposed to testable theories (Lawler & Ford, 1993). Meehl (2002) contends that as far as meta-theory's own theory appraisal is concerned (and in keeping with the sparse literature on workable meta-theoretical frameworks), it "is surprisingly thin as to criteria" (p.345). Can it too be scrutinised along with other theories? These are difficult questions to answer especially as the enterprise of meta-theorising is not really a developed one, which is obvious when one views the myriad definitions of what constitutes a meta-theory in the first place. Does meta-theory precede theory development and act in a capacity of guider or does it seek to rectify after the fact (Berger & Zelditch, 1993)? Deliberating about such meta-scientific issues is not a seemingly straightforward one and "disciplined thinking *about* science appears to be much more difficult to achieve than disciplined thinking *within* it" (emphasis in the original) (Rozeboom, 1970).

Lawler and Ford (1993) state that meta-theory busies itself with both, in that it "not only initiate(s) theorising but often guide(s) extensions of extant theories in implicit, unacknowledged ways" (p.174). One can of course argue that there are no presuppositionless theories and as such are either implicitly or explicitly guided by some sort of metatheory, this type of meta-theory is of course not a formalised one (Lawler & Ford, 1993). A parallel between theory/experiment and meta-theory/theory can perhaps be drawn in which experimentation is the test and theory the driving force for theory development (Willer & Markovsky, 1993) where theory becomes the test and meta-theory the driving force within meta-theory development. Bjorklund (1997), in devising for cognitive development a meta-theory concerning developmental biology, contends that his is not a theory of theories but rather an overarching framework with principles and assumptions which may or may not be subject to empirical validation. His argument for a necessary framework for cognitive development is warranted at a time when in this field there is "an overabundance of details about children's performance on trivial tasks in unreal situations" (p.144), a common lament within the discipline of assessment as a whole (John, 1980) and similar in sentiment to what Turner and Roth (2003) refer to as "large body of practical activity, but not successful theory" (p.5). Diversity and specialisation can both be accommodated within an overarching framework. The rationale underlying the attenuated Madsenian metatheory framework (to be discussed below) is similar in sentiment to that written by Abrams and Roth (2004) "a metatheory should provide an alternative framework for asking particular questions, not a complete explanation for all phenomena" (p.100).

However, before embarking on a 'metaforay' it will be necessary to first highlight the differences between three interrelated meta concepts; metatheory, metamethod and metadata (Zhao, 1991).⁷⁰ What is the point of theorising about extant theories? Within metatheory, three avenues to pursue according to Zhao (1991) are the needs to further develop and attain greater insight into theories within specific research areas, to serve as prelude to new theory development and lastly to overarch some or all theories within any particular discipline. Metamethod, is the study of methods used within the specific research area and Zhao (1991) relates three types of studies that emanate from metamethodological endeavours; methodological presumptions necessary to carry out studies, evaluations of the strengths and weaknesses of various methods used in research as well as the systematising of procedural rules. Lastly, meta-data-analysis is the study of results of various studies, collating findings (the raw data) into one grand result (the meta-data). Meta-data-analysis consists of understanding the underlying assumptions utilised in analysing various data sets (as differing statistical assumptions may well result in divergent conclusions), the comparison of different forms of data as well as synthesising the results from a spread of studies (Zhao, 1991). This has resulted in the appended meta-analysis in this study although methods have been devised which synthesise narrative or qualitative results too. In contrast with Zhao's (1991) opinion in which he states that meta-studies are often confused with philosophical studies on epistemology and ontology, the author maintains that, in keeping with Madsen's metatheory fundamental epistemologies are paramount to the understandings of various models and theories and can be comfortably accommodated in a development of a

⁶⁸ Although as time progresses this seems less likely as all disciplines are so rich and varied in the branching of sub-discipline areas of expertise and specialities. It is hardly conceivable that such an undertaking would be lightly entered into to.

⁶⁹ Referred to as minitheories by Faust and Meehl (2002). There is no standardised definition of what meta-theory is and hence synonyms of this sort arise.

⁷⁰ Much of the 'meta' literature has as its focus, sociological metatheory. Zhao (1991) is one such example, for whom meta-studies centre around sociology.



meta-theoretical framework devised for the purposes of situating dynamic assessment within intelligence. Rozeboom (1995) highlights three levels from which a scientific meta-theory can be successfully conducted:

- I. the substantive level in which existing theories are critiqued. The aim is to evaluate and improve on these theories (akin to Madsen's HQ calculation discussed below)
- II. methodological meta-theory attempts to build an appraisal system of sorts according to which varying theories can be compared and analysed, thus bringing a semblance of empiricism to the technique
- III. the foundations level is concerned with basic operational questions about fundamental issues (the philosophical levels which predicate many theories either implicitly or explicitly)

According to Aldridge, Kuby and Strevy (1992, p.683) "a metatheory is a paradigm or world view" in which the social sciences are more closely tied to general views about the basic nature of human behaviour (hence the need to discuss the role of fundamental epistemologies and ontologies as they pertain to theories in general and also the need to state the author's own epistemological and ontological leanings in chapter 2). Of note here and reason for inclusion of this quote is to highlight the blurred boundaries which are often encountered in the social science literature. Is a meta-theory a paradigm? Are the two concepts to be used interchangeably? Perhaps a more accurate rendition of what a meta-theory is as conceptualised for this study is one that "gives the big picture or may be described as the umbrella under which several theories are classified together" (Aldridge et al., 1992, p.638). These authors equate meta-theory to paradigm⁷¹ and mention four such paradigms which effect educational psychology in particular; the organismic paradigm in which person is considered paramount and includes notables such as Piaget, Janet and Gesell; the mechanistic paradigm; in which process as mechanism is paramount and includes among others Galton, Hall, Watson, Skinner and Thorndike; the dialectical paradigm; in which individual-environmental interaction is considered as propeller of conflict and change and is exemplified by, among others, Vygotsky, Luria, Hegel and Marx and lastly the contextual paradigm in which pragmatic concerns are voiced in a milieu of eclecticism and methodological pluralism and include among other pragmatics Peirce, James, Dewey and G. Mead.

This strict demarcation of people into categories, is of course, almost a past-time for philosophers! When viewed in this manner it becomes increasingly difficult to link certain theories with certain proponents of the broader movement, and often the case may be that, as is clearly illustrated within dynamic assessment, models and presuppositions are encased within dialectical, organismic or contextual frameworks only to be delivered using a mechanistic world view (Aldridge et al., 1992). What can one conclude from this? Possibly that the mechanistic paradigm is considered more worthy of scientific status as regarded within the natural science model of conducting good science. What message is this sending out to practitioners and researchers? Possibly that it is worthwhile (perhaps even noble) to theorise and understand phenomena within a progressive framework⁷² but that the method used as ultimate arbiter of the theory or model's veracity is one of mainstream understanding of what it means to conduct science. Hence the need for statistical manipulations, robust measures and empirical verifiability. What happened to the ensconced framework within which the theory was originally embedded? Not to worry, at least it looked good!⁷³ Swanson (1988) refers to this pluralist scenario in a critical way in which incommensurable orientations are mixed without due consideration for the, at times, contradictory claims underlying these various affinities in an effort to offset the inherent weaknesses evident in any one theory.

Meta-theory may also refer to the second-order approach to the study of theories developed within dynamic assessment in this instance (Mautner, 2000; Zhao, 1991). It is a view from above, so to speak, a higher order appraisal of theories aiding in the direction and guidance of the construction of empirical theories, also referred to as orienting strategies (Berger & Zelditch, 1993a).⁷⁴ Dynamic assessment theories and those encompassing models or precursor models (Mouton & Marais, 1988) will be assessed. Swanson (1988), in attempting to move towards a metatheory of learning disabilities, states the need for meta-theories as they provide the framework necessary for continued research and offer services such as the understanding of events, provision of an organising framework (Berger, Wagner & Zelditch, 1989) and revelation of the complexity of performance within seemingly simple tasks. Viewing the meta-theoretical landscape from a view above, below and around the theories and models that make up such a vista, allows one the luxury of a fuller scope of what is going on within the field, but simultaneously confounds such a straightforward attempt to order the theories into some sort of workable framework. Weinstein and Weinstein (1992) detail types or ways of conducting meta-theorising and differ in terms of their reflective practices:

- meta-theory seeks to achieve a deeper understanding of theories;⁷⁵

⁷¹ Used interchangeably with the term research programme (Berger & Zelditch, 1993a; Turner, 1993).

⁷² Social constructionist, feminist and other such critical views.

⁷³ The author states this with some jest and tongue-in-cheek, but sadly this is an all too common phenomena.

⁷⁴ Some philosophers have used the phrase "grand theory" when obliquely referring to what at times seems to be a meta-theory of sorts but do not advocate it as such or much less develop some sort of formal or systematic framework for it.

⁷⁵ Not much of an empirical approach one might add; contrast this to the meta-theoretical framework or systematology offered by Madsen as well as the use of his hypothesis quotient (HQ) in determining the numericised empirical value of theories.



- meta-theorising serves as prelude to new theory construction
- meta-theorising strives to create perspectives that overarch theories seeking novel theoretical integrations
- contrasting agendas for meta-theorising which include meta-theories. These seek to prescribe the prerequisites for theory generation and meta-theorising that seeks to study existing theories (Swanson, 1989; Turner, 1986 in Weinstein & Weinstein, 1992)

In sharp contrast to the authors' predilections in terms of philosophical leanings towards certain issues as elaborated upon in chapter 2, yet in keeping with progressive trends within the reality of how science is conducted (as opposed to any prescriptives), Weinstein and Weinstein (1992) uphold the notion that meta-theorising cannot be undertaken by those remaining firm to foundationalist views on science because the undertaking as such "elucidates the structure of difference, not the unity of differences" (p.141). Through the clarification of suppositions inherent within any theory, meta-theory aims to dismantle such (latent or manifest) alliances in order to bridge or separate those theories that are in need of such restructuring. The manner of such restructuring is, of course, open to debate. Whether or not such an endeavour is an "anti-dote" to abolitionist theorising can likewise be debated. There is no tension here between a positivist type stance in studying social phenomena and doing so in various means, some of which are construed as post-modern, and this is due to the affirmation of the notion that objectivity can be enclosed in a subjectivist psycho-social realm.⁷⁶

Berger, Wagner and Zelditch (1989) maintain that theoretical work starts with either implicit or explicit adoption of a metatheoretical position which in turn is re-negotiated depending on the verifiability of the theories subsumed within it and in this sense it is a reflexive process. In their view, the growth of theoretical knowledge and not merely the relation of theory to observation is intertwined with its predecessors, theoretical alternatives available and competitor theories. Five relations between theories within any one theoretical research programme (or paradigm⁷⁷) can be expanded upon and include:

- elaboration - which entails the growth, development and increase in scope, precision and empirical adequacy of a theory. A theory is more elaborate if it is more comprehensive or possesses greater analytical power or has a firmer empirical grounding but still shares the same basic concepts and principles of the theory it is trying to replace (Berger & Zelditch, 1993). Elaboration consists of three main goals; theoretical goals in which the explanatory domain is larger than the domain originally covered by the initial theory, analytical goals in which the theoretical structure and model is formalised and empirical goals in which observational corroboration and empirical consequences are increased (Berger & Zelditch, 1993). Theory growth occurs and results in the growth of its "assumption and prediction sets" (Jasso, 1993) but presumably due to the criterion of parsimony (discussed above) a smaller assumption set and larger prediction set would be advisable. However, a monotonic scale is evidenced for assumptions in which early theories are characterised by numerous and ever-increasing sets of assumptions as observational data is encompassed within any given theory, only to be trimmed and pruned later on as certain postulates become redundant or simply prove to be incorrect. Thus fewer postulates evidence a more mature theory. On the other hand, an increase in predictive power is an ever-increasing phenomena and thus theory growth regarding postulates and predictions operate in different manners (Jasso, 1993)
- proliferation - in which a theory is deployed in novel terrain and encompasses new concepts and assertions. In other words the range of explanatory power increases beyond the range originally explicated by the initial theory. The initial theory's concepts and principles might be carried over into the new theoretical explanation and modified to suit the new theory (Berger & Zelditch, 1993)
- competition - in which a theory is able to lay claim to territory already claimed by another theory and to subsequently explain the same phenomena in new ways. Competitive theories offer explanations about the same explanatory phenomena but differ widely in their explanatory mechanisms and their chosen concepts and theoretical principles. Faced with contradictory predictions, competitive theories serve theory growth in a manner in which opposing plausible explanations might well co-exist in time yet be reformulated at a latter stage (Berger & Zelditch, 1993). Barendregt (2003) is at pains to delineate competitor theories which seek to eliminate other theories with co-explanatory theories which do not seek to eliminate but to merely offer alternative renditions of the occurrence of phenomena. For instance, as with the case discussed in chapter 2 concerning bridge principles and intertheoretic reduction, it is maintained that there is a difference between intralevel and interlevel theory explanation. Intralevel theories can either evidence theory reduction in which bridge principles are in fact accommodated and theory replacement in which one theory is able to explain away facts and concepts at its own level taking into account the explanation yielded by another theory. Dawson (1995) discusses competitive theories by noting five variations on this theme; intertheoretical entailment in which one theory is logically entailed by another; intertheoretical inconsistency in which one theory logically entails negation of the other;

⁷⁶ Positivist objectivity maintains that objectivity can be achieved; postmodern thought says otherwise, nothing is what it seems to be, and has to be contextualised before it can even start to resemble something coherent. A middle-range stance is preferred, perhaps close to the views espoused by critical realism (it's developer and most vocal supporter being Roy Bhaskar) which places more emphasis on ontological properties as opposed to epistemological properties in the understanding of how science is conducted (Maree, 2003; Outhwaite, 1987).

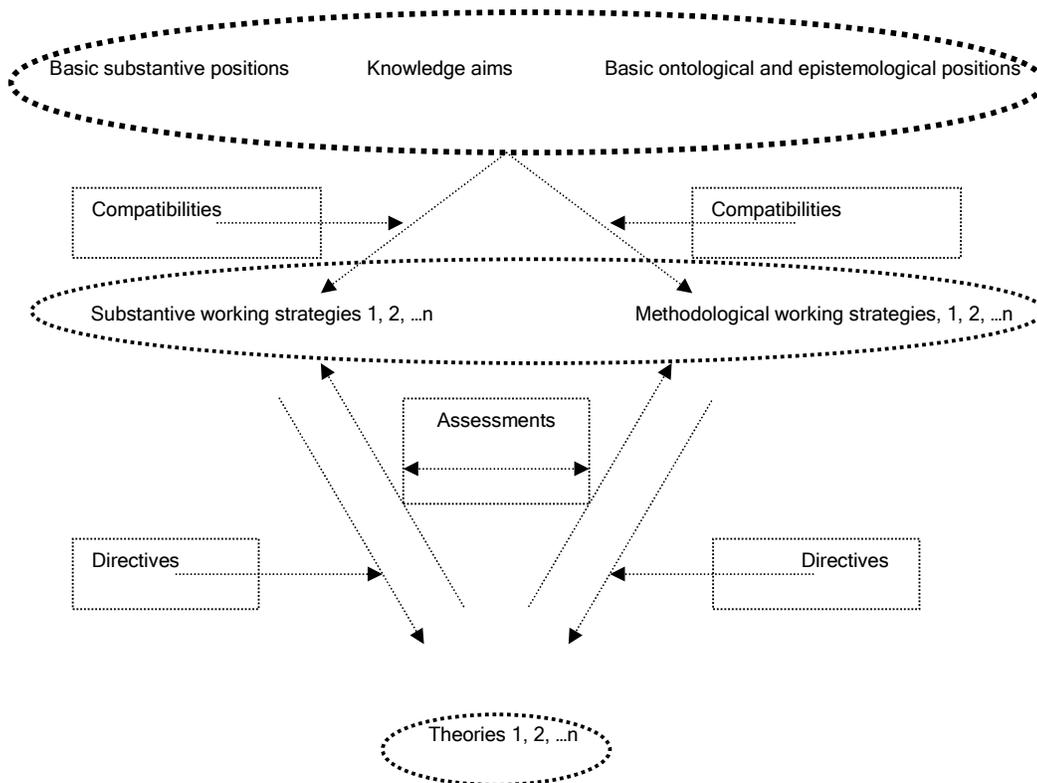
⁷⁷ Typically Kuhn's terminology yet contested as a reliable term which is considered by some as too vague (Stegmüller, 1976).

intertheoretical complementarity in which the negation of the one theory implies the other; independence arguments in which proofs are warranted if no logical relations are said to exist between the two and lastly, conjunct deletion which entails the deletion of propositions in the one theory and in so doing transforming it into the other theory yet simultaneously maintaining the empirical consequences. However, as it at times the case with psychology, interlevel theories operate at different levels of explanation and can both be considered as explaining correctly the same phenomena albeit at different levels, hence they are not competitive theories

- variation - involves the construction of closely related theories which although similar on many accounts, puts forth different mechanisms of explanations on certain issues. Theories are variants of one another if similar concepts emanate from shared concepts and if they address similar explanatory problems. Although very similar in the domains about which they theorise, variant theories often employ different explanatory mechanisms in describing similar processes. It often occurs that variants will replace initial theories or integration results as the one theory is subsumed by the variant theory (Berger & Zelditch, 1993)
- integration - in which at least two different theories or theoretical arguments are able to combine into one coherent account the details of partially overlapping areas of concern. Given two theories and their respective explanations concerning phenomena, a third theory might well integrate both explanatory mechanisms into a third unique yet more powerful explanation. Three types of integration serves theory growth and concern the aforementioned relations; integration of variants in which conditionalisation is utilised to specify the conditions under which each variant theory's process operates; integration of proliferants, in which the different interrelations of the different theories are consolidated; and integration of competitors in which two competing theories are combined into a novel third theory harbouring concepts not found in either one of the original two theories (Berger & Zelditch, 1993). The authors also make a case for the integration of independents in which two competing theories neither share nor make use of similar explanatory mechanisms and principles. Integrations represent major leaps in theory growth, however losses too are incurred when concepts from initial theories are left out during integration

Berger and Zelditch (1993) illustrate their conceptualisation of elements within an orienting strategy (meta-theory) and is reproduced here in figure 40 below. Notice the similarities this shares (especially the feedback process between the third and second levels) with the overarching taxonomic systematology framework devised by Madsen below.⁷⁸

Figure 40 Elements in an orienting strategy (Berger & Zelditch, 1993, p. 18)



⁷⁸ Once again, no mention is made of K.B Madsen's work in meta-theory development which does strike the author as odd, especially since this framework had been developed in the late 1950's.



3.7 Summary

Dynamic assessment is to be placed within an overarching meta-theoretical framework in order to appreciate the path it needs to follow if it is to remain a fruitful field within the broader discipline of psychology. However, before meta-theory is critically evaluated, notions of theories, models and the progress of science needs to be viewed. Psychology, as with any formal discipline has arisen in contexts which have ineluctably formed much of its core armament and the deployment of such armament has had profound effects on the manner of its methodology and its subject area. Science has a long informal history and spans time since Greek speculation through to more formally recognised flourishes emanating during the Enlightenment. The history of natural science is very much a history concerned with explanation into the why and how of nature. Scientific explanatory mechanisms can be largely grouped according to ontological and epistemological concerns and include the deductive-nomological, causal and pragmatic approaches. Law-like models of explanation were sought in investigations allowing for a rendering of nature that was clearly comprehensive. Such core approaches to the understanding of reality finds its way into subsequent psychological areas of concern. Causal explanatory models became the *modus operandi* of many psychophysical experimentation which employed robust measurement techniques. Through such mechanisms psychology was understood to be causal, law-like and knowable where truth could be accounted for. Scientific theories blossomed forth from reductionist, realist and instrumentalist accounts of nature which likewise found conduits pliable enough in psychological research.

Social sciences, in an attempt to emulate their natural science counterparts, strove to become relevant, timely and progressive resulting in the many branching avenues pursued by various sub-disciplines within the ever-burgeoning field. Early on, the disparate areas of interest were to spread off into directions necessitating different techniques of meaning extraction and understanding, thus dividing the field into “hard” and “soft” areas. The continuing crisis of identity pervaded many research programmes resulting in a backlash of criticism of methodology purportedly narrow and close-minded. The “science” in social science was being questioned. The necessary hierarchy of phenomena to be studied and those deemed fit to be studied was and still is a contentious debate. Clear dichotomies polarising niche areas were an unfortunate consequence of the need to redress essential areas of concern. Tension from atop this false hierarchy permeated down to the lowest rungs resulting in the discipline’s ill-defined status as science. Objective observations, hypotheses, precision, verification of findings, measurability and prediction (Barker, 2001) were the benchmarks of a progressive science and fields which did not avail themselves of such techniques were considered as non-science. In this manner, psychology has never really made for itself a unique development and progressive path to follow. Dynamic assessment, it is argued, can in fact do so as it consists of a number of elements in its favour as a new sub-field within psychology. Historical contexts, philosophical preoccupations and critical notions of what it meant to practice psychology were developed alongside the mainstream culture of psychology.

The realm of pseudoscience cannot be equated with the aforementioned non-science as psychology might well develop in a non-scientific manner, or rather it should be stated in a non-natural science manner. The role of theory and model too have been uneasily displaced from one context to another and this is no more clearly seen than in the rationalisation utilised within experimental designs and research hypotheses in many psychological fields but most specifically within psychological assessment. The very elements of theory, concept and terminology development are paralleled along natural science concerns. From psychophysical research to areas of environmental concern, psychology has travelled a varied path in trying to obtain a unique progressive identity which it has hitherto not been very successful in attaining. A tentative remedy for this situation can be seen in the works of theoretical psychologists which can aid and guide the development of the discipline along a new trajectory but is a field not strongly supported by the majority of practitioners. Explanatory mechanisms are still by and large natural science oriented and this state of affairs is reflected in the literature where a disproportionate number of treatises argue the debate of philosophy of science as it pertains to natural science with a resounding emptiness in the social science disciplines. An area rife with contention within psychological theorising is the question of theory appraisal which as with the above mentioned aspects takes much of its argumentation from natural science theory appraisal. Hardly anywhere in the literature is psychological theory appraisal discussed without due recognition of natural science theories. Has psychology not cast itself a role in singular status discipline without need to continually go back and compare itself to natural science models? Philosophers of natural science are routinely brought in to discuss the pertinence of theory appraisal and theory discussion in general within the social science realms. Before meta-theory can develop as a tentative framework, theory itself needs to be discussed. One can see the social sciences continued identification with natural science and as such a meta-theory will have to take into account this relationship as past theories have been built with a natural science framework in mind. Meta-theory seeks to achieve greater understanding of extant theory, serves as an introduction to new theory construction (without the need to utilise natural science models), to create areas of mutual integrative concern for various schools within psychology, specifically assessment and to set agendas for further theorising. Such a meta-theory will now be looked at.

3.8 K.B. Madsen

K.B. Madsen (d. 2004) was a professor at the Laboratory of General Psychology at the Royal Danish School of Educational Studies (now amalgamated with other institutions and as such no longer exists as an independent unit) in Copenhagen, Denmark. A school psychologist, having taught and lectured he was the author of numerous texts in psychology written and

translated into Norwegian, Swedish and Danish⁷⁹ (Madsen, 1971; Royce, 1975). Thankfully some of his work was written in English. That Madsen's model of meta-theory was chosen was due primarily to the scarcity of *psychological* models of meta-theory in the literature and the close affiliation that his model evidenced with the naïve views originally conceived by the author (as noted above, theoretical psychology itself as sub-discipline within psychology is not at present universally formally accepted as being mainstream). It is once again of interest to note that Madsen among others was influential in establishing meta-theory within theoretical psychology as a discipline within the larger subject matter of psychology and that this effort was largely but not exclusively backed by researchers from the Netherlands and Scandinavian countries.⁸⁰ In this regard, dynamic assessment as an assessment model within intelligence research may be geographically more aligned with just such a meta-theoretical approach (since many researchers emanate from Germany and the Netherlands). As has been noted, scientific traditions emanating from specific geographic locations mould, to a certain extent, researchers' conceptions of their field of study (Li & Kunzmann, 2004). Learning potential as understood by the writings of Vygotsky and Binet both espousing intellectual growth and modifiability of the individual (Sternberg & Jarvin, 2003) with a move away from static assessment and meta-theory seeking to bridge both disparate and similar theories can be viewed as movements away from the traditional, conventional and mainstream focus that psychology has exemplified throughout its history with emphasis being placed more on the physiological aspects of the discipline. Critics of both dynamic assessment and meta-theory development alike caution researchers embarking on excursions taken down either path.⁸¹ Criticisms of dynamic assessment will be discussed at length in chapter 4.

Madsen concentrated on psychology as a discipline and originally closely studied over 20 theories of motivation within psychology utilising his "systematology" taxonomical approach (1968; 1973).⁸² He conducted a comparative study of theories of motivation, culminating in his "Theories of motivation". Madsen regarded his treatise as an application of the philosophy of science specifically the meta-science of psychology. He later added over 20 more theories of motivation in a revised book entitled "Modern theories of motivation" (1974) yet he did not relegate his framework to theories of motivation only as can be evidenced in this historical rendition of psychology as discipline as viewed from a meta-theoretical perspective (Madsen, 1988). One possible reason as to why Madsen's work is not more often cited within the current literature may be that much of his work was not written in English and possibly that he conducted much of his research from the 1960's till the late 1980's. Madsen's systematology, a framework he employs throughout his analysis of various schools of psychology, is a theory-focused framework and fits well within his purported structure of theory analysis within psychology. However, this particular study does not concern itself with a broad range of theories within psychology, but a range of models within dynamic assessment and intelligence. The framework as formed and set-out by Madsen then will have to be attenuated to suit the purposes of this study. Before attenuating his framework, however, it will need to be presented in its original and unadulterated Madsenian form.

The youthfulness of theoretical psychology as a (contentious) scientific enterprise might also be a contributing factor towards this lack of interest.⁸³ This study could possibly serve another subsidiary function - the resuscitation of Madsen's work confined not only to North European psychology but to a wider Western audience, more specifically a South African audience. Madsen (1985, 1987b, 1992) in his systematology relies upon the philosophical works of Kuhn (1962)⁸⁴ in his development of a meta-theory within psychology (Brandt, 1984) and also acknowledges the influential works of other notable philosophers of science among them Lakatos, Hanson and Popper as well as the Swedish meta-scientists Radnitzky and Tornebohm (Madsen, 1971; 1976; 1984a; 1985; 1987a; 1987b; 1988). However, it was the work of Sigmund Koch who inspired Madsen to a large extent in the rendering of his systematology approach (Madsen, 1987a; 1987b). Koch's theoretical studies resulting in the edited six

⁷⁹ Intelligence assessment within the Nordic countries (Sweden, Denmark, Finland and Norway) has been influenced by European and American psychometry where measurement and empirical research have received much attention (Carlstedt, Gustafsson & Hautamäki, 2004).

⁸⁰ In order to avoid confusion, Scandinavia is, strictly speaking, the combination of Norway and Sweden but culturally this extends to Denmark, Finland and Iceland with the Netherlands referred to as Holland. Also of note is the long history of content analysis in Scandinavian countries which is a technique utilised in appendix 2 of this study (Rosengren, 1981).

⁸¹ No-where in the literature so far surveyed has it become evident that a link has been established between dynamic assessment and meta-theory in terms of basic philosophies underlying either academic pursuit. Appendix 2 details the formal content analysis of completed questionnaires that were returned to the author in which question eight asked for opinions on the matter of meta-theory development for dynamic assessment.

⁸² Fourth edition. The first edition appeared in 1959 (Madsen, 1971).

⁸³ It is nevertheless curious that treatises such as Gholsen, Shadish, Neimeyer and Houts (1989) fail to mention Madsen. This edited volume can in many ways be construed as a work which sought to place the psychology of science on the map so to speak and nowhere is Madsen cited as a contributor in helping to establish psychology of science as part of the data level of his metascience. Royce who previously cited Madsen in other works and vice versa is cited in the edited volume however.

⁸⁴ The author has expressed her views regarding the utility of Madsen's framework, a framework to be refined for use within this study. Madsen's penchant for the works of Kuhn might seem to be at odds with the original philosophical leanings discussed in chapter two in which it was asserted that personal preferences sway towards realist interpretations of reality and the search for scientific "truth" which seem to stand in stark contradiction to the relativist views propounded by Kuhn (Scruton, 2004). However, Madsen's choice is made. Of course it is an inescapable fact that many contradictory views espoused by many eminent philosophers do nevertheless yield such insights that are difficult to ignore and although the author tends to lean in one direction as opposed to another, this does not imply that many useful insights emanating from these opposing views are simply ignored. Much of what Kuhn, Hanson and Feyerband have to offer are so insightful that is at times extremely difficult to offer counter-arguments in the author's favour. Having said that, the preferred philosophical positions on many issues are traceable to Hempel, Oppenheim and E. Nagel (although as with the aforementioned statement there are a few issues with which the author takes exception and can legitimately employ arguments from the opposing camps).

volumed "Psychology: a study of a science" (1959) was considered by Madsen a major endeavour in theoretical psychology (1985; 1990). Unbeknownst to Madsen at the time of writing his classic on motivation in 1962, the word systematology had already been used as far back as 1782 by a German philosopher and later in 1929 by an Austrian philosopher (1987b). Madsen's overview of theories of motivation was eventually published in Wolman's 1973 book "Handbook of general psychology".

Although Madsen himself does not mention dynamic assessment nor learning potential per se in his writings⁸⁵ his ideas and frameworks concerning models, theories, meta-theories and overarching philosophy of science are malleable in terms of its applicability to this area of assessment. The original notation utilised throughout Madsen's work is reminiscent of the times during which he worked (typical stimulus-response nomenclature is used) and may be considered as too reductionist by current and more phenomenologically inclined scholars. Madsen's terminology and discussion of systematology and other taxonomic schemes is at times puzzling as various models at times overlap in terms of labels, meanings and functions between and within books and articles but the author feels that this might be due to language translation from the original Danish and at times one feels awash in the hierarchical chains that he constructs. For instance in discussing three major issues Madsen will plunge into these separate issues each with an additional four sub-issues and in turn these issues contain their own sub sub-issues till one is so far down the hierarchy that the point of origin has been lost somewhere at the top. As such it is deemed necessary to tabulate his views on certain issues in order to bring a sense of coherence to his ideas. These tables (or organograms) are presented throughout this section. Nevertheless, specific and singular attempts are made to develop for psychology specifically, a meta-theoretical framework from which to view its multitudinous theories and models of which dynamic assessment is used in this particular study.

3.9 Madsen's meta-science and meta-theory

This section will focus exclusively on how Madsen envisioned, conceptualised and eventually realised his own understandings of meta-science and meta-theory within science and specifically within psychology. The choice of Madsen has been outlined above and although innumerable studies conducted within the philosophy of science have been penned, viewpoints concerning the growth, development and progress of theories per se will have a Madsenian flavour and cognisance is taken that his approach and emphasis on many issues within this area are open to debate. Before attention is give to Madsen's definition of science and meta-science his scheme of how science progresses⁸⁶ is discussed. Science is not practised in a vacuum and although the impression is often given that science as an enterprise grinds onwards towards improved explanations of how things are, science is really the accumulation of many individuals pursuing knowledge-gathering functions in various guises for various reasons. At the heart of the scientific enterprise is the individual personality placed within a specific scientific community itself housed within the broader governing general scientific community which is vulnerable to the time and location of its particular setting within history (a relativist tinge is evident here with influences from the meta-science branch of the psychology of science, Feist, 1995). This placement of individuals and scientific practises contextualises scientific practise⁸⁷ and leads to a better understanding of how and why theories and models are conceived as they are. This nesting of hierarchies is typical of Madsen, a technique he so often utilises throughout his writings. Table 8 presents Madsen's scheme of embedded scientific practice.

Table 8 Classification of the three sub-systems within which theories are produced (Madsen, 1987b, 1988)

Systems level	Historical meta-theory	External history of science	Combined external and internal history of science	Internal history of science
Cultures and societies ¹		Cultural history of science and socio-economic history of science		
Scientific community ²			History of the scientific community	
The researcher ³				Biographical history of science

1 = Sociology of science

2 = Social psychology of science

3 = Psychology of science

⁸⁵ As far as this author is aware.

⁸⁶ Progress here connotes the development of the method and does not necessarily indicate that current findings are any "better" than previous findings. The term "progress" merely indicates change through time. A hint of relativism creeps in here.

⁸⁷ Which is why the author felt the need to include her own predilections and leanings on various issues in chapter 2. It is thus in keeping with Madsen's approach of understanding how and why theories develop as they do.

3.9.1 Madsen's definition of science

Before meta-science can be disclosed as a definition, science as a concept needs to be unpacked and as such Madsen (1985; 1987a; 1988; 1990; 1992) delineates the three components or processes contained within the definition, namely;

- Empirical research
- Theoretical thinking
- Philosophical thinking

Radical empiricist philosophers of science have advocated that *empirical* observation via the senses be the key to the practice of science (Smith & Katz, 1996). Among many such positivist empiricists are included Comte and Mill followed in turn by the less restrictive empiricism of a later age exemplified by the neo-positivists such as Russell, Wittgenstein and Carnap. *Theoretical* thinking was included in the arsenal of techniques used to derive the truth by the logical empiricists and as such included aspects such as hypotheses testing and explanatory modelling of data and is evidenced by philosophers such as Hanson, Popper, Bunge, Radnitzky and Törnebohm. *Philosophical* thinking was melded into this mix and was introduced in order to provide a philosophical framework encompassing the empirical and theoretical aspects and such frameworks include the paradigms of Kuhn and the metaphysical research programme of Popper. Madsen (1985; 1987a) includes in this level the philosophy of the world (which includes the ontological world hypotheses and overall meta-model) and the philosophy of science which encompasses epistemological, metatheoretical and methodological issues. In sum, Madsen concludes a definition for science as follows "the social-cultural system of individuals who are engaged in empirical research, theoretical and philosophical thinking" (1985, p.4).

3.9.2 Madsen's definition of meta-science⁸⁸

As with the definition for science, Madsen expostulates his framework for meta-science in similar vein. The classification follows three levels of abstraction namely; the empirical, theoretical and philosophical levels (1985; 1987a). The *empirical* or data level of this science of sciences (meta-empirical level) include several disciplines such as the history of science and a sub-discipline within this area of the history of psychology. Another meta-discipline within this empirical meta-level is the comparative study of theories which formed the basis of Madsen's 1959 original treatise on the comparison of theories of motivation. It important to note at this stage that Madsen's terminology of this comparative study of theories is located within the empirical meta-level and contributes to the overall general meta-science of the discipline being studied, in this instance psychology. To illustrate: the empirical level of a study deals with the data gathering of an experiment for instance and the theoretical levels deal with the hypothetical details concerning the nature of the data and the results that are yielded in this process. The meta-level would be when more than one such theory is combined with others of similar nature and an overall result is obtained.⁸⁹ Imagine this scenario taken one step higher in a hierarchy where the empirical data now becomes a meta-empirical data level consisting of comparisons of theories where the data are already formulated theories. In other words the theory becomes the data. Madsen is careful to delineate exact differences between historical studies and comparative studies (which are by nature historical). Studies of historical theories also make use of other material in addition to the texts used with the purpose of explaining the development of such theories, whereas comparative studies make use of only the theoretical text and in so doing contribute to the meta-theory of that discipline. This comparative study is termed "systematology".

The *theoretical* level of the meta-level of the meta-theory level refers to the varying types of theories of particular disciplines, so instead of dealing with one theory on the ordinary level, the meta-level tackles many theories simultaneously as its theoretical level (in comparison to the one theory dealt with at any one point in time for ordinary level theory). Madsen (1985) cites the theories of Kuhn (1962), Lakatos (1970) and Hanson (1957) as examples of meta-theory level concerns. Being theories of science rather than philosophies of science, states Madsen (1985), results in these theories being more testable as opposed to more prescriptive philosophies of science. Theories of science then belong to the meta-theory level of meta-science. The philosophy of science is housed within the *philosophical* meta-level of science and is derived from the philosophy of knowledge or epistemology. Philosophy of science emerged after the advent of science during the Renaissance as formal discipline and as such it is a newer discipline. This level of science is predicated on prescriptive activities for the deployment of scientific ideals in terms of how science should be practised (1987b). Ways of doing science, methodologies utilised and conclusions reached

⁸⁸ The reader may experience the same confusion often encountered in texts dealing with extra dimensions in space. For instance we know what a four-dimensional object looks like via its "shadow" in three dimensions. But to imagine what the object looks like in four dimensions we have to strain somewhat to conceive of what it might possibly look like. So it is with the meta-level of science, which looks exactly like its normal theory counterpart but has added dimensions to it.

⁸⁹ It is hoped that the reasons behind conducting a meta-analysis on South African studies within dynamic assessment is becoming clear (see Appendix 1). This statistical approach is more in keeping (at least philosophically) with the nature of this study and with Madsen's meta-theoretical approach.

should be governed by rules and norms which are partially established by fundamental philosophies.⁹⁰ The next levels of meta-scientific inquiry deal with similar issues but on a more hypothetical level and is not as prescriptive as the meta-level; prescriptive meta-science hoping to explicate the process necessary for proper knowledge acquisition as opposed to more descriptive meta-science with less emphasis in this area of concern (Houts, 1989; Shadish, 1989). Figure 41 illustrates the framework of ordinary science discourse as commonly understood. This figure is followed by Figure 42 which in turn illustrates the framework of meta-science discourse as understood by Madsen.

Figure 41 Illustration of ordinary science discourse with examples

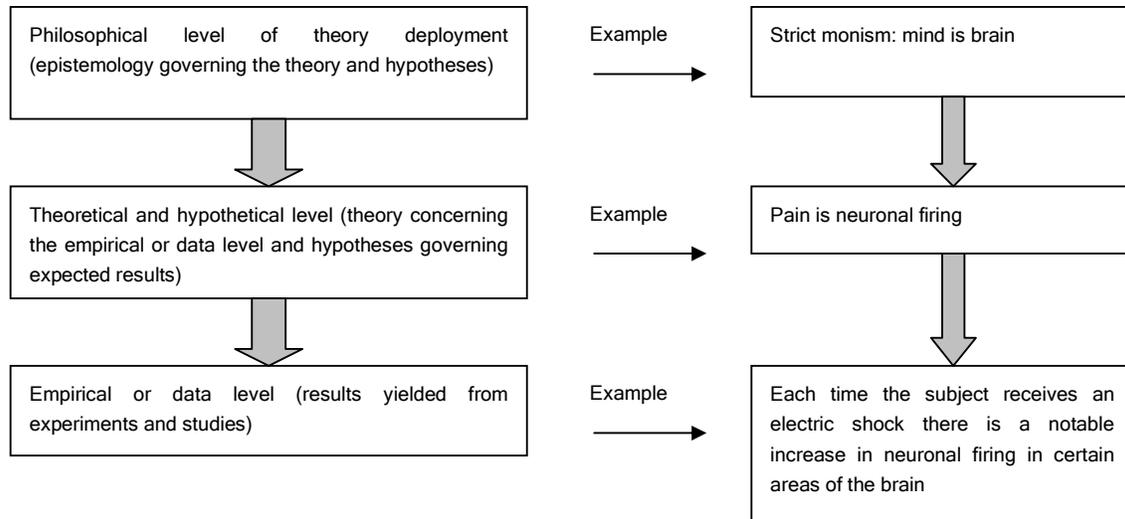
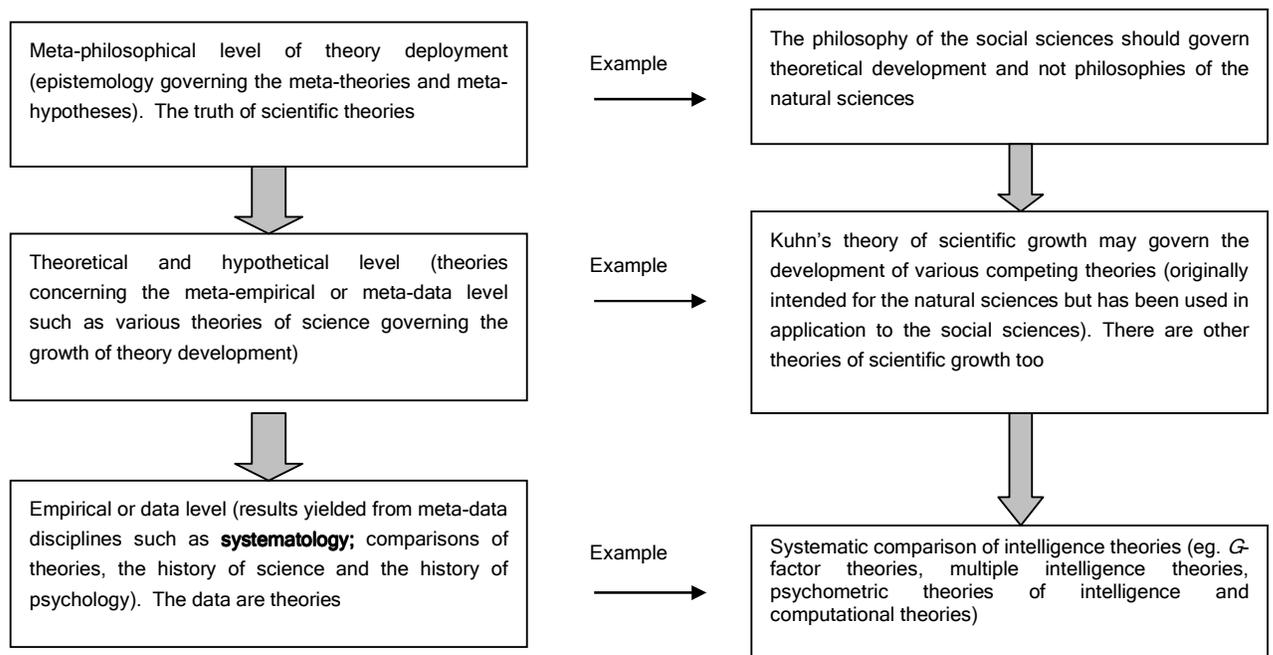


Figure 42 Illustration of meta-science discourse with examples



⁹⁰ It is just such basic epistemological queries which plague dynamic assessment within intelligence research and if proper care is taken to place dynamic assessment's forebears in terms of philosophies perhaps the field would not be quite so saddled with methodological and empirical problems.

3.9.3 Madsen's view on theory

Whilst discussing at length the history of psychology within a meta-scientific perspective, Madsen (1988) views a theory as both

1. a narrowly defined set of hypotheses with associated models and
2. as a broad scheme encapsulating a fully-fledged scientific text including
 - a. meta-theses or a set of deeper philosophical issues
 - b. a set of hypotheses related to the model via explanatory statements
 - c. a set of descriptive data-theses

The latter definition of expanded theorising functions along three axes (or expanded discourse), namely the process yielding the theses, the linguistic category utilised in the theses and purpose (or function) of the theses and is illustrated in table 9.

Table 9 Level of discourse according to classification scheme from Madsen (1988)

Classification	Level of discourse	Meta-theses	Hypotheses or model	Data-theses
The processes		Product of philosophical thinking often divergent in nature	Product of theoretical thinking often convergent in nature	Systematic observations, tests or experiments
Linguistic categories		Philosophical and scientific hypotheses, the former not capable of verifiability or falsifiability but the latter indeed so. Includes epistemological, meta-theoretical and methodological statements about science	Propositions or hypotheses consisting of empirical and theoretical terms. Can refer to unobserved processes therefore constructed descriptions of concepts	Purely descriptive statements
The purpose (function)		To provide and create a basis for understanding a conceptual framework for both the hypothesis and data levels**	To systematise information via explanations (causal and functional) and/or to interpret coherent meanings	To present data, knowledge and information which includes both specific (isolated data) and general information (relations between data)

** Madsen (1988) supports the notion that for any length of time, certain implicit assumptions and meta-models are not directly stated due to large-scale acceptance of these views, until that is, a crisis emerges in which these very implicit assumptions need to be stated and aligned within the larger arena of research (in whatever context) (Lichtenstein, 1980). The fact that dynamic assessment finds itself at just such a juncture in its history is testament to the need for such a revision of basic assumptions which will no longer suffice as implicit but need to be made manifest in order for progress to continue. Kuhnian influence is quite evident in Madsen's conception of the above framework.

3.9.4 Madsen's view on models⁹¹

Employment of analogous explanations of the functioning of human beings within larger contexts was originally the technique utilised in early models of human behaviour (Madsen, 1988). To simplify understandings of how humans functioned it was necessary to ground the comparisons in everyday experiential contexts, thus was borne the system of metaphors widely employed as explanatory mechanism throughout formal psychology's history as well as that of science in general (Bailer-Jones, 2002; Brooks, 2003; Cervone, 2000; Crosson, 1985; Dawkins, 2001; De Mey, 1989; Draaisma, 2000; Jordaan, 1989; Penrose, 2002; Woolley 1933). These explanatory mechanisms were offered to suit the reigning paradigm and adapted as knowledge

⁹¹ This discussion on models is placed here in order that flow of thought is preserved. Models will again be discussed under section 3.5.2 below as it pertains to the hypothetical stratum within Madsen's systematology.

improved about certain phenomena (Daugman, 2001). Before returning to Madsen's view on models within the history of psychology a brief digression into the role of metaphor as foundation for model development and heuristic for consequent theory development is warranted. A variety of metaphors of epigenesis for instance, render scientific explanations of intellectual growth. Metaphors have served formal science in general for over 150 years often resulting in prescient veridical theories, among others the following is a brief list offered by Draaisma (2000):⁹²

- John Newland (1837-1898), English chemist; his analogous piano keyboard metaphor was utilised as heuristic for the grouping of elements into groups of eight (octaves) as each eighth element was a repetition of the first. This system was presented before Mendeleev's compilation of the periodic system of elements
- Elie Metchnikoff (1845-1916), Russian biologist who utilised the analogy of how larvae wrapped themselves around wood shavings and ingested them to aid in the explanation of how the immune response to infection was explained leading to the discovery of the body's immune response to infection (phagocytes consuming bacteria)
- Albert Einstein (1879-1955), German-Swiss theoretical physicist, who, by employing a thought experiment was able to analogise the relativity of light through ideating on what it might be like to hold up a mirror to one's face during a speed-of-light travel on a ray of light
- William Harvey (1578-1637), English physician, who described the heart as a pump and thus developing the theory of blood circulation

Within psychology the role of metaphor too as has had profound consequences for theory development

- Various metaphors permeate psychological theorising (Leary, 1990a) as well as intelligence assessment (Carlson, 2002) and include among others the view of the mind as understood from the view of a machine (although Gergen (1980) is of the opinion that mechanistic metaphors have severely impeded the growth of psychology as a discipline as it splits the organism from its environment and thus concentrates on the mechanics of action versus person), an electronic computer, a homunculus, a mirror or blank slate and so on (Boden, 1988; Daugman, 2001; Groome, 1999; Leary, 1990b; Pribram, 1990); William James' stream of consciousness and Freud's ice-berg analogy of the conscious and unconscious (Madsen, 1988).

These forms of heuristics are divided into two types, namely theoretical and empirical, the former playing the role of organiser, systematiser and bringer of coherence to contradictory empirical findings whilst the latter yields new fruitful avenues of further research. The former notion is often decried as not sufficiently "scientific" within the field of psychology (Feist, 1995; Kukla, 2001) yet has had profound consequences in the realm of the natural sciences. Madsen (1988) refers to pre-scientific cognition as analogous in explication but these types of comparisons are not always considered "scientific" (Draaisma, 2000). Analogies and metaphors are helpful aids in seeking an initial explanatory mechanism within any domain of knowledge acquisition but as Gentner and Jeziorski (1989) caution, analogies like the phenomena they seek to describe involve a certain degree of skill when utilising them. Human brains are better equipped at handling data in pictorial or visual form similar to pattern recognition (Dwyer & Dwyer, 1987; Kline, 2004) and thinking about phenomena in terms of similar phenomena might ease the understanding of what is being explored (Eysenck & Keane, 1993; Mayer, 2002, 2003; Shermer, 2005). There is however a fine-grained difference between the roles of analogy, simile, model and metaphor with metaphor considered by some philosophers (Indurkha 1992 in Maree, 1995) as encompassing an unconventionality when interpreting or understanding a phenomenon. Metaphors can be similarity-based (the comparison view; Montuschi, 2001) or similarity-creating (the interactive view of metaphors; Montuschi, 2001) in which the former are based on already known similarities whilst the latter create the similarities between phenomena not hitherto seen or understood (Maree, 1995). Models, are seen by some as special types of metaphors (Crosson, 1985) which are more developed in terms of bridging explanatory predicates to mathematical or logical theories.

Models can also be classified according to a framework set out by Madsen (1987) and is illustrated in table 10 below.

⁹² Dates from Porter (2003) as well as Millar, Millar, Millar and Millar (1996).

Table 10 Framework for model abstraction and function

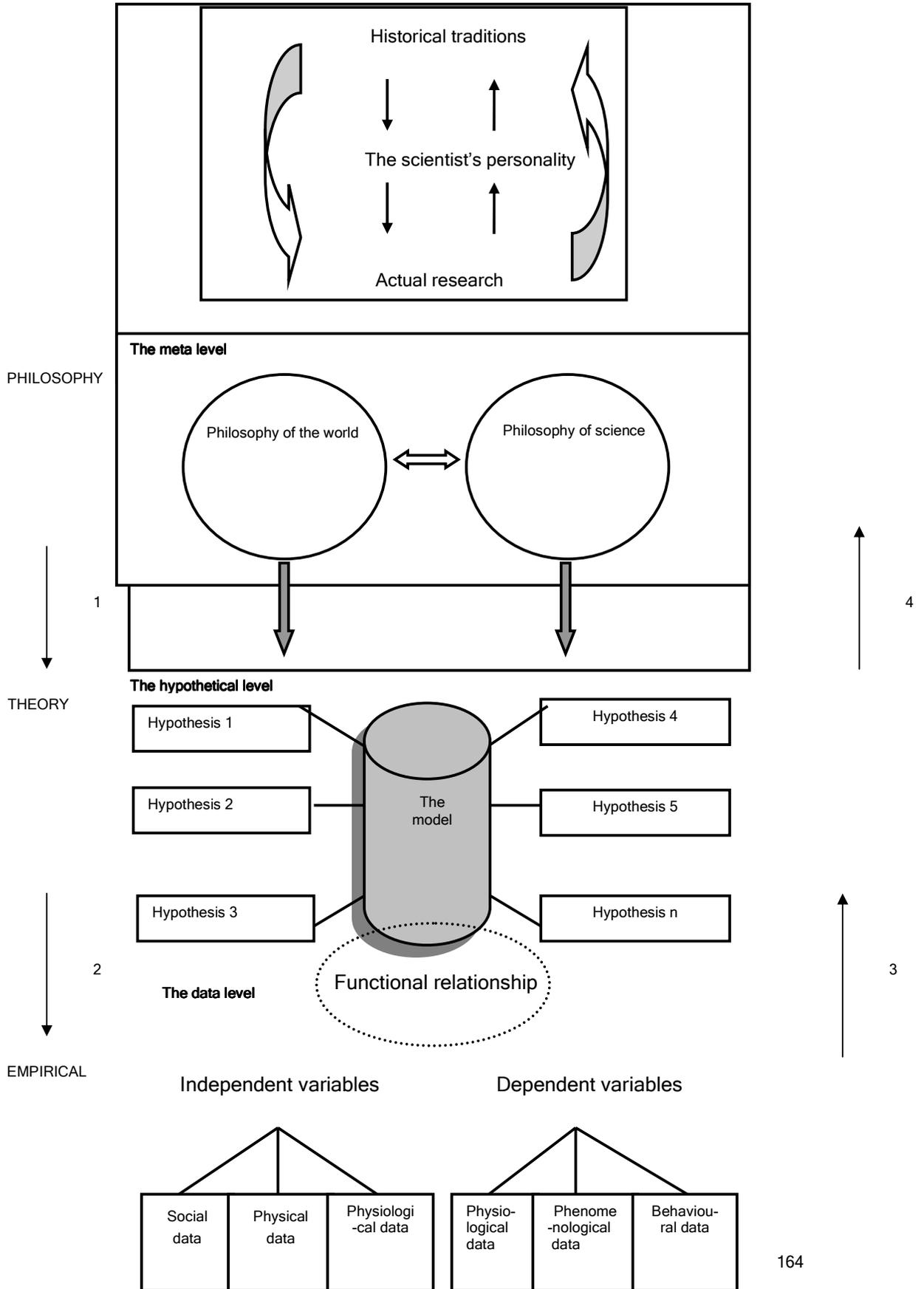
Degree of abstraction	Purpose (function)
Material - three dimensional	Descriptive - schematic descriptive representation
Graphic - two dimensional	Explanatory incorporating hypothetical variables: <ul style="list-style-type: none"> • Constructive models • Physiological models • Mentalistic models
Simulation (computer programmes)	Meta-models (ontological)
Mathematical (symbolic systems)	

Regarding the degree of abstraction Madsen (1988) uses the globe as an example for material three dimensional models and a map as instance of a graphic two dimensional model. He states, moreover, that descriptive-schematic model representations belong to the data level as they are often of the “black-box” type as they contain no hypothetical variables. Explanatory models on the other hand provide instances of hypothetical variables which are classifiable as constructive (containing pure thought hypotheses), physiological (containing surplus physiological meaning) and mentalistic (containing surplus mentalistic meaning). Meta-models, states Madsen (1988), are typically implicit “global models” and provide the background framework for the understanding of the presuppositions inherent within the models. The word systematology has been highlighted above in Figure 42 as this is where this study has positioned itself. The various theories and models within dynamic assessment now become the data in this meta-theoretical exploration of dynamic assessment and intelligence. This systematological approach itself becomes an area of inquiry to which attention shall now be turned.

3.10 Madsen’s overarching framework: Systematology - a comparative meta-theoretical taxonomy

This section will study Madsen’s overarching framework for viewing theory within meta-theory. In essence, three levels form the basis of this framework, namely, the meta-stratum, the hypothetical stratum and the data-stratum. Each level will be discussed in turn. Figure 43 presents Madsen’s overarching scheme of science within society. Madsen (1987b, 1988) devises this classification scheme in which three system levels interact with three types of historical meta-theory. The system levels include cultures and societies, the scientific community and the researcher. Historical meta-theory engirds the external history of science, the combined external and internal history of science and the internal history of science.

Figure 43 Madsen's overarching scheme. From Madsen (1971), p.7; (1973), p.698; (1976), p.301; (1984), p.201; (1985), p.3; (1987a), p167; (1987b), p268; (1988), p.29; (1992), p504.





Systematology is used as a synonym for meta-theory in the writings of Madsen (1968) and becomes the level at which empirical data gathering is conducted. Madsen (1971; 1976; 1984a; 1985; 1987a; 1987b; 1988; 1992) is of the opinion that in order to be classed as empirical meta-data, theories need to encompass three levels of abstraction, namely the;

- i. The meta-stratum-theses such as propositions about the philosophy of the world⁹³ and philosophy of science
- ii. The hypothetical stratum consisting of hypotheses and explanatory models⁹⁴
- iii. The empirical of data stratum which consists of data theses such as general functional relations between dependent and independent variables as well as consisting of propositions

3.10.1 The meta-stratum

As is the case with theory and data, the one influencing the other in order to better inform; so too does this feedback take place within meta-theory and meta-data models (recalling the fact that this meta-model is merely one place higher in the hierarchy of the practice of science). Meta-data feeds back information to the meta-stratum which directs and steers meta-empirical findings. "This interaction between the overall philosophical steering and the rectifying empirical research lies at the heart of scientific growth" (Madsen, 1988, p.30). The theses at this level can be divided into meta-theses about the world (ontological meta-theses) and meta-theses about science (philosophy of science).

Ontological meta-theses

Theories of human behaviour (in all their generalities and scope) often do not make explicit their underlying implicit assumptions governing belief in fundamental issues about the nature of human-kind, a cautionary note taken heed of, hence the need for Chapter 2 so as this author is not similarly accused. Madsen (1988) maintains that as a result of the reigning philosophy of science and the practise of psychology as discipline within a positivistic framework during the early rise of the subject, many thorny issues were simply not dealt with or at the very least were only partially formulated as these issues were not considered scientific enough. These issues consisted of, among others:

- The conception of man⁹⁵

Humans can be divided according to many conceptual classifications, but among those included by Madsen (1988) are the biological, social and humanistic conceptions. These conceptions tie in well with the current debates in the intelligence and dynamic assessment literature and have formed the bases of questions included in the questionnaire in appendix 2. That man is a product of his genetic heritage is without question, the emphasis to be placed on this dimension is however questionable. Is man resultant from environmental and societal impingements? Man can be considered a product of his own unique culture that he builds around him.

- Psycho-physical theory

Madsen (1988) employs a sliding scale to illustrate the various classifications according to which the mind-brain debate can be exemplified. On the extremes of this scale lie the materialist (mind is matter) and spiritualist (mind is spiritual substance) conceptions and placed squarely in the middle is dualism (consisting of both mind and matter). Both substances co-occur and are manifest in the body and its consciousness⁹⁶ and the relation between the two can be divided into interactionism, parallelism (interacting in synchrony) and identity (neutral monism where the two supposedly opposing substances are regarded as the same type of substance). Between the outer limits of materialist and spiritualist views on mind-brain and the middle-of-the-road views lie intermediate views referred to as epiphenomenalism (consciousness is considered an emergent property of the system⁹⁷) and hyperphenomenalism ("animism" in which physical brain processes are considered as a result of the spiritualist substance).

⁹³ Which is why the nature of the questions in the questionnaire sent to 100 dynamic assessment practitioners across the globe contained the specific blend of questions it did, a blend which was construed as "too difficult" in some instances to be answered. See Appendix 2 for further information pertaining to this endeavour.

⁹⁴ Where dynamic assessment may often be found as these models cannot not always be considered fully-fledged theories.

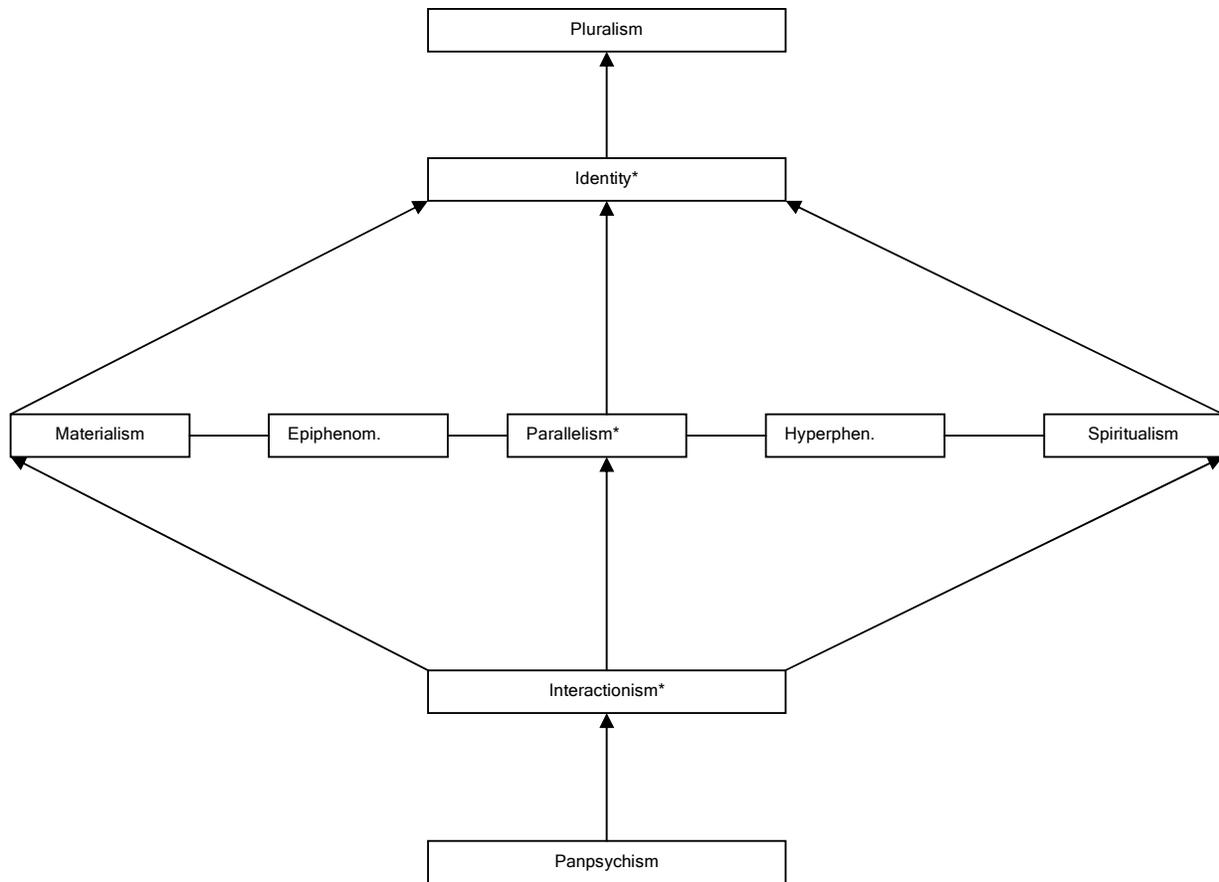
⁹⁵ Madsen utilises terms such as mankind as opposed to humankind. The author has decided to leave it as such as it becomes too cumbersome to continuously make the changes required to suit more modern tastes. Expediency dictates this choice.

⁹⁶ Chapter 2 briefly outlines the debate within this particular area of concern.

⁹⁷ Madsen (1988) refers to this emergentist view as "bi-phenomena".

Two other conceptions which are mentioned but are not included in Madsen's sliding scale are panpsychism (in which animate and inanimate objects are considered conscious) and pluralism (encapsulating systems-theoretical notions and ideas⁹⁸). Madsen's psycho-physical theory is illustrated in figure 44.

Figure 44 Madsen's (1988) conception of psycho-physical theory within the ontological meta-theses as part of the meta-stratum



*Dualism

⁹⁸ Such as larger systems subsuming smaller systems; for instance, the eco-system includes human beings which are made up of smaller systems such as organs, themselves made up of cells which are in turn made up of DNA which are made from proteins and so on (Capra, 1997). Pertinent to this issue of consciousness is the conceptualisation of consciousness as neural mechanism in which a synchrony of neural firing evidences a pattern of firing sequences which is what constitutes the basis of memory. It is not the firing per se which is memory or consciousness but the order, pattern and interconnectedness in and with which these firing sequences take place "a brain-wide synchrony" (Zimmer, 2005, p.288) (Calvin, 1997; Capra, 1997; Greenfield, 1997, 2001). This is an idea stretching back to Donald Hebb's (1949) cell-assembly hypothesis (Calvin, 1997) which, although insightful, did not predict the true trajectory of the course of events (Corballis, 1985) but did manage to synthesise both localist and holist tendencies in his work (Harnish, 2002). It did give to the intelligence debate views on intelligence A (innate and unknowable) and B (phenotype) (Elliott, 1988) (see chapter 2) .



- Human freedom of action

Whether human actions are free or causally determined yields two opposing arguments which seek to settle the dispute, that of determinism and indeterminism. Determinism includes mechanical determinism in which a single cause results directly in an action and dialectical or dynamic determinism in which actions are a result of varying interactions between causes. Indeterminism includes probabilism (in which actions are random but can be predicted with the use of stochastic rules and classical indeterminism in which actions are determined by free will and are thus unpredictable).

Philosophical meta-theses

As opposed to ontological meta-theses, philosophical meta-theses usually receive more explicit coverage within psychology theories but this is not to say that all such philosophical issues are likewise explicitly stated. Philosophical meta-theses can be divided into the following theses:

- Epistemological theses

Epistemological theses are of the more abstract type and are usually the ones which are implicitly assumed or implied within theories. Regarding the philosophy of science and its relation to psychology, two issues within epistemology have been investigated, namely, the origin of cognition and the relation of cognition to reality. Two major theories concerning the origin of cognition include empiricism (observation is paramount) and rationalism (thinking is paramount). Combinations of these abound, the most popular being logical empiricism, a product of the analytical philosophers' Vienna Circle (who held in favour both theory and observation; Outhwaite, 1983) and Karl Popper's critical rationalism (Machamer, 2002; Van Lil, 1987). Oriental philosophy has bequeathed a third dominant epistemology, namely, intuitionism⁹⁹ (described by Madsen as more holistic yet ascribing right hemispheric dominance to this type of thinking as opposed to left hemisphere dominated rational thinking). The relation of cognition to reality attends to the issue of the existence of reality and its independence from human knowledge about it. Three main views include (i) realism - reality does exist independently of our knowledge of it and through naïve realism man can achieve the goal of objective truth and the more tempered consideration of critical realism which advocates knowledge of reality but knowledge which is achieved through the faculties of the brain and sense organs (ii) idealism - one such independent reality exists outside the cognising agent. Realism and idealism are often correlated with the extreme views on the mind-brain debate, namely realism with materialism and spiritualism with idealism although such a relation is not logically necessary (iii) instrumentalism - which is poised between realism and idealism. It is a view which acknowledges an objective independent reality but that the only "truth" available to us is pragmatic truth, i.e. truth that is useful for our actions and in a manner very similar to Popper's critical realism.¹⁰⁰

- Meta-theoretical theses

Meta-theoretical theses deal with the hypothesis stratum and concentrate on scientific cognition, dealing with the function and purpose of theories. Madsen serves an eclectic mix of the "ideals of science" by synthesising the different meta-theory ideas of Habermas (naturalistic, hermeneutical and critical sciences), Maslow (controlling and taoistic sciences) and Allport (nomothetic and idiographic theories). The result is a coverage of all three scientists' thoughts concerning the ideals of science and is considered below:

a) Nomothetic ideal

These ideals are based on natural science laws (nomos referring to law) of nature, which, when translated to meta-theory, become highly probable hypothetical contenders for how things work and was the ideal adopted in the early foundational stage of psychology as scientific discipline. Predictability and explanation are prime considerations in this ideal.

⁹⁹ These epistemological categories are aligned to those advocated by Royce (1964, 1973) as well as Royce, Coward, Egan, Kessel and Mos (1978) as utilised in chapter 2 as second axis when describing the author's philosophical leanings.

¹⁰⁰ Much like the world we inhabit in which it is hypothesised that multitudinous dimensions pervade our space-time reality but as we are simply not attuned to such dimensions by virtue of our present evolutionary make-up these dimensions are not available to us in the sense of our observing them in any measurable way. That they cannot be hypothesised about (such as string theory for instance) does not limit us in trying to think about them at least. Our four dimensional self-hood is constrained by its own existence, very much like the brain trying to understand itself; it is an encapsulated system. A retort to this would be for another brain to investigate my brain, but this system too is self constrained by virtue of the fact that the "brain observing the brain" system occurs within a larger yet contained system. And so on down the spiral. Can reality ever be known in such constrained systems? If this argument is taken to its logical conclusion then the answer has to be "no". However that this type of argument is valid is a question that needs to be asked. A slippery slope argument if ever there was one.



b) Hermeneutic ideal

The hermeneutic ideal seeks to interpret and understand texts (be they biblical, legal, philological or psychoanalytical in nature) and is mainly concerned with intentional relations whereas the nomothetic ideal strives for causal relations which are sought for as an ultimate end-point explanation. Both nomothetic and hermeneutic, although distinctly varied in goal, strive for coherence and a need to add to the total system of information.

c) Idiographic ideal

The idiographic ideal takes its cue from humanities disciplines such as biographical history and seeks to analyse and describe unique and individual case-studies.

- Methodological theses

These are perhaps the most explicitly stated and formulated theses in this framework. Methodological theses are concerned with the choice of methods and the choice of data language to be used. They include the following:

a) Research methods

Wundt's Leipzig laboratory bequeathed to psychology its scientific disciplinary status just as Galton's method bequeathed a test method to the discipline (although as will be evidenced in chapter 4, the experimental nature of the research design within psychology was to change, specifically with the introduction of the treatment group as variable and was to impact on the future development of the discipline in terms of the statistical techniques utilised). Psychology developed in tandem with natural science methodology, namely, the experiment and the degree of variable control manifested in varying degrees ranging from strict experimental designs through to more qualitative designs. The degree of quantification of results in part also determined how robust results were. The test method was regarded as variant of the experimental method.¹⁰¹

b) Data language

Just as close philosophical affiliations predispose researchers to define and research their area of interest in manners differing from their counterparts advocating other philosophical affiliations, so too is this reflected in the data language employed to discuss the research results. Ranging from highly subjective qualitative phenomenological descriptions of results to strict objectivist and empirical descriptions of results, the data language will often mimic the research design. Fundamental philosophical issues regarding mind-brain, consciousness, dynamic vs. static assessment, nature-nurture issues and more besides filter through most often in implicit fashion resulting in assumptions which pervade subsequent models and theories. The meta-stratum level thus far discussed is illustrated in figures 45 (a) and 45 (b) below.

¹⁰¹ Perhaps this was the logical limit of this newly found discipline, an issue of contention today.

Figure 45 (a) Madsen's (1988) meta-stratum (A) Ontological meta-theses

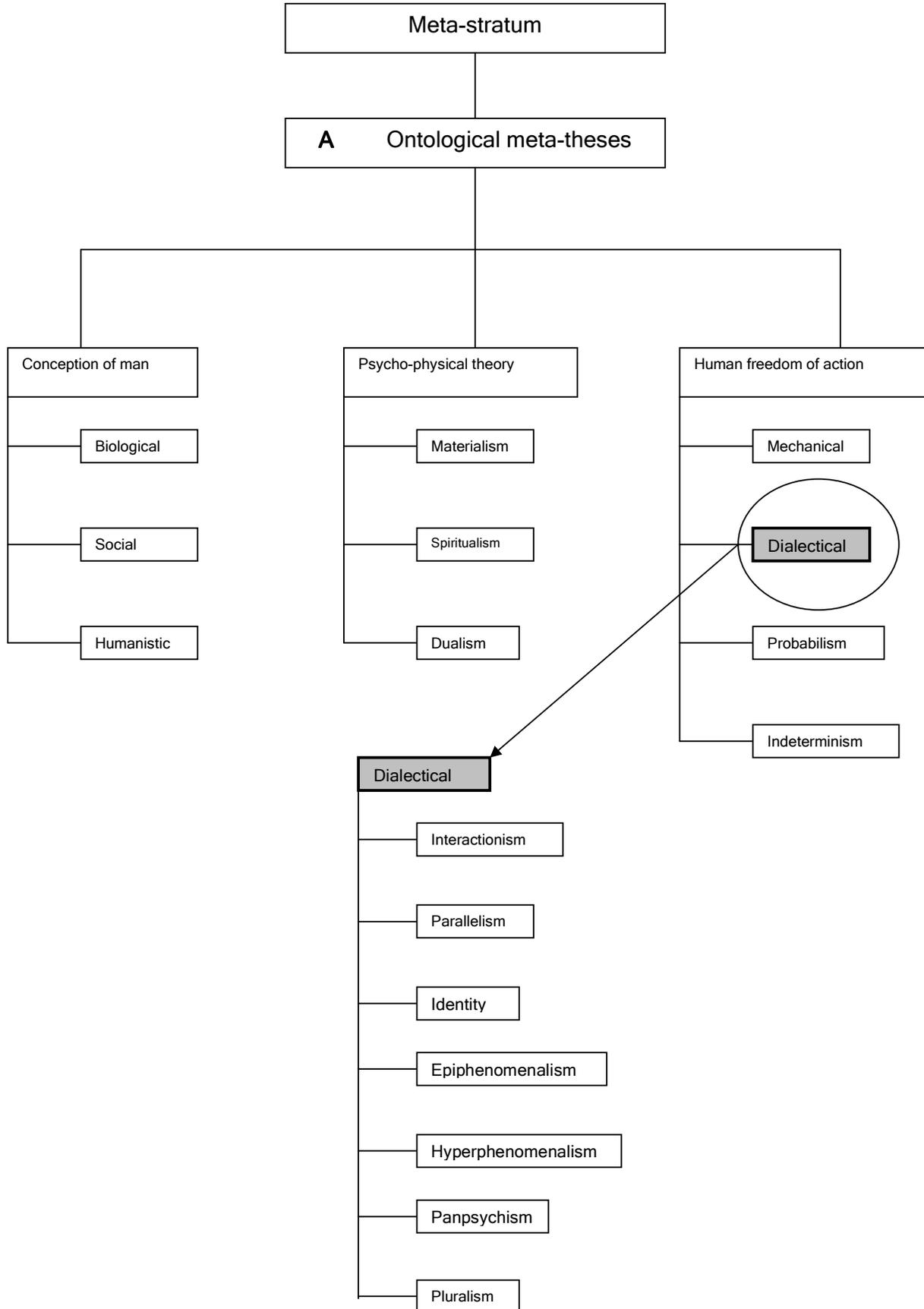
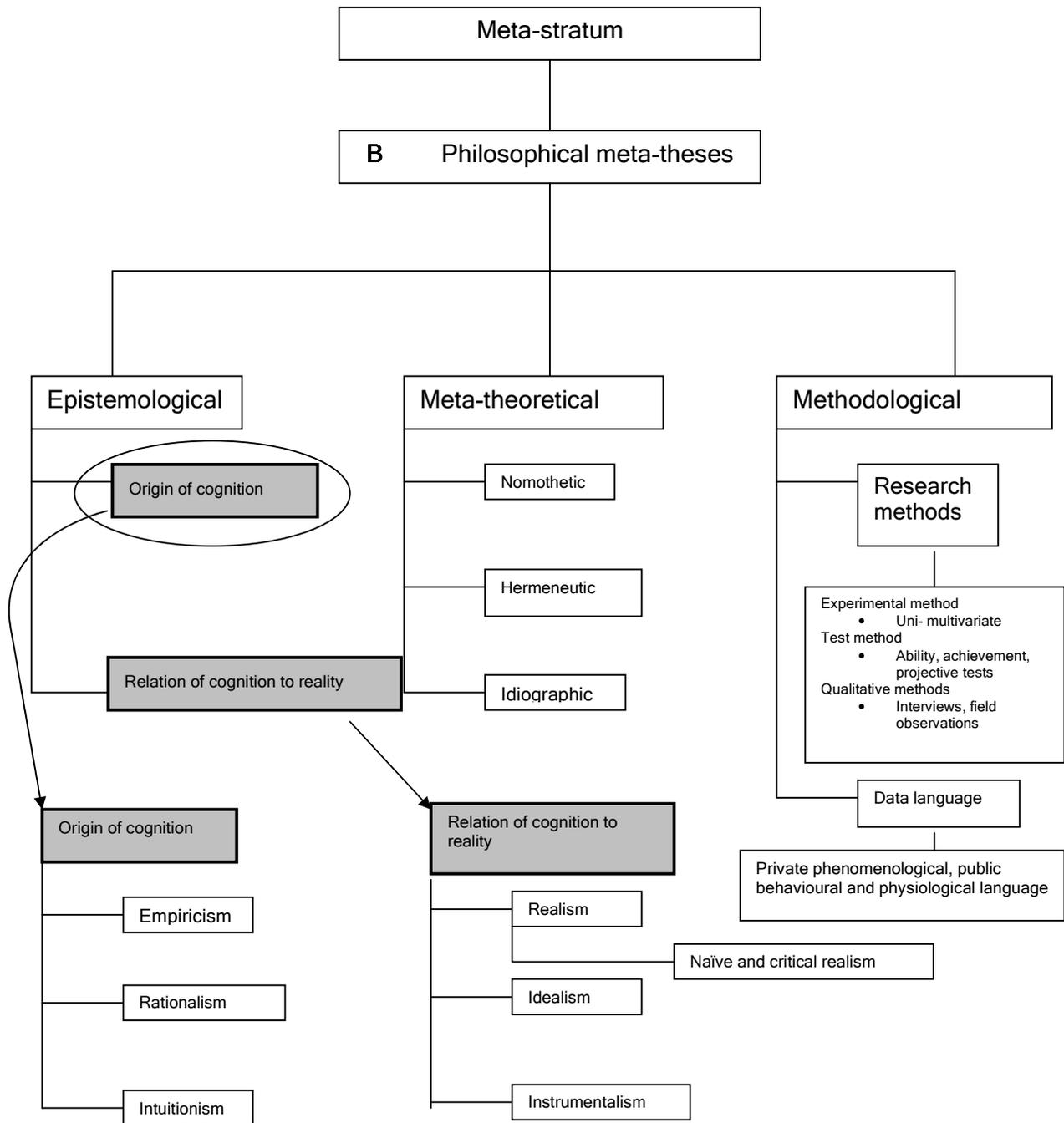


Figure 45 (b) Madsen's (1988) meta-stratum (B) Philosophical meta-theses



3.10.2 The hypothetical-stratum

The next level consists of H-level formulations of hypotheses and constructions of models serving as explanatory agents of the data level. This level is the constituent of theory (Madsen, 1992). The overarching meta-level is dominated by M-type methodology and principles regarding empirical data as well as meta-theoretical principles governing the theory one level down. Philosophical, epistemological and ontological presuppositions form part of this level of abstraction within the model (1985; 1987a; 1987b; 1992). Madsen contends that it is often the case that these presuppositions are not explicitly formulated but remain implicitly hidden within the principles and hypotheses. The hypothetical stratum forms the centre of Madsen's systematological framework and mediates between two levels, the overhead meta-stratum and the lower data-stratum. It consists of constructs employed as explanatory mechanisms seeking to shed light on observables within a system. This level consists of the following:

- Hypothetical terms
- Scientific hypotheses
- Hypothesis system

each of which will be detailed below.

Hypothetical terms

Hypothetical terms are employed to provide coherence to a system of observables for which there is no current explanation. Providing some sort of linkage between dependent and independent variables, hypothetical terms are often placed in the position of intervening link. For instance, during a dynamic assessment intervention, it is hypothesised that mediation is the intervening variable (a construct hitherto ascribed to some measurable and quantifiable intervention), a causal link tying the pretest and posttest scores in a causal loop. Intervening variable and hypothetical link are somewhat different in meaning, the former referring to empirical and operationally defined link; the latter referring to hypothetical constructs which have what is termed "surplus meaning" over and above that which is accounted for by the intervening variable. In order to better classify hypothetical terms Madsen structures these terms into three main sub-classifications:

- The ontological reference of hypothetical terms

Knowledge of the intervening variable is invariably coloured by the researcher's ontological assumptions whether they be made explicit or not. Background knowledge (conscious or unconscious awareness) and assumptions filter through to the nature of the hypothetical variables and can be classified into three categories:

- a) Mentalistic hypothetical terms (H_m)
- b) Organismic hypothetical terms (H_o)
- c) Constructive hypothetical terms (H_c)

Mentalistic hypothetical terms refer to constructs concerning the brain or mind; behavioural outcomes of mental states such as "consciousness", "intelligence" and "potential" and are typically employed by researchers and psychologists whose meta-model is of a dualistic, hyperphenomenalist or spiritualist type (refer to the meta-stratum above). Organismic hypothetical terms refer to processes and structures in the brain within the organism as a whole and might include such terms as "impulse control" and "patience" and are typically employed by researchers and psychologists whose ontology is materialistic, epiphenomenalist and neutral monistic (refer to the meta-stratum above). Constructive hypothetical terms are analogised versions of explanatory systems and can be construed as crude metaphors as was discussed above in Madsen's conception of model and are chosen by researchers and psychologists whose ontology is pluralistic or neutral monistic.

- The existential form of the hypothetical variables

Hypotheses may take on various forms and can include:

- a) Process terms
- b) Structure terms

Process terms are hypothetical terms which serve as explanatory mechanism for some sort of change occurring within the organism and include processes such as changes in perception, thinking and planning for instance. These process terms refer to features which are often of brief duration as opposed to structure terms which refer to features which can span a life-time.



Structure terms are a hypothetical terms which serve as explanatory mechanism for structural features of a system such as factors of intelligence, knowledge and acquired skills.

- The function of hypothetical variables

Hypothetical terms may have various functions and include

- a) Dynamic variables

Dynamic variables are mobilising (process and functional) variables which activate or seek to mobilise energy in some manner, such as understood by terms such as “needs”, “drives” and “motives”.

- b) Directive variables

Directive variables are those which have organising, controlling and regulatory effects and include cognitive processes and structures and include terms such as “thinking”, “memory” and “perception”.

- c) Vector variables

Vector variables are a combination of dynamic and directive variables which result in functional and process variables which seeks to activate and control. Constructs such as “instincts” would be an example of such a vector variable.

Scientific hypotheses

Hypotheses are underlying assumptions and scientific hypotheses seek to add credence to such terms by validating the hypotheses with scientific explanations. Hypotheses in general are abstract and pertain to levels of explanations which may or may not be empirically validated. However, this is not the case with scientific hypotheses which are formulated in such a way as to make them testable. The same cannot be said of general hypotheses which is what distinguishes the scientific hypothesis from the general hypothesis. Two types of scientific hypothesis are pertinent to the study of psychology in Madsen’s scheme, namely,

- the ontological classification of scientific hypotheses

Ontological scientific hypotheses can be further divided into

- a) existential hypotheses

Existential hypotheses are hypotheses which proffer the existence of some sort of hypothetical construct which later assumes a more functional and structured role within the hypothetical system. Learning potential as hypothesised construct might begin as existential hypothesis only to assume a functional role in the model.

- b) functional hypotheses

Functional hypotheses proffer the existence of a function or effect of some hypothetical construct and form functional relations between two or more sets of hypotheses. Functional hypotheses postulate either causal or stochastic processes between hypothetical variables and Madsen is of the opinion that functional hypotheses form the most important category of hypotheses in general psychology.

- c) historical hypotheses

Historical hypotheses follow the historical development of a system and include the many and varied hypotheses concerning human development and various developmental intelligence models.

- the meta-theoretical classification of scientific hypotheses

According to Madsen, there are two types of scientific hypotheses (meta-theoretically speaking) and include:

- a) theoretical hypotheses



Theoretical hypotheses contain only hypothetical terms and as such do not relay information about relations and functions in and between empirical data and hypotheses. They are by virtue of their nature untestable as they offer no means for evaluation. Such terms are notated as H-H and refer to hypothetical (transempirical) terms (see the section dealing with Madsen's HQ system in which this is employed).

b) empirical hypotheses

Empirical hypotheses offer at least one descriptive term which refers to either an independent or dependent variable along with an hypothetical term. Such terms are notated as H-S or H-R referring to stimulus and response hypotheses respectively (see the section dealing with Madsen's HQ system in which this is employed).

The hypothesis system

The role of the hypothesis system is to mediate between the data-stratum level and meta-stratum level and to systematise information in a number of ways already mentioned above in the meta-stratum discussion. As discussed, three types of systematising explanations can be employed, nomothetic, hermeneutical and idiographic. Madsen extrapolates within the nomothetic explanatory sub-system and be formed in one of two ways:

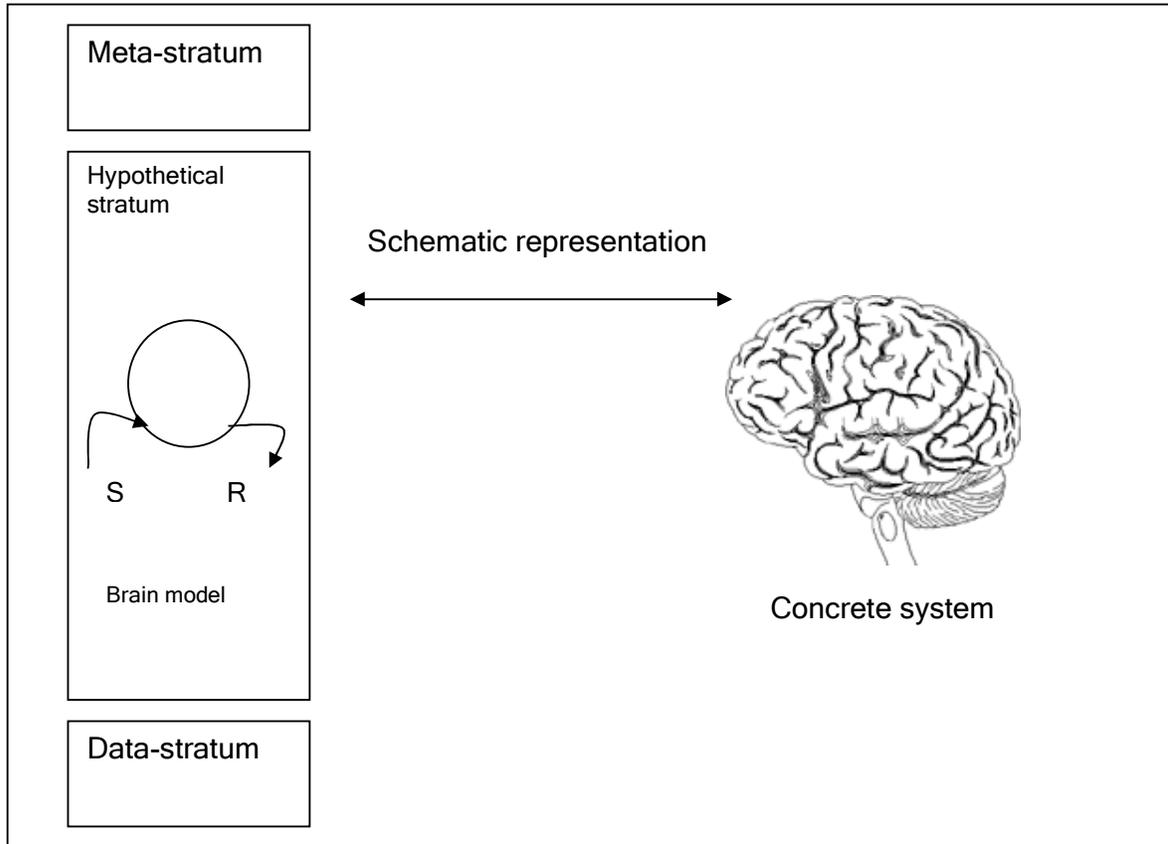
- deductive explanatory systems

This system utilises deductive arguments in a logical deductive structure common to most scientific arguments and takes the usual form of at least two premises followed by a truthful and valid conclusion. Madsen cites Hempel's deductive explanatory system as the standard account lauded by the logical empiricists which has become mainstream practice within psychology and intelligence research more so. Madsen's affiliations with relativist philosophers and theorists (Kuhn, Feyerband Hanson) do seem odd then when supporting such a nomological-deductive account of science (Estany, 2001).

- model explanations

Models serve to explain the functionings of other systems by means of schematics which often take the form of two and three dimensional illustrations. Figure 46 illustrates the level at which models function.

Figure 46 Schematic representation of model functioning, after Madsen, 1988



Brain picture courtesy <http://www.ccn.stir.ac.uk/brain.gif>

Interestingly enough, Madsen (1988) conceptualises mathematical (formula) deductive systems as special instances of analogy-model explanations. Hence simulations as well as graphic models suffice as explanatory mechanisms and it is at the level of model that dynamic assessment most often finds itself lodged. Models can be classified according to their

- a) degree of abstraction

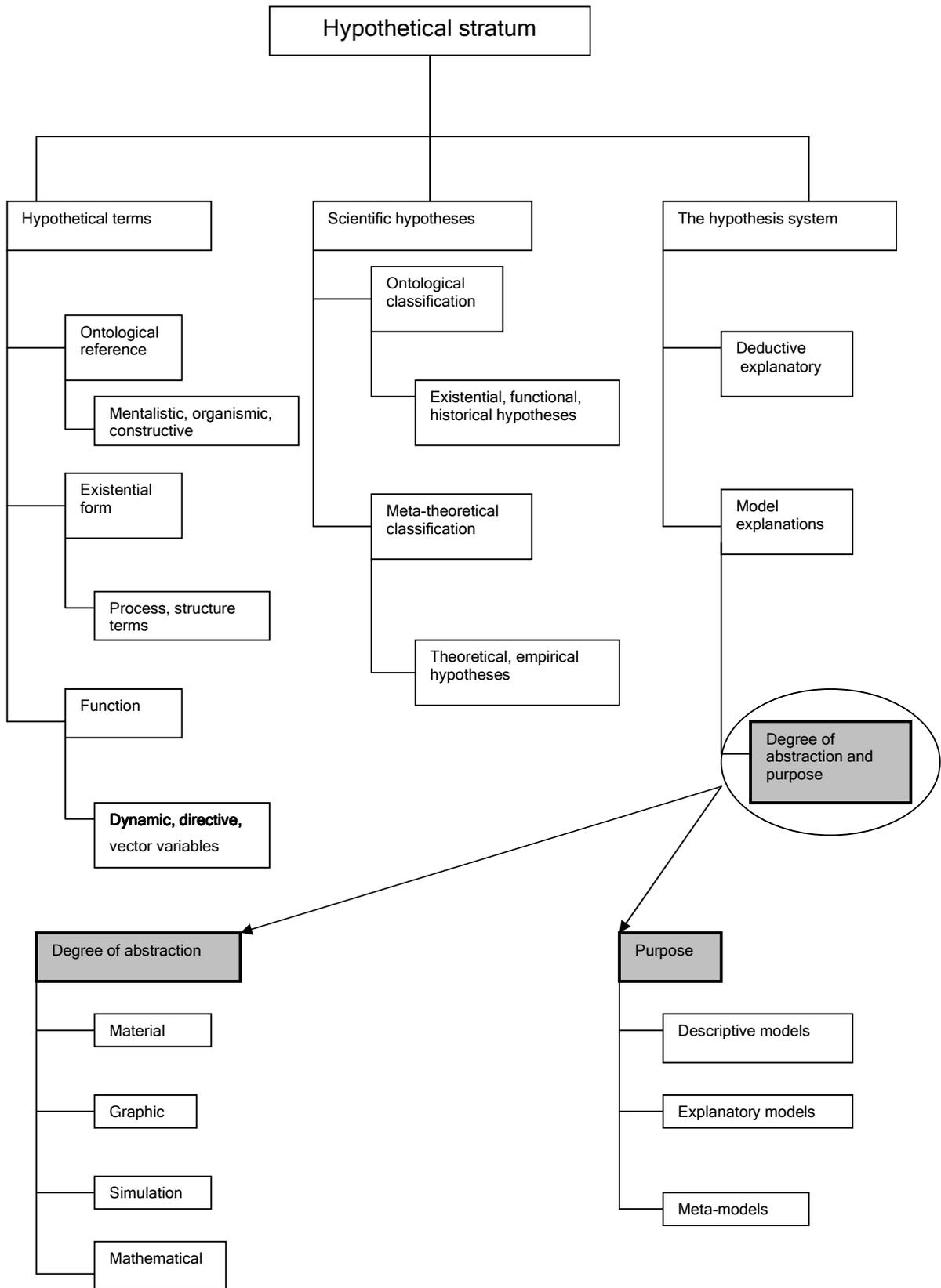
Levels of abstraction include the material three dimensional (globes to describe the earth), graphic two dimensional (maps to delineate the lay-out of land), simulation (computer software programmes) and mathematical models (symbolic deductive systems).

- b) purpose or function

Models' purposes range from providing a schematic descriptive representation of a part of the world, through explanatory models employing constructive models, physiological models and mentalistic models through to meta-models where they are usually implicitly implied throughout the explanatory system as global models.

The hypothetical-stratum level thus far discussed is illustrated in figures 47 below.

Figure 47 Madsen's (1988) hypothetical-stratum



3.10.3 The data stratum

The data stratum is no different from the preceding stratum descriptions in terms of the composite blend of both abstract and concrete aspects but is, as a whole, the most concrete of the three levels. The data stratum can be divided into:

The abstract data-stratum

The general data theses are contained within this section and contain exclusively descriptive terms which are highly abstract but contain no hypothetical terminology, and Madsen cautions the reader as to the somewhat confusing make-up of very abstract data terms in comparison to more concrete hypothetical terms and that the same term can be employed in both meaning categories depending on the context and usage. The empirical or data level forms the lowest level in this diagram and includes concrete descriptions from experimental trials and observations and also includes abstract relations which exist between various data. Madsen's short-hand describes this as the D-level. The data level stratum within this meta-science model is the core or essence of how science is practised within the socio-cultural milieu. The data stratum itself is divisible into more and less "abstract concretisation" of empirical information depending on the nature of the datum, either referring to something observable such as impulse control during a test situation or an aspect much less observable and hypothesised to exist such as the construct of potential as assessed during dynamic assessment. The abstract data theses are enmeshed with empirical relations which play out in one of two relations, namely, functional relations and correlational relations. Functional relations detail the interplay between dependent and independent variables and can be illustrated thus: pretest result (independent variable) in addition to intervention (independent variable) leads to a posttest result (dependent variable). In psychology there are two kinds of empirical relations, namely:

a) functional relations

These are relations between dependent and independent variables, notated as follows:

S → R

The functional relationship within a dynamic assessment set-up can be viewed as:

Pretest → Intervention → Posttest

b) correlations

Correlational relations relate two or more dependent variables in a manner in which the relationship is statistically measurable but not necessarily causal. The relation between performance on an intelligence test and performance on a mathematical test may be correlated due to the possible *g*-loading of both types of assessments and can be illustrated as:

R-----R

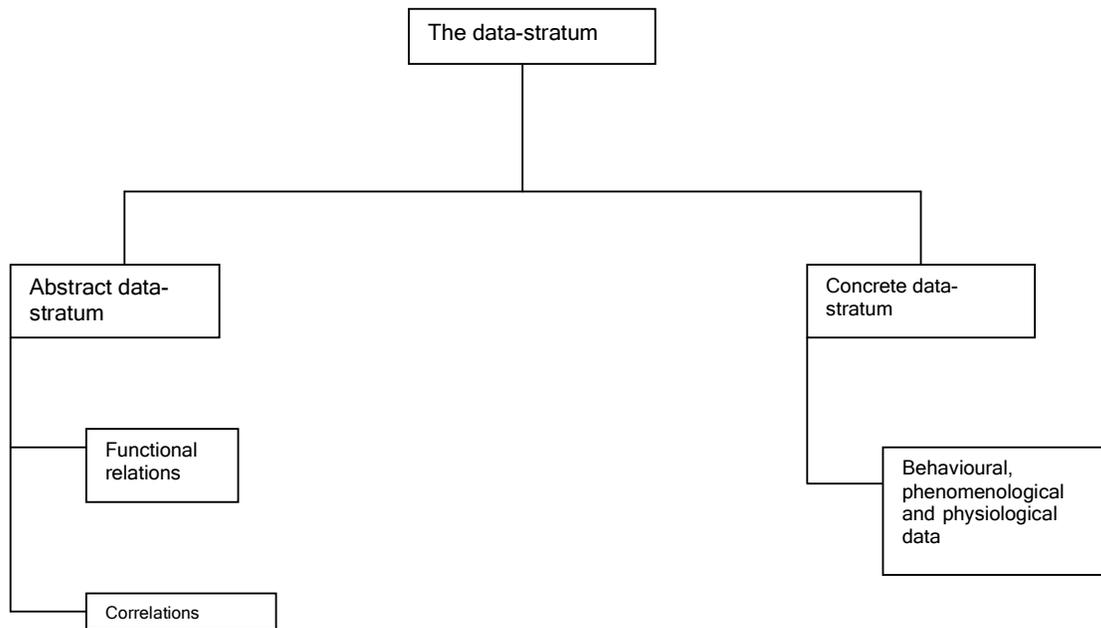
The concrete data stratum

Specific data theses reflect observables in the system and included among others descriptions of behavioural, phenomenological and physiological data. Madsen's firm contention of science as proceeding from top-down as opposed to bottom-up may be unjustified as both manners of approaching any discipline are necessitated, yet this is his contention when he states:

Modern meta-science does not regard science as built up from the bottom but rather from above - from the meta-stratum. It is the meta-stratum that governs the formation of the hypothesis-stratum and the methods for collecting the data that is then used in the revision of the hypotheses or still better, science has arisen and developed in an interaction between philosophy and the empirical facts (1988, p.59).

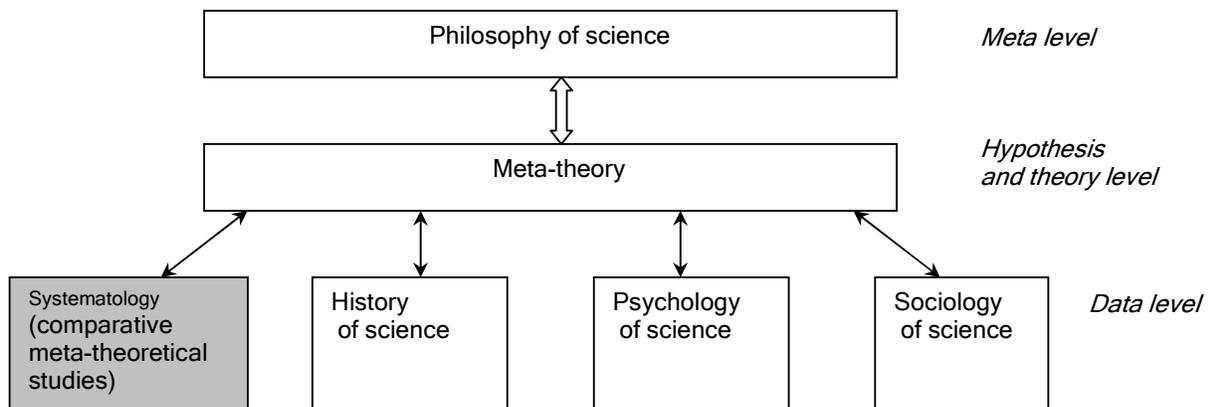
The hypothetical-stratum level thus far discussed is illustrated in figures 48 below.

Figure 48 Madsen's (1988) data-stratum



The composite scheme highlighted in the various figures above culminates to form Madsen's systematology. This is itself positioned within another larger encapsulating system and can be seen in figure 49 below. Systematology (shaded in this figure), forms one designated area of study alongside studies of the history of science the psychology of science and the sociology of science each of which belongs to the data stratum of this meta-theoretical conceptualisation. This data-stratum itself is placed below the meta-theoretical level which is superseded by the philosophy of science.

Figure 49 Diagrammatic illustration of the interrelations between various meta-scientific disciplines



The philosophy of science inspires theories about science (the theory level) which in turn produces data and can be empirically tested under various disciplines of scientific research. From Madsen (1987a), p.170; (1987b), p.263; (1988), p.12.

3.11 Considering the whole text

Many philosophically-inclined writings, schemes, models and theories are written in formats in which clear systematisation of these texts is not explicitly delineated and compartmentalised. In order for such texts to be numerically quantifiable and in order to determine its testability Madsen developed what he refers to as the hypothesis quotient (HQ) facilitating the quantification of ratios. Hypothetical theses as well as empirical hypotheses are counted and summed in a fairly straightforward equation. However limiting this approach may at first glance appear, it does capture the essence of the leaning of a theory in terms of its affiliations with data, hypotheses or philosophies. Various theories (or models as the case may be) emphasise one or more of the three strata discussed above in varying detail and hence display different levels of testability. Philosophical texts are by their

nature irrefutable and however incontestable their advocates may be, no serious attempt can be made to prove any claim emanating from the text (logical analyses aside). Texts emphasising hypothetical constructs without the requisite data to support these conjectures avail themselves of testability criteria, but once again only to a point, as they highlight models and hypotheses more so than testable data. The most testable would of course be theories with larger bases of data. Madsen works towards hierarchically arranging theories within the discipline and has chosen for his efforts a scheme which is commendable although could be construed as flawed depending on the view taken towards science as whole and psychology as sub-discipline within science. These objections have been noted above in this chapter. Nevertheless Madsen offers a point of departure which, however reductionist, is unconvoluted and thus easy to replicate.

3.1.1.1 The testability of theories - hypothesis quotient (HQ)

Natural science dictates ensure that theories are testable and in the process of constructing hypotheses in order to further this goal, quantifiable results¹⁰² make manifest something intangible in terms of perceiving a result to be either confirmatory or not. Psychology's hard science approach following in the wake of natural science advancements initially hoped for an empirical, quantifiable and verifiable heritage to which later was added Popper's falsifiability as further criterion (a criterion which has had serious argumentative flaws lodged against it, Worrall, 2002). To add credence to the theories of motivation Madsen set about devising a "quantitative estimation of the testability of a theory" (1984a, p.185) and labelled this approach the hypothesis quotient. As this approach forms the bedrock of his systematology it is considered prudent to briefly outline the process.

Madsen (1984a) acknowledges his indebtedness to philosophers of science who preceded him in terms of studies regarding testability criteria seeking to advance science in general. Among these giants are cited Popper, Bunge, Radnitzky, Törnebohm and Kuhn (granting the tension between Kuhn, Bunge and Popper). Madsen attests to the greater affinity Kuhn's theory has to reality (factual development) as opposed to the Popper and the logical positivists preceding him. The testability of an hypothesis or theory itself is prone to error¹⁰³ and as such Madsen (1984a) states that just as with all testability criteria his own can be considered as merely preliminarily acceptable.

Defining HQ

An hypothesis according to Madsen (1984a) is a formalisation of the functions between two or more variables, at least one of which is an hypothetical or transempirical variable (an unobserved intervening variable) and "is based on the proportionate relation between theoretical and empirical hypotheses" (Madsen, 1988, p.61), the method through which theories are numerically tested. Two types of hypotheses present: a theoretical hypothesis and empirical hypothesis. Theoretical hypotheses include the functional relations between two hypothetical variables and are designated as H-H hypotheses. An empirical hypothesis includes the functional relation between at least one empirical and one hypothetical term. The empirical term can be one of two conditions, namely dependent (response variables) or independent (stimuli variables).¹⁰⁴ Hence, empirical hypotheses constitute one of two types, H-R and H-S hypotheses. In the present scheme no place is allowed then for purely empirical hypotheses as the definition has to include one hypothetical construct. To accommodate relations between two or more empirical variables then, the term datathesis is added. This subsumes the functions between S and R variables as well as the correlation between two or more R variables.

Interpreting HQ and it's calculation

Due to the lack of explicit laws¹⁰⁵ (which resulted in incomparable theories) within the various theories of motivation surveyed by Madsen, HQ was redefined as an estimation of the potential explanatory power of the theory. Theories with greater explanatory power, writes Madsen (1984a), is in keeping with Popper's conclusion that this leads to better testable theories and so explanatory power is directly related to testability. HQ can be calculated by summing all theoretical hypotheses and dividing it by the sum of both dependent and independent partly empirical hypotheses and looks as follows:

¹⁰² See chapter four which discusses this quantitative imperative.

¹⁰³ Relativism, it seems, forever lurks under the surface. The author's avowed dislike of relativistic stances has already been stated in chapter 2 but she also acknowledges that it is an inescapable fact of science and scientific discovery.

¹⁰⁴ It is clearly evident here that Madsen deployed technical terms from the prevalent literature of the time. Theories of motivation were firmly ensconced in behavioural terms. That these theories were more amenable to this type of analysis is telling of the period in which these theories were developed. These types of theories are no longer the dominant forces they once were as psychology as a larger discipline now encapsulates a much broader view of human functioning. Another reason then that the fragmentary discipline should be mindful of the role of theoretical psychology and the streamlining effect that can result from efforts in mapping out future trajectories for the discipline.

¹⁰⁵ Still problematic. How many laws proffered by psychology are known? After all the quantification of data is gathered psychologists sit with heaps of data. That quantification is at fault is untrue as it is merely a tool. That the wrong tool is being used may seem to be a better question to ask. Or is it the discipline as currently defined that is at fault? Is this the reason dynamic assessment sits so uneasily within the domain of intelligence assessment (itself questionably seated within psychology). And so a cascading effect ensues.

$$HQ = \frac{\Sigma(H-H)}{\Sigma(H-S) + \Sigma(H-R)}$$

The higher the hypothesis quotient the lower the testability of the theory and is evident from the equation. The fewer hypothetical terms the better in terms of the theory's testability. For explanatory purposes Madsen's calculation of HQ as applied to two of Freud's theories¹⁰⁶ will be briefly discussed.

Madsen (1984a) utilises Freud's chapter 7 in his "Interpretation of dreams" (1900/1953) as basis of HQ determination of Freud's theory by analysing the text and thus extracting theoretical (H-H symbol) and empirical (H-S/H-R symbols) constructs. Firstly, there is a well argued rationale surrounding the choice of Freud's particular theory, followed by a systematic deconstruction of Freud's passage in which identification of constructs is made based on quotations taken from the passage and lastly a formulation is evidenced in the form of the HQ equation. Madsen concludes with a calculated HQ of 2 for this particular theory detailing six theoretical constructs, two stimulus empirical constructs and one response empirical construct;

$$HQ = \frac{\Sigma(H-H)}{\Sigma(H-S) + \Sigma(H-R)} = \frac{6}{(2+1)} = 2$$

This HQ is one of the higher among HQ's that Madsen (1984a; 1987) has investigated (an HQ of 0 has been calculated for Skinner's theory, which when viewed from this perspective is an obvious score) and is testament to the lower testability of the theory, although being one of Freud's many theories it is not indicative of the testability of his other speculative theories. This theory is contrasted with Freud's theory of anxiety (theory presented in Freud's "Problems of anxiety" 1926) in which Madsen (1984a) evidences an HQ of 1.2, thus allowing for greater testability when compared to his above-mentioned topographical (structural) theory. The nearer the theory is to the empirical level the greater the testability and hence the lower the hypothesis quotient. An example of the derivation of this HQ utilising Freud's theory of anxiety is given:

1. H (anxiety) - S (trauma)
2. H (perception) - S (danger)
3. H (perception) - H (expectation)
4. H (expectation) - H (anxiety)
5. H (anxiety) - H (defence)
6. H (defence) - H (change of perception)
7. H (drives) - H (perception)
8. H (super-ego aggression) - H (perception)
9. H (anxiety) -R (flight)
10. H (anxiety) - R (conscious experience)
11. H (anxiety) - R (organic processes)

There are two stimulus empirical statements, $\Sigma(H-S) = 2$; $\Sigma(H-R) = 3$; $\Sigma(H-H) = 6$. Therefore

$$HQ = \frac{\Sigma(H-H)}{\Sigma(H-S) + \Sigma(H-R)} = \frac{6}{5} = 1.2$$

In essence both the chosen theories evidence higher quotients when compared to the average HQ of other theories analysed by Madsen. The closer Freud's alliance to empirical bases though the higher the testability and lower the HQ become. The mistake of confusing testability with heuristic value should not be made, cautions Madsen, since un-testable theories are often superseded by their guidance values. HQ is a measure of a theory's testability and not of its heuristic value (Madsen, 1984a). It is hypothesised that this might well be the case with various dynamic assessment models within intelligence research. The notion of the methodology of the science-model as closely followed within natural science disciplines comes into question in areas in which this model may not necessarily be the best available model to follow. The case for theorising thus leading to empirical validation is stated in this chapter. Arguments against this reductionist exercise might well be grounded in the dissatisfaction of reducing complex theorising which may well not yield to quantification. Yet if a study such as this is to progress with even a tentative explanation of where dynamic assessment finds itself within the broader intelligence field, such reductionism is at times warranted.

¹⁰⁶ In agreement with Brandt (1984); Ettin (1984) and Schaeffer (1984); the author also thinks that had Madsen chosen another theory with which to explain the derivation of HQ, it would have been initially better received than in fact it was. Madsen (1984b) in his retort to commentaries makes this admission though; but defends other critiques about his systematological approach.



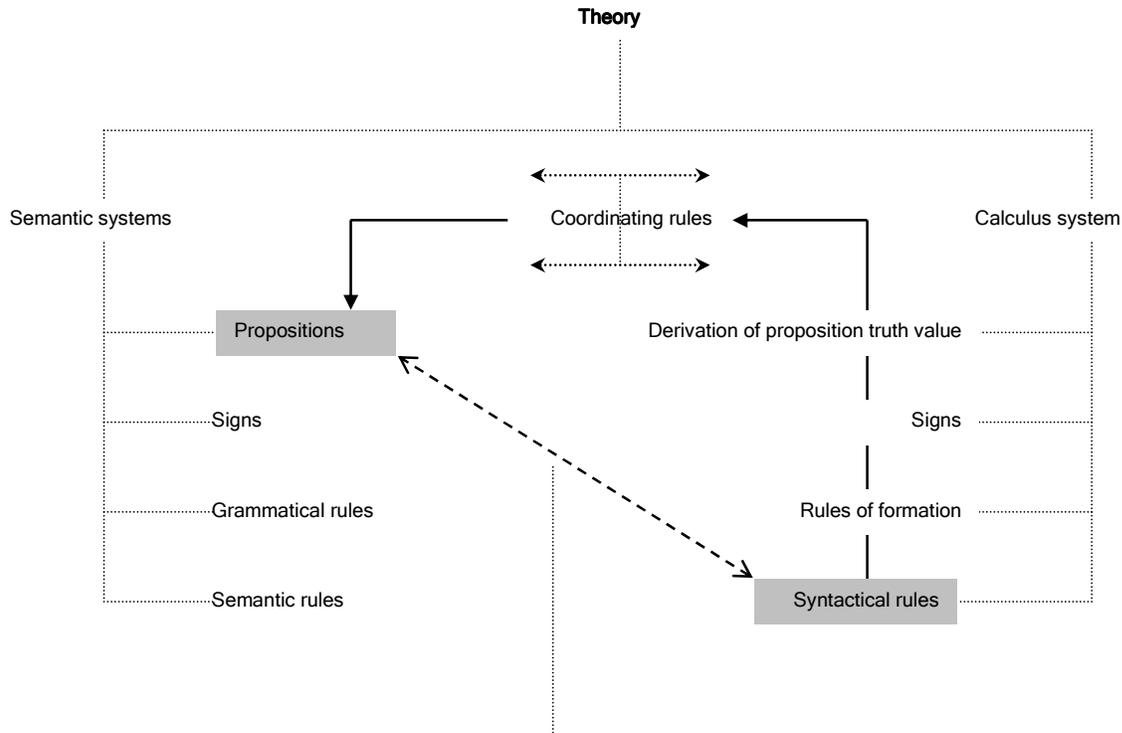
Madsen's HQ will be employed as a guideline only along which to estimate the plausibility of various dynamic assessment theories and models within intelligence assessment. Knowing that various strands of thought within these defined areas (dynamic assessment in intelligence research) are theories and models, this equation cannot necessarily be applied directly to all manner of dynamic assessment intervention models. As has already been mentioned such theory/hypothesis testing within theoretical psychology and specifically within meta-psychology is not often cited in the literature. The study does not concern itself directly with the tools of analytical philosophy¹⁰⁷ nor with the philosophy surrounding theory confirmation but considers a smaller area of investigation - dynamic assessment theory/model within intelligence research and the degree to which these theories/models are testable and their subsequent placement within a meta-theoretical framework.

However, a very brief explanation of theory formation is now addressed, but is done so from an analytical positivist view (Ross, 1964¹⁰⁸). Theories are customarily made up of a semantic system and a calculus system and are linked via co-ordinating rules. The semantic system concerns itself with general propositions of the theory whereas the calculus system concerns itself with the derivation of these propositions in terms of their logical truth value. Rozeboom (1970) refers to the calculi system as the uninterpreted theory and by assumption implies that the semantic system is the interpreted system. The system is diagrammatised in figure 50 as follows:

¹⁰⁷ The calculus of which is beyond this text. The logic involved in theory analysis, confirmation and revision can be sought in the following texts: Chalmers (1999); Harré (1988); Hempel (1970); Hempel and Oppenheim (2000); Howson (2001a); Huber (2002); Meehl (1992; 2002); Nagel (1979); Oldroyd (1986); Ross (1964); Suppe (2001a; 2001b; 2001c); Trout (2001b); Tuomela (1973) and Wilson (1985). For a very brief introduction to the types of logicism (verificationist and falsificationist) involved in theory appraisal McMullin (1978) offers a comprehensive overview.

¹⁰⁸ Inspired mainly by Hempel, Copi and Carnap who were concerned with, among other things philosophical, the logical analysis of language in which metaphysical statements are regarded as meaningless due to the lack of empirical refutation and lack of confirmation as well as concerned with the logical positivist criterion of meaning and the defining of theory as a formal calculus in addition to meaning (Mautner, 2000; Rozeboom, 1970). Upon further reflection, it may well prove fruitful "to have a go" at just such an analysis pertinent to dynamic assessment theories at some future date. Were the thesis of reductionism as set out in chapter 2 to become a reality in the future, there might well be a place for the logical positivist account of theory deployment, although currently, "the dream of the logical positivists that theories should have finished formal architectures has practically never come true in psychology" (De Jong, Bem & Schouten, 2004, p.280). It does seem rather ironic that analytical philosophical renderings of psychological theories have not really become routinely accepted within the broader discipline of psychology as this method is deemed as exemplifying the tools of philosophy - clearly at odds with current mainstream natural science approaches and methodologies within psychology (Jordaan, 1989). If analyses of this type is a rigorous a task as purported to be, what better complement of tools to utilise along with a natural science approach? "By and large conceptual housekeeping in mainstream psychology is directed/overshadowed by methodological considerations" (Jordaan, 1989, p.54). Chapter 4 discusses this in greater detail.

Figure 50 Formal analytical view of theory constituents (Ross, 1964)



Provided the theory is true, members of the syntax rules of formation are transferred into true propositions in the semantical system via the coordinating rules

In deploying his own framework for the express purpose of deriving a numerical quantity for testability, Madsen investigates the major theories throughout the history of psychology such as classical experimental, Gestalt, reflexology and classical behaviourism, psychoanalysis, humanistic, Marxist, contemporary behaviourism and mainstream psychology. The work culminates in hypothesis quotients for each major theoretical school or paradigm within the various branches such as chosen theories representing each school; a theory of learning, a theory of motivation, a theory of personality and a theory of cognition each subsumed within the greater school or paradigm of mainstream psychology.

3.12 An amalgamated Madsenian framework

The Danish meta-scientist, K.B. Madsen developed for psychology a suitable framework from which to numerically assign values of comparative worth and testability so that the discipline would in some measurable manner become more suited to empirical investigation. That this quest be considered ignoble by some anti-reductionist psychologists is an issue with which these said psychologists will have to grapple. The obvious advantages and disadvantages have been highlighted and will not be looked at again. His devised systematology offers an integrated and tightly packed consideration of how theories should be assessed and was himself very much influenced by various meta-theorists and philosophers of science. The appealing aspect of this framework is his frank contention that theories cannot be developed without due consideration for fundamental and underlying (and at most times, implicit) affiliations. Ranging from the greater historical and geographical areas of theory origin to the most minute datum detail, Madsen's systematology evokes appreciative applause for his attention to these wide ranging concerns from which the framework is built. His systematic and reductionist trends are in fact what makes this framework tenable in the first place. It is not a purely philosophical system entreating the researcher to ensure governability of thought and neither is it a purely data-orientated closed system in which theory development is considered isolated from the larger environmental systems. This framework has found for itself a niche for which this study offers fertile grounds for its own development and deployment of a meta-theoretical framework for dynamic assessment and intelligence. However, in its current form, the framework will need tweaking and realignment of sorts if it is to wield even greater impact for dynamic assessment models and theories. As befitting as Madsen's framework is to a study quantifying theory testability across the entire psychological spectrum, a similar framework is warranted for model and theory testability within dynamic assessment. The point of concern for this study when seeking such a framework as offered by Madsen is that the field of dynamic assessment is itself



unsure as to its theoretical/model/schematic status.¹⁰⁹ Consequently any framework built for the express purpose of comparing various models will have to concede to the field's rather blurred status in this regard. Now that Madsen's framework has been delineated the next step will be to attenuate the framework somewhat so as to allow for the utility of another framework developed for dynamic assessment. However, before Madsen's framework is attenuated for this study's purposes, chapter 4 and its focus will need to be considered. The framework will thus appear at the end of chapter 4 in an amalgamated and attenuated format.

3.13 Conclusion

This chapter has sought to make clear that underlying fundamental explicit but usually most often implicit assumptions predominate within a theory or model to the extent that original conception of an idea, the choice and implementation of a methodology, and the final analysis of data are all influenced by these assumptions. Depending on the nature of the study, various tools can be utilised to realise certain outcomes but more often than not, such assumptions form the study and thus partially form the outcome. Epistemological and ontological psychological positioning of theories and models propel the nature of ideas concerned within the theories and models. Due to psychology's scientific past and ever-increasing scientific present, natural science based approaches to the study of phenomena have largely overshadowed other ways of gaining access to knowledge. However, the more recent past (within the past two decades) burgeoning ideas within epistemological and ontological alignments (which have thus become more manifest as researchers have taken to understanding their positioning better) have started to rear their heads in a number of research areas and concern themselves with issues of constructing the reality as it is researched for instance. This is an idea which has woven its way through the fabric of many psychological sub-disciplines, including psychological assessment.

Such renewed attention to underlying philosophies concurs with concomitant changes in research domains (what can be researched) and how such domains are studied. Natural science's contribution to the study of the world (in any form) is breathtaking and wholesale importation of ideas, methods and ways of understanding has been partially successful within areas most aligned to physiological psychology. It is maintained that this method of gaining knowledge is a worthwhile endeavour and should be encouraged. However, within the same large domain rests avenues of pursuit within psychology that have not flourished to the same extent as the former. This is most likely due to the thoughtless importation of natural science methods into areas not akin to such analysis. Dynamic assessment is one such area of concern. The scientific status of this sub-discipline is not in any danger of being toppled because a scientific manner of such assessment is tenable – just not in the manner as described within the natural sciences model. The continuous overlap between domains has resulted in various fractures within the discipline leaving a trail of misguided epistemologies. Explanatory mechanisms within science, social science and psychology were discussed. A running concern within each of these areas was the nature of the theory and its verifiability in terms of its being falsified or verified (Popper's concerns). What counts as a science? What counts as a scientific idea, notion or finding? Natural science philosophy (whether scientists are implicitly or explicitly aware of it) seems to work for the vast majority of studies. As mentioned, technology attests to at least a partial amount of success: no science - no technology. Can the same be said of psychology though? This was highlighted and debated within the discussions.

A scientific psychology need not lend itself to a pure realist/rationalist approach because as is known, natural science studies themselves do not always progress on a firm footing of realist foundations. Relativist approaches appear everywhere within science and should likewise be an accepted feature of a scientific psychology without necessitating the strict rules and regulations that govern most natural sciences. Psychology as a broad discipline is fragmented partially owing to its mixed status as natural science / social science being predicated on natural and social science models simultaneously. It is suggested that if it is to remain in any manner a directed research area worth pursuing it should refocus and align itself according to the specificities of subject area. Hence, certain types of research can move into dedicated natural science domains and others can, for instance, move into more qualitative social domains (they need not necessarily be qualitative). A split would not be an entirely bad thing. This will go a part of the way in trying to dislodge psychology from a strict and essentially overly simplistic model of being "at the bottom of the pile" so to speak, allowing it more room to develop and become a more credible knowledge gathering enterprise without having to dance to the tune of natural science rigour.

Psychology theories are notoriously more difficult to define, analyse and interpret due to the overwhelming number of impinging variables, which is not to say that natural science theories are not likewise swamped with impinging variables; psychology's variables are overly intractable. In order to overcome such stalemates, techniques are employed to simplify models and theories in as parsimonious a fashion as possible which inevitably leads to reductions which are themselves questionable - but such is the nature of investigation within this realm. Definitions, concepts and propositions are necessary to make sense of raw data but

¹⁰⁹ One need only read the results of the informal content analysis conducted on the responses to the questionnaires in Appendix 2, which although only based on a small sample and thus not in anyway appreciably generalisable, does not allow for a conclusive conclusion to be drawn regarding dynamic assessment's status as scheme, model or theory.



are far more layered in terms of meaning than is the case within the natural sciences (an electron, although not yet observable, can be detected in a bubble chamber; Close, Marten & Sutton (2002); so even as a concept it is understood - in fact it was understood long before it was ever detected; Gribbin (2003); however, ID, ego and superego are not as easy to understand for obvious reasons). Just as theories in psychology are difficult to define, so to are they difficult to appraise, for the most part, because they are usually not tested in replication studies if at all. No replication, no theory proof or disproof. Just a remaining theory at the edge of the discipline. Do psychological theories strive for parsimony as conventionally understood in the natural sciences? Or do they make for themselves their own framework of parsimony? Is parsimony even attainable in complex theories? These issues among others were discussed before attention was shifted to the meta-theorist's work with which this study concerns itself.

Meta-theory is not a developed science within the social sciences much less within psychology. Many notions of what constitutes a meta-theory exists but very few carefully worked out models rendering an effective service exist. One thoughtful exception is the meta-theoretical framework devised by K.B Madsen. The framework cannot be imported without judicious consideration for the reasons necessary for utilising the framework, moreover, the framework needs to be attenuated for the purposes of this study. Various ideas concerning meta-theory exist; it serves to cumulate and study whole fields of endeavour; it serves in a constraining capacity; it allows for recommendations to the existing repertoire of theories in order for the discipline to progress further and it makes available insights gleaned from a view from above so to speak as opposed to views within from the morass of activity down below. Some posit that such endeavours are doomed from the start and that no such study can take place until the field under investigation has reached a necessary level of maturity. This thesis takes the opposite view by stating that it is precisely this view that is needed at dynamic assessment's juncture in history in order to offer it guidance in its future trajectory. Dynamic assessment, although youthful is most certainly not a novel set of theories and methods.

The singular taxonomic systematology devised by Madsen spans a great deal of what is considered important for epistemology and ontology within psychology and although Madsen focused on theories of motivation, his framework can nevertheless be utilised for the study of dynamic assessment. His encapsulation of governing cultural environments, scientific communities and the individual researcher in tandem with external histories of science (socio-economic), the combined external and internal history and the individualised biography of researchers attest to this breadth of scope of concern. Madsen's views on science and meta-science consist of empirical research results, the theory behind these results as well as the reigning philosophical thinking. He subsequently viewed theory as a set of defined hypotheses with requisite models accompanying them as well as a broader understanding of more mature scientific theories; i.e. his scope included developed and developing models and theories. In pursuance of a fully inclusive meta-theoretical model, he considered the meta-level, the hypothesis/model and the data with each level adhering to specificities of their own. Madsen views models as differing on the nature of abstraction and resultant purpose (or functions). Are models descriptive, explanatory or meta in nature? The discussion on Madsen's systematology included his views on the meta-stratum including ontological meta-theses and philosophical meta-theses. His hypothetical stratum consists of hypothetical terms, scientific hypotheses and the hypothesis system. The data stratum is described through considerations of abstract data and concrete data. His writings, although consistent, are quite nebulous in terms of an order to his work and the discussion on Madsen's meta-theoretical framework is a more concise effort in attempting to order it. To test theories and thus assign a numerical value to various models and theories, Madsen devised a simple formula for assessment, which may at first glance seem oversimplified especially for less defined theories and models such as evidenced in dynamic assessment. Nevertheless his attempts are laudable. The hypothesis quotient includes two dimensions; the hypotheses and empirical instantiations (independent or dependent variables) of these hypotheses as reflected in the specific theory at hand and is only an estimation of the potential offerings a theory can tentatively make. The calculation in no way attests to a theory's absolute quantified status as law-like. Fewer hypotheses indicate a more testable theory and thus a lower quotient and vice versa. As mentioned Madsen's HQ will be utilised as a guide only in an attempt to build for dynamic assessment a similar type of rating system.

Chapter 4 The fundamentals of assessment: prime considerations

4.1 Introduction

This chapter investigates the prime considerations within intelligence assessment; namely mathematical, statistical and measurement predicates upon which is based psychological assessment and measurement and attempts to link these prime considerations with the core philosophy of dynamic assessment. Such a view entails a re-evaluation of basic premises within intelligence assessment. These prime considerations can be classed under the rubric of quantification, often referred to in the social sciences as “the quantitative imperative” (Michell, 1997, 1999; Niaz, 2005). A re-evaluation is deemed necessary even though assessment has continued unabated from its earliest inception right throughout the twentieth century and on into the twenty first century. From the outset is made explicit that the mathematical formulations, logical derivations, statistical conceptualisations and the theorems, proofs, axioms and subsidiary deductions emanating from basic and proven tenets are not in question.¹ This chapter can only be read within the context of the preceding chapters, as many arguments employed and highlighted in those chapters are equally pertinent to the issues discussed in this chapter. Looking back on basic epistemological and ontological issues and discussing in turn issues such as reductionism, relativism, the nature of intelligence, the place of assessment within psychology and psychology as a discipline within the humanities as well as studying this method of knowledge acquisition as a form of a greater whole entitled “science”; illustrates how this all converges on assessment and where it is currently situated today.

4.1.1 The quantitative imperative²

Understanding assessment thus necessitates a re-look at its prime considerations, its foundations, the nascent fertile grounds from which it sprung and grew into what it currently is; a noble and at times misguided effort at assigning numerals and applying variegated statistical techniques to what are purported to be reified constructs in the hope of fulfilling a utopian ideal of equality and success. It has had a chequered past and has often fallen far short of these ideals which, at times, have yielded precisely the opposite results. As discussed in chapter 3, the growth of psychology as a discipline occurred in varied contexts since its inception as formally credited science. Having assumed the mantle of robust firmly entrenched natural science rigour and method as well as assuming hermeneutic, humanistic tendencies emphasising the notion of cultural relativism; pointing out the deficits of the hypothetico-deductive method of analysis and growth and seeking to amalgamate in some sort of consilient redress of the whole plethora of what it means to gather knowledge and what it means to “know”; the path towards psychology-as-science is indeed circuitous. As part of this same journey, is the questioning of psychology’s mathematical, statistical and measurement past to which attention is now turned. Psychology’s eclectic array of research efforts and individuals has been cited as a main reason as to why the quantitative imperative has been incorrectly aligned with positivism (Michell, 2003) even though naïve positivist methodologies have been cited as one among many reasons as to psychometrics’ less than sparkling reputation among some work forces at certain periods in time (Sehlapelo & Terre Blanche, 1996). Carnap was influential in putting forth his ideas within frameworks, stating that no framework could be judged as right or wrong, because it was not an assertion. The need to move science along in terms of bettering the whole enterprise necessitates various methods which seek to do just this, as long as the framework is useful. Stevens’(1946) scales have been incorrectly assumed to have emanated from a fully fledged and developed methodological framework. There was no fully developed framework but rather an attempt at a solution (Michell, 2003).

The goal of this chapter is not in anyway meant to construe the great edifice of mathematics, statistics and measurement as ill-founded and ill-used within the psychological arena. On the contrary, without much of what the afore-mentioned has offered psychology during its history as recognised subject of inquiry,³ a considerable amount of research would never have seen the light of day. What is contested though in this chapter is how these methods of inquiry are used within assessment and it will be argued that through thoughtful consideration of what in fact forms the foundation of assessment, dynamic assessment will be offered a place within the broader arena of assessment conducive to its own development. That dynamic assessment will replace any mainstream assessment is of course absurd and its main mission has always been one of concurrent usage throughout its history alongside conventional assessment. The move away from observable entities (whatever might be construed as an entity, let for instance, a test score be such an example), to statistical abstractions or artefacts somewhat spurned early psychologists’ work towards the infusion of statistical methods in order to aid in theory-building, thus moving

¹ The author is neither a logician nor a mathematician and as such will not even attempt to understand the finer grained truths emanating from such studies. However, the author can also not ignore the influence these aspects of study have had in the history of assessment and psychometrics.

² Michell (1999, 2003b). From the seventeenth to nineteenth century’s distortion of the Pythagorean philosophy of measurement (Barrett, 2005).

³ The author hesitates to use the word “science” here as a backlash of criticism is expected; although the use of the word as applicable in this context can be defended.



further away from what Michell (1999) refers to as “foundational issues of quantification” (p.104). Issues of inference and problematic measurement resulted in psychologists’ increasing tendency to rely on statisticians’ models bringing into question the validity and acceptability of the quantification of psychological constructs much later on. This situation now finds itself once again in a position of debate within the origins and development of dynamic assessment (and dynamic assessment’s resultant reluctance to make decisions on static one-time assessment scores alone; Wiedl, Guthke & Wingefeld, 1995). In order to systematise this inquiry into the prime considerations of the fundamentals of assessment the chapter is divided into four subsidiary sections, namely the;

- mathematical foundation
- statistical foundation
- measurement foundation and
- psychological assessment foundation

These considerations will be critically discussed from both historical and intelligence assessment points of view and the role they have played in the formation and continued use of assessment instruments across the globe. A main hypothesis is that dynamic assessment has not quite found its place within the broader intelligence assessment framework due to the misunderstandings of what in fact is meant by “measurement” and when assessed in this light, it becomes obvious that dynamic assessment is fundamentally, philosophically and psychologically a theory not aligned with traditional manners of assessment. This situation is made even more intractable when one considers the almost pure utilitarian value of many intelligence assessment tests (Barrett, 1998). In this instance, Barrett (1998) fervently states that intelligence as a trait construct has much pragmatic value but little causal theoretical backing (and hence little scientific value), mostly due to the enterprise of poorly thought out measurement. Figure 51 depicts the roles played by each fundamental level within assessment and how each level interfaces with every other level. The diagram illustrates the multi-directional flow of how measurement levels being predicated on a statistical foundation, itself predicated on a mathematical foundation, need to inform one another. Problems emanating at the measurement level as is often the case within behavioural assessment must be traced back to its statistical and mathematical foundations if any solution is to be found. Solving issues at only one level will not bolster the cause of re-looking assessment within psychology for instance. This is not to say that any level cannot function on its own, that is counter-intuitive and incorrect as evidenced in the argument that statistical manipulation cares nothing for the level of measurement those numbers are levelled at, “even if the numbers are the purest nonsense, having no relation to real magnitudes or the properties of real things, the answers are still right as *numbers*” (Hays, 1981 in Maxwell & Delaney, 1985). Statistics reflect the numbers not the constructs yet numbers indicate where we are, what we do, how much of it there is and is imperative to science, as science would be “impossible without an evolving network of stable measures” (Wright, 1997b, p.33). The job of adequate representation is the psychologist’s job. The chapter concludes with suggestions as how best to realign the basic fundamentals within the theoretical framework of dynamic assessment within intelligence. The foundations as illustrated here are not to be viewed as subordinate to any of the remaining two levels, (this is not a hierarchical arrangement) as the statistical foundation might well proceed from the measures taken before statistical manipulation can proceed. This diagram merely serves to position the three realms of entities or foundations upon and from which psychological assessment “results” or scores derive. As Eves and Newsom (1965) state, a particular philosophy can be equated to a process of refinement and ordering of experiences and values and in so doing find relations which are normally considered disparate and find differences between things which are normally considered the same. Hence, a philosophy is essentially a description of a theory concerning the nature of something. Lazarsfeld (1977) posits four main reasons why quantification (or as Pawlowski, 1980 refers to it, “quantophrenia”) was becoming an increasingly important part of the social sphere as early as the seventeenth century:

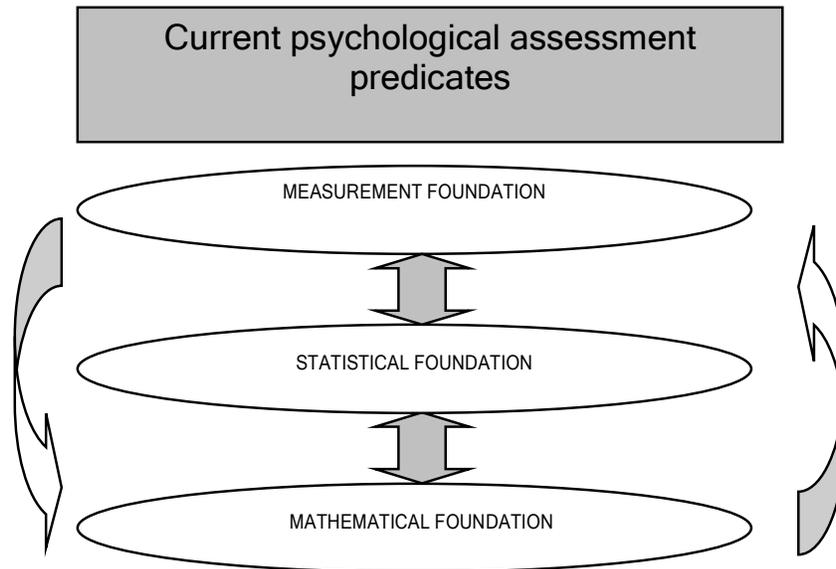
- i. the rise of capitalism
- ii. the prevailing Baconian spirit (see chapter 3)
- iii. a pressure to derive accuracy similar to that of the natural science in social endeavours and
- iv. the increasing role of public administration and planning due to burgeoning population sizes - much to do with insurance and the role of money and taxation

Ramul (1963) furthermore adds that psychological quantification, measurement and ratings were being practised well before such physicalist notions of measurement were discussed within the psychological domain and cites early usage of ratings and measures within areas such statistics, vision, memory, attention and thought (in which the “velocity of thought” was already being pondered as early as 1750, a prescient notion of the speed of neural conductivity perhaps?). Most measures of psychological import go back only as far as the start of the eighteenth century and of such measures carried out, only a few were

conducted by persons considered psychologists.⁴ The need for quantification in psychology proceeds along the following very narrow rationale (Schönemann, 1994):

- science is defined by its quantification via concatenation of its constructs
- any discipline wishing to call itself a science must adhere to this principle
- only if this is so can the discipline be called a science
- psychology does not possess constructs which are quantifiable via concatenation of these constructs, hence
- psychology is not a science

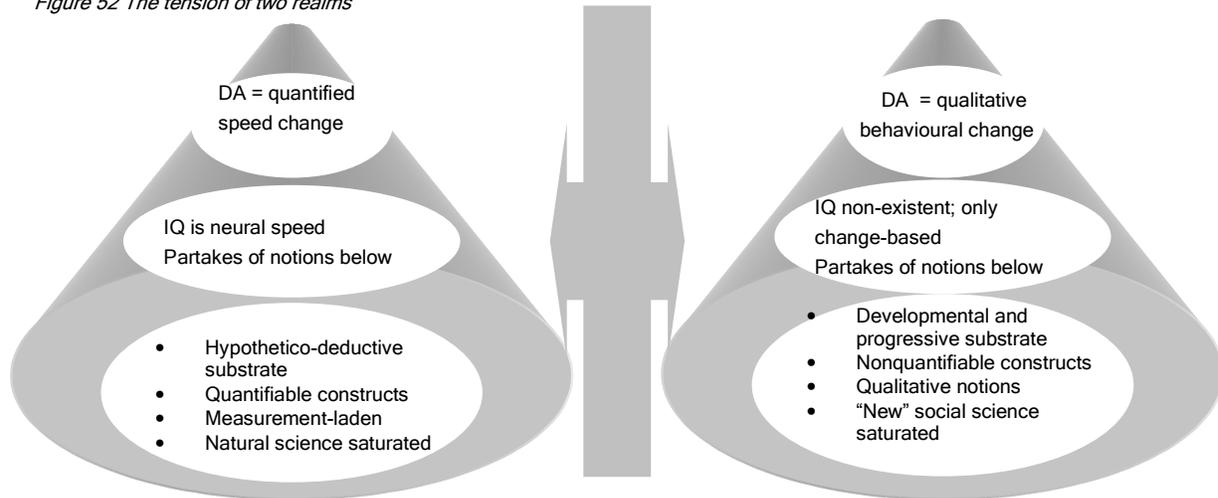
Figure 51 Assessment predicates



Measurement rhetoric has done a grave disservice to the psychological assessment enterprise and no manner of sophistication as exemplified by statistics and mathematical modelling can ever hope to rectify a problem which is clearly insoluble from this approach; the problem needs to be solved from another level entirely. Currently psychology exists within many realms, each purporting scientific accuracy, reliability and validity. These terms can lose their meanings very easily if misapplied in a variety of contexts. Psychological assessment should either reside in the domain of quantitative hypothetico-deductive development or it should reside in the nonquantifiable realm and progress in its own manner akin to the progress evidenced in the natural sciences. Neither realm is "right" or "wrong". To talk of the correctness of such realms is tantamount to misdirected and misinformed notions of what it means to conduct enquiries via truth-seeking mechanisms. Dynamic assessment lacks a fit or is at the very least a poor fit in the measurement models adhered to throughout the history of mental test theory. It is possible that it has found itself located in the wrong realm, which can hardly be the fault of dynamic assessment researchers and practitioners as they are merely fitting in with current mainstream concepts of what it means to practice psychological assessment. Figure 52 illustrates the tension of two realms.

⁴ Ramul probably means that, although they were not formally designated as psychologists, they could for all intents and purposes be considered as such.

Figure 52 The tension of two realms



Dynamic assessment is currently placed in both realms, thus resulting in tension which has yet to be resolved. It must choose for itself a realm in which to lodge and grow. However, this is not only a fault of this sub-discipline but a pervasive trend within the whole of the psychological discipline. Michell's (2001) pointed criticism is levelled at the misguided efforts touted by the social science measurement effort which he envisions as "instances of the scientific method applied to psychology" (p.211). The manner in which psychometrics is taught, he says, subverts the scientific method. Once again, it is necessary to reiterate that the tools of a trade are not necessarily at fault; it is the incorrect tools which are being applied which is very much at fault. The difference inherent in scientifically experimenting *a priori* and instrumentally going about the practical work of extracting scientific concepts are the two tasks of quantification (Michell, 1997). Yet utilising the instruments before one has worked out the scientific basis for measures is "pretence of science" (p.359). There is thus a call for a change of tools-to-trade within this thesis which will hopefully go some way in alleviating the current misfit of appliance and trade. Utilising psychometric tests to test psychological constructs is not a proven mechanism and remains at best hypothetical (Michell, 2001). Rumbings of the soundness of psychometric measurement can be traced back to the first quarter of the twentieth century and is thus hardly a new concern for psychometrists (Maraun, 1998; Stevens, 1946). However, due to "big business" psychological measurement and the seeming lack of mathematically trained psychologists, the utilisation of measurement and the ever increasing sophistication of statistical techniques (Barrett, 2002, 2003; Blinkhorn, 1997) the characteristic lack of enthusiasm for test use by psychologists is evident (Maraun, 1998; Michell, 2005). The following citation partly sums up the essence of this chapter.

"The post-Second World War methodological consensus in psychology combined a variety of elements thought to be necessary for scientific rigor, such as null hypothesis significance testing, [see below] Fisher's work on experimental design and analysis, [see below] and classical test theory [see below]. This consensus occurred at a crucial time in the history of psychology. Patterns of funding for universities and research were undergoing unprecedented changes in the USA and the effect was to set in concrete a methodological consensus that owed more to the values of window-dressing than to any values implicit in logical positivism" (Michell, 2003, p.16).

The issue of meaning-ladenness and measurement of a construct is one suffused with confusion (Barrett, 2001). Utilising \square as measured construct is tautologous, as psychometricians assign \square to the supposedly quantifiable construct "intelligence" and then seek to measure \square . Upon locating it along a continuum of "less-to-more" \square is upheld as existing (Maraun, 1998) which is clearly absurd. From this point onwards, the robust and sound statistical techniques used become ever more detailed and inherently presumptuous in terms of manipulating \square in manners which falsely bespeak of its existence. Constructs are identified *a priori* as existing, but as to whether they do or not is another philosophical question altogether. Added to this is another erroneous Pythagorean assumption that all attributes are quantifiable (Barrett, 2005; Michell, 1999, 2003). Assuming that \square measures intelligence in some manner, this will need to be done within a context of rule-bound associations which are themselves products of human behaviour. Rules are not empirical facts or findings, but a constituted set of instructions to follow (Maraun, 1998).⁵ Having identified supposedly quantifiable constructs, techniques further constrain interpretations of findings in such a manner as to lead to the acknowledgement of a methodological artefact and nothing more (Barrett, 1998). This turns the artefact into instantiated fact which is a leap not always scientifically condoned. What occurs here, states Michell (1997), is a gross

⁵ As Barrett (2001) states "the IQ variable is created - which is constructed ad hoc from one or more of the constituent technical concepts. IQ is taken as a 'measure' of 'intelligence' by some, but not by others. This is because the meaning of IQ is conflated with specific technical concepts and then mapped onto an arbitrary concept of 'intelligence'. What's worse, the biometrical geneticists then try and use IQ as a proxy for intelligence in order to 'find' genes for IQ test scores" (p.37).

instance of “thought disorder” which is defined as blindly following on in a tradition of delusion of simply not acquainting oneself with methodological concerns as it pertains to psychological measurement. This, states Michell (2001), subverts the scientific enterprise and is an instance of psychometrics playing the role of a diseased method (pathology of science). Notwithstanding these strictures, there is also the very important point of having no identifiable and workable common unit or metric according to which to measure the purported construct (Barrett, 1998; Kline, 1998). Measurement is theoretically assured if, as Wright (1997) maintains, the following is adhered to:

- to measure is to infer (which is precisely the leap made within assessment, otherwise there would be no point)
- measures are obtained by stochastic approximations
- measures are one-dimensional
- measures are counted in abstract units which are of fixed sizes (note that the abstractions are of fixed size which need not necessarily indicate that the constructs being measured are of fixed size - for more see below)
- measurement results are not influenced by outside influences

Clearly, the path followed by the natural sciences works for the natural sciences and it could work for varying sub-disciplines within psychology, only if stringent rules are followed and implemented. There are areas amenable to such treatment and there are areas which are not amenable to such treatment (Barrett, 1998; Borsboom, 2005; Michell, 1997). It is the job of psychologists to piece together their discipline and sort out these issues and to determine the fit between theory and data or nomological network (Strauman, 2001). It is clear that psychology's challenge is to account for its scientific status in a manner unlike that of the natural sciences (neither observability nor levels of analyses will suffice as method of knowledge-gathering as psychologists work with inference from observation and theorising; Strauman, 2001). Maraun (1998) follows Wittgenstein's arguments in terms of rule-based measurement and argues against what he considers a conflation of conceptual and empirical issues when rendering the measurement issue within the context of construct validation theory, as espoused by Cronbach and Meehl (1955). Measures, states Maraun (1998), and how we measure are not empirical issues but logico-grammatical issues and attempts at solving conceptual issues via rule-based measurements (predicated on human judgment as to what constitutes these rules) cannot be equated. Cronbach and Meehl's (1955) predicates for construct validity are based on a nomological net of concepts derived from this particular philosophy of science and is applicable more so to the natural sciences than to the social sciences. The concept of validity within the natural sciences is redundant (Kline, 1998) as concepts are public domain (more obviously manifest) and detecting error in what is purportedly measured is far easier within the natural science domain (which is not to say that all constructs within the natural sciences are amenable to immediate comprehension). Laws (statistical or deterministic) are set forth (akin to rule-following) according to which observable properties are somehow related or observables are related to theoretical constructs or different theoretical constructs are related to one another. Empirical discovery and construct validation are not synonymous and the more of one does not imply one obtains more of the other. For instance poor scores on a maths test at school does not necessarily indicate lack of proficiency in maths years later. The score changes over time but the construct of underlying maths intelligence has not changed (at least this is the perception within mainstream assessment). Regarding intelligence, Bardis (1985) states “discovering a functional unity by means of correlation has nothing to do with inventing a faculty and attaching a label to it” (p.219). The limit of substantive utility has been reached and according to Barrett (1998), the purely methodological approach no longer suffices as means of measurement.

Think of g and one immediately knows of much research attesting to its (at least) statistical existence but substantive psychological theory has yet to account for it, not to mention trying to explain what it is in a psychological sense (Kyllonen, 1996). Barrett (2005) asks the pertinent question of whether modern psychometrics, which is now so dominated by psychological statistics, has lost its way in terms of concentrating on the substantive thesis underlying the statistical thesis. Recall Meehl's (1967) lament on the lack of fit between theoretical and substantive theories in terms of null hypothesis significance testing. A visual aid might assist in more fully understanding why this is so. Figure 53 illustrates the nebulous area occupied by a construct entitled “intelligence” and the well defined area encompassing a small section is the partially identifiable construct validation. Figure 54, adapted from Oberhauer (2005) whose model is delineated for working memory as construct can easily serve as generic model in this instance. His model illustrates succinctly the circular movement involved in theory-measurement-data-theory as process towards construct definition and validation.

Figure 53 Empirical construct validation and meaning-ladeness of true construct

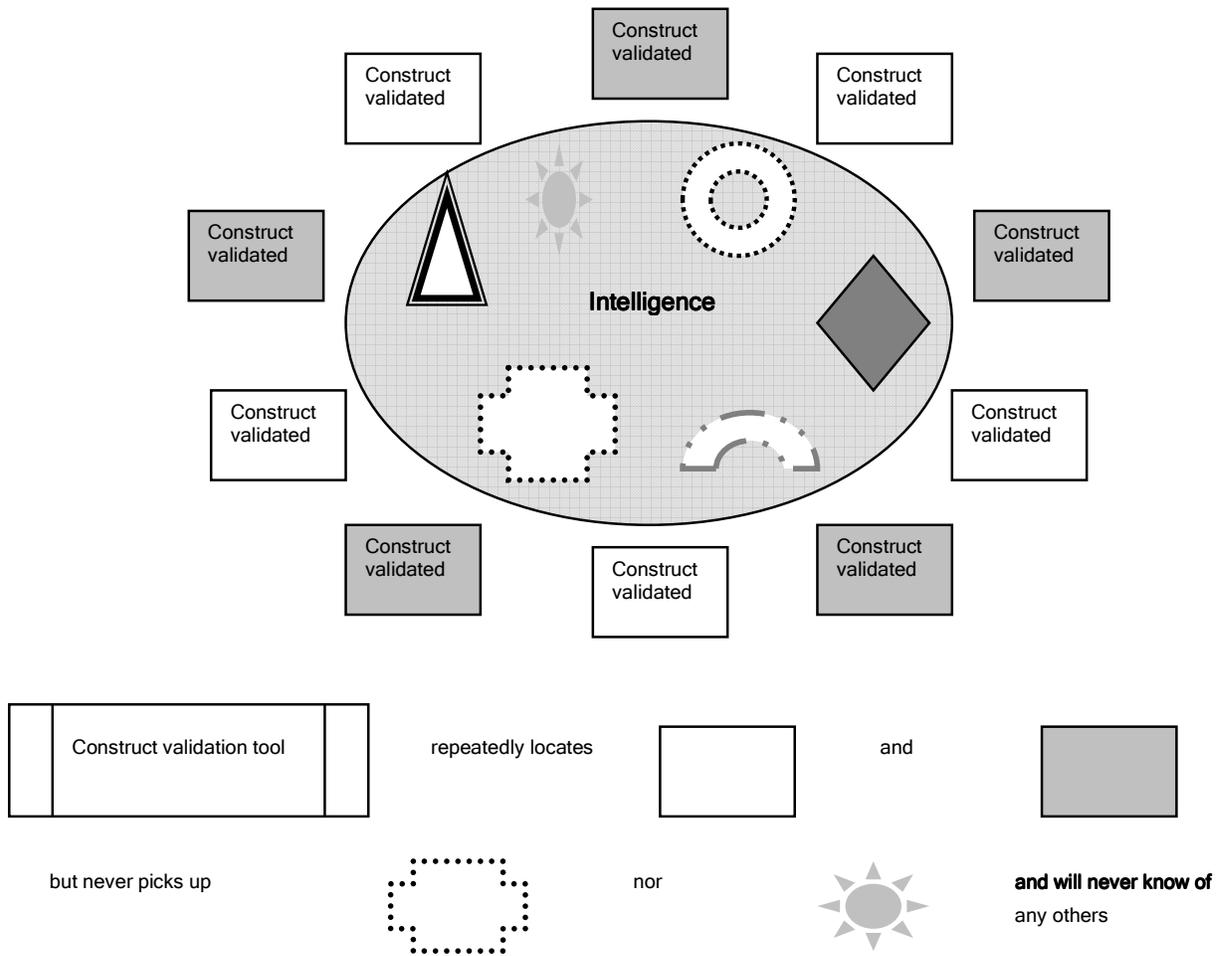
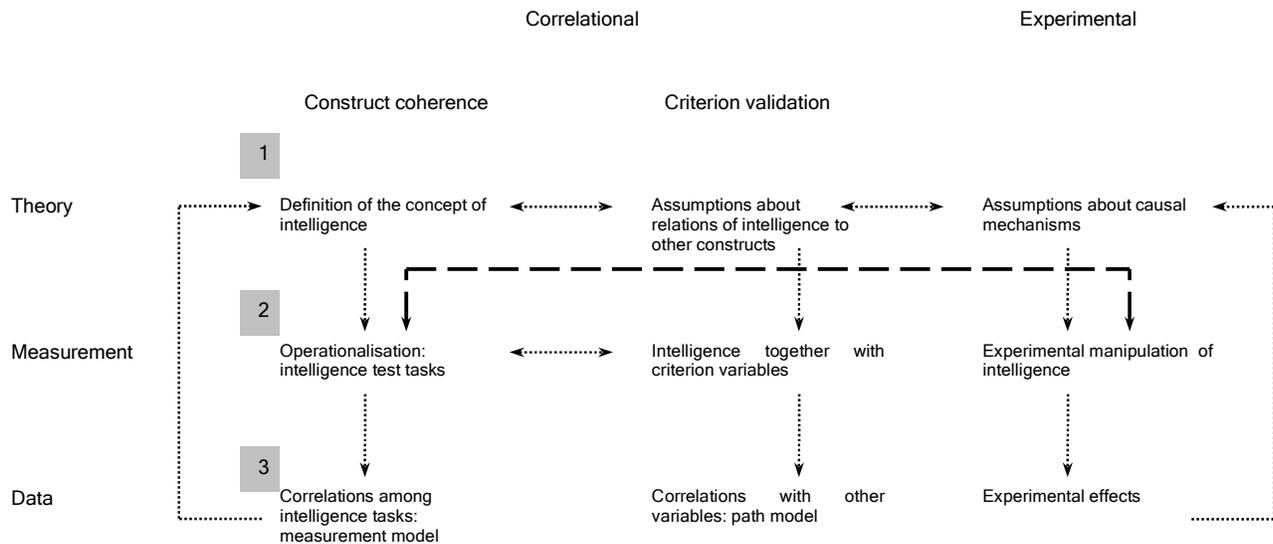


Figure 54 The vacillation between theory, measurement and construct validation (adapted from Oberhauer, 2005, p.394)



However ...

- 1> It is obvious that there is no such standardized definition
- 2> This has been referred to in this thesis but has not been delved into, yet there remains considerable paucity of research attesting to the validity of the actual tests utilised in various test batteries
- 3> Circular argument which is pervasive across the discipline including dynamic assessment construct validation; mentioned on numerous occasions

This figure is of course hypothetical and merely a thought experiment. The construct of “intelligence” is validated a number of times purely because the exercise happens to locate a number of similar constructs housed within the concept. The tools of the trade are precise enough to be able to locate a number of these correlating constructs and are thus able to conclude that the entire construct is possibly one large square shape. The reality is strikingly different! The areas “picked” up by tools of validation are indeed there but make up only a small percentage of the actual existing construct, which for the mean time, is beyond the reach of the technique. Probable models themselves are also merely tools for digesting information content which add little more in the way of credence to a latent attribute, assuming that such a latent attribute exists in the form we speculate about (Barrett, 2001). Building models predicated on constructs which have no doubt been validated as the above figure illustrates is precarious because a number of assumptions are being made, among others;

- that the construct in fact exists (it is possible that it likely exists in some form or another)**
- any tool utilised to search for the construct is purpose-built (with constructs being perpetually inferred; Utley, Haywood & Masters, 1992). This is tautologous as it is speculated that σ exists, a technique is brought in to search for σ which was custom built to find σ . Given this logic, it will in all likelihood find σ .⁶ *A priori* considerations will lead a technique to search for an *a priori* concern (Williams, Zimmerman, Zumbo & Ross, 2003)
- the nature of σ is housed within the context of science progression itself lodged on the bedrock of any variety of philosophy of science schools
- searching for σ is loaded before we ever start the investigation
- σ is found and conclusions are drawn in favour of its empirical existence
- σ is rarely noted for being an extension of science practice dictated to by the above considerations

⁶ Spearman having designed factor analysis to locate for his theory of intelligence a common factor of intelligence also went about developing classical test theory (Maraun, 1998; Williams, Zimmerman, Zumbo & Ross, 2003). This can hardly be considered a co-incidence. His endeavours never seemed to question the very utility of quantifying the very constructs he sought to measure (Michell, 1997).



- σ is defined according to the construct definition used. Any construct is as equally valid as any other provided one has followed the strict tenets as laid down by whatever model or school one happens to endorse. This unfortunately still tells us nothing about the actual underlying nature of the construct (Barrett, 2001).

** if σ exists, it may exist in the form which is amenable to extraction via techniques employed to extract this particular meaning from it.

In such a case, neural conductivity might well function as intelligence correlate. The correct tools should be employed to determine this. Tools are not to be borrowed from domains from which they were clearly not designed. Psychometric measures and physiological measures are bound to correlate but how sure can we be that this is not another instance of Meehl's (1990, 1997) crud factor? Measurement of characteristics follows directly on from definitions of both measurement and the characteristic at hand (Maraun, 1998). The logic of quantification in the social sciences usually runs as follows (and is perfectly acceptable practice bar one is in the correct domain in which one can apply such strategies); the first level of quantification would be one of assigning constructs numerics, in other words the metrification of constructs which proceeds with summarising statistics of these numerical counts and ending with the mathematisation of supposed theoretical entities (Meehl, 1998). As has been argued above, the logic flowing from this argument is sound enough given the correct circumstances in which to practice such dealings. However, the first premise is flawed if one considers that numerosity of entities supposedly existing is both right or wrong depending on how one views the situation. Numerical assignation as discussed in 4.2.2 below relates to this discussion on measurement, hence the need to include in this larger debate the mathematical foundation of measurement. Ross' (1964) delineation of the formal analytical view of theory constituents (section 3.11.1 in chapter 3) is similar in nature to his numerical assignation discussion pertinent to this chapter. Depending very much on the nature of a strict one-to-one isomorphic⁷ relation between hypothesised constructs and their numerical counterparts, some schemes are more amenable to numerical assignation than others and this assumes a common unit across the discipline thus making it axiomatic (see section 4.2 on mathematical foundation below) (Barrett, 2000). Most often, as is the case within psychology, the question as to whether one can even assign numerals to nonquantifiable traits is never even asked (Barrett, 2003). Hypothetico-deductive means of investigating nonquantifiable aspects exist and have existed for a considerable time (such as facet theory within intelligence research and cellular automata; Süs & Beauducel, 2005). The question then is, why such methods cannot be employed within psychometrics (Barrett, 2003).

Figure 55 illustrates Ross' (1964) paralleling of formal physical systems and number systems. Changes in the one system do not necessarily mean that similar changes will manifest in the other system. This is so due to the formal properties of both systems which can differ in ways not understood even though translation rules have been put in place. Figure 54 needs to be understood in tandem with figure 52 as well as figure 55 which is the more modern rendition of this tension between the physical-to-measurement concern. Figure 56 illustrates clearly the divide spanning the construct g which is an hypothesised notion only, although supported on numerous occasions, it is supported via a nomological network of mechanisms which were purpose-built to find g . Scores on measures of tests proposing to test subject matter exists as such. However, the link between the measure of g (as defined by and searched for by techniques developed to do just this) and scores on subject matter tests is highly contentious. Is there a mismatch between ability as predictor and task performance as criterion (Ackerman, 2005)? Chapter 2 detailed the author's preference for the notion of g and all that it entails philosophically. This is upheld as the position taken. There is no contradiction inherent here. G , if it exists, can most likely be probed via methods and tools amenable to such probing. Currently the wrong tools are being used. The deployment of theoretical models, schemes and other myriad conceptual frameworks of how the brain functions during intelligence tests are premised on statistical findings which are derived from mathematical artefacts themselves products of underlying mathematical models which guide the interpretation of psychometric test findings.

The assumption underlying this path of inference is that there seems to be a one-to-one correspondence or mapping of theoretical mathematical entity and actual brain functioning and processing. Granted that sophisticated neurological and less subtle invasive techniques were not yet developed during the heady days of Pearson's statistical development and analysis of data, a mathematical tribute and contribution to the measurement of human abilities was conducted in part by Galton and all their successive followers within the same tradition (Nunnally, 1978). Notwithstanding the great leaps forward in intelligence research the role of mathematical modelling to fit data at a behavioural level is perhaps currently overstated or perhaps slightly outmoded. Mapping local brain functioning on a one-to-one basis with models developed from this type of data would seem to be more parsimonious than what has gone before; advocacy of parsimony is not an underlying assumption within the preceding statement, as it is erroneous to think that in all spheres both behavioural and social sciences will indefinitely progress in a parsimonious fashion. Cognisance is taken of the fact that by the mere introduction of any one measuring technique immediately

⁷ Isomorphism is more strictly speaking the relation regarded as a one-to-one mapping. Homomorphic mapping includes other elements not mapped onto the construct (Michell, 1999). Homomorphic measurements include various objects than can be assigned the same number.

delimits the area of investigation in terms of what can and cannot be observed and/or derived. No one single technique known currently to either sphere of natural or social sciences can indeed derive all aspects of a function or structure.⁸ There are however methods and models available today which come closer to doing just this as opposed to the continuous utilisation of some assumptions inherent within models followed from the early days of twentieth century psychophysics

Figure 55 Measurement theory and theory of positive real numbers (Ross, 1964, p.60)

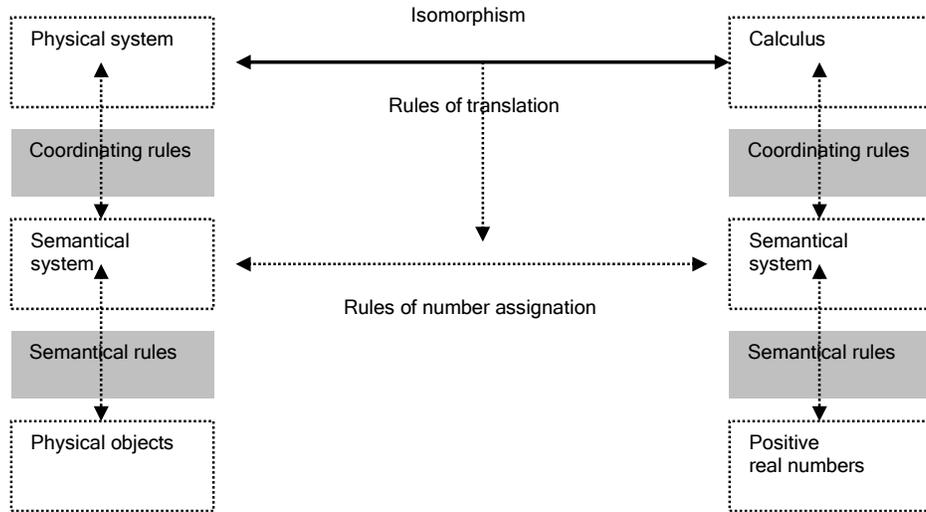
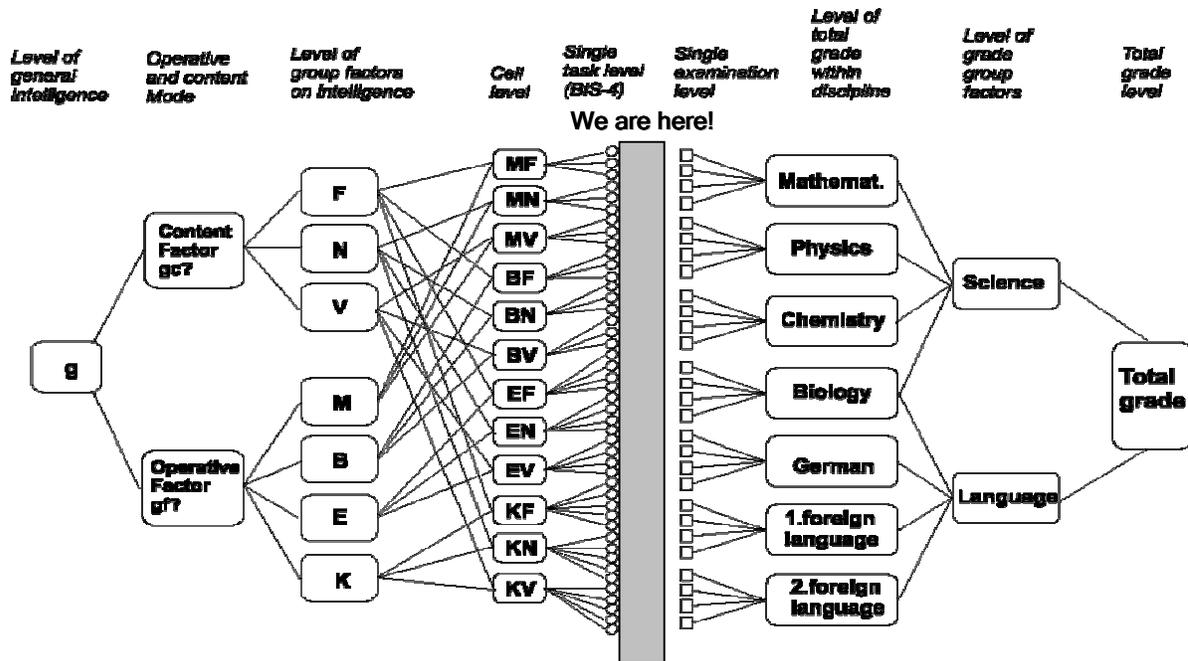


Figure 56 Baret's (2000, p.41) utilisation of Wittman's (1997) illustration



⁸ Although this remains foreseeable within the ambit of scientists to rectify and it is believed that this can occur. Time will tell. What may initially look doubtful may well in time become mainstream practise.

4.2 Mathematical foundation

Prelude

Much of early psychological work was conducted within the confines of calculability and predicated on pillars of mathematical and statistical solidity.⁹ It thus assumed great prominence within the broader field of natural philosophy, especially with the pioneering works within psychophysics, decision-theory, game theory, learning and information processing and fields dominated by the propositions of mathematical psychology (Coombs, Dawes & Tversky, 1970; Dalmedico, 1997; Estes, 1975; Luce, 1997, 1999; Ratcliff, 1998; Reber & Reber, 2001). It is within the concept of science that mathematics originated (Brouwer, 1983a; Russell, 1983) and functioned as “the handmaiden of scientific enquiry” in the nineteenth century (Maddox, 1998, p.9). As such, a link between chapters 3 and 4 is forged. However, the area of mathematics was not always conceived of as abstract and academic. Until the seventeenth century, it was considered as little more than a mechanical trade serving only to aid other trades (Uglow, 2003). This is very much the same historical path followed by measurement the history of which is not traced as originating with mathematics but within trade and construction (Wright, 1997). Mathematical models have been employed within psychological research since the mid-nineteenth century and have continued unabated since, finding for itself particularly fertile grounds in application since the 1950’s (Coombs, Dawes & Tversky, 1970). Measurement theory necessitates some sort of reality (the psychological construct for instance) and the requisite mathematics (IRT or CTT modelling) to describe the reality or a model assumed to reflect the empirical situation (Wille, 1994). Gould (1998) maintains that the fascination humans have for numerical regularity is closely tied to our propensity to dichotomise objects in nature (see chapter 3) but that in certain instances this need for numerical patterning has resulted in “our overinterpretations, run[ing] so far beyond what nature could possibly exemplify, that we can only postulate some inherent mental bias as a driving force” (p.36). The enterprise of logic and mathematics, is of course, one of the more aloof disciplines calling itself a science as it focuses in on itself more often than appealing to outside issues (Quine, 1983b).

The link between Abraham Wald’s decision theory,¹⁰ Von Neumann’s game theory¹¹ and the early work of Neyman and Pearson (Dalmedico, 1997; Kline, 2004) frames early to mid-twentieth century mathematics, statistics, science and social science studies. The axiomatic scientific method in which one proceeds from an undefined but accepted truth can be linked to the axiomatised system developed for mathematics. Probability (Boyer & Merzbach, 1991) in turn is utilised by the social sciences and psychology in how deductive thought envelops reasoning and how logic can be utilised to rationalise arguments and “prove” certain conclusions (reference here is made to the logicist school in mathematics). Mathematician Andrey Kolmogorov maintained that probability as a mathematical discipline should be derivable from axioms just as algebra and geometry were derived from axioms (<http://en.wikipedia.org/wiki/Kolmogorov>). In fact he did just this; axiomatising probability and thus satisfying Hilbert’s call for the axiomatisation of mathematics (Boyer, 1991; Grätzer, 2002), the discussion of which is to follow. Theories of probability have played major roles within the psychometric discipline ranging from Bernoulli, Poisson and Bayes to more modern-day renditions of probability such as those encountered within item response theories which will be discussed below. The role played by probability within the natural sciences differs however where error has often been ignored or if corrected has subsequently utilised alternative models (Borsboom, 2005). Issues of change, probability and error preoccupy social scientists - perhaps we should reconsider our position on these aspects? The connection between “truth” and “proof” is, however, a loose one (Benacerraf, 1983b). The mathematical intricacies utilised within probability models for instance are a point in case as probability theory is widely employed as manner of inference within psychological research and without such models “it would not be possible to measure the reliability of inferences and decisions” (Stigler, 1999, p.2).

Mathematics was considered the pinnacle of scientific thought during the early twentieth century (Dalmedico, 1997), yet mathematics’ edifice was starting to crack as a number of fundamental issues were being grappled with, some more successfully than others (Coveney & Highfield, 1995). Logic, as with mathematics, can be divided into three main periods of development, from classic Aristotelian logic through two thousand years till the algebraic or symbolic period beginning with Boole and ending with Hilbert and culminating with the metamathematical or modern period exemplified by Gödel (Rucker, 1987). The misperception on the part of psychologists and measurement specialists within psychology as to the sound edifice of mathematics is a gross misunderstanding on their part. Not only can foundational mathematics said to be grappling with a number of its own issues, but the wholesale transference to a realm in which is not always suited is tantamount to academic

⁹ Itself a questionable notion as mathematics like any other formal discipline has its fair share of detractors opting for a more relativist notion of what it means to engage in mathematical science. Ethnomathematics, for instance, is an area devoted to the understanding of mathematical tendencies throughout cultures and throughout the ages (Ernest, 1994)

¹⁰ Any one of a suite of theories seeking to describe for a system of decision making, how in fact decisions are made. These theories include information pertaining to probabilities, mathematical approaches based on game theory and probability theory as well as more subjective aspects such as attitudes and beliefs (Reber & Reber, 2001). Wald’s work, among others has led to broader acceptance of Bayesian probability (http://en.wikipedia.org/wiki/Bayesian_probability). (See below for more on Bayes).

¹¹ Perhaps the most commonly known instance of game theory to psychologists would be the prisoner’s dilemma. Game theory seeks to describe a system for the making of moves and decisions within simple to more complex games (Reber & Reber, 2001). Game theory postulates have been applied to areas such as interpersonal interactions, economics and international affairs.

sacrilege. This has resulted in another perception, espoused by Schönemann (1994) for instance where “mathematics seems to have been singularly ineffective in the social sciences” (p.151). Measurement theory (see below) is largely predicated on axioms and like mathematics, proceeds from such, as yet, improvable givens. Aesthetically pleasing axiomatised measurement systems, says Schönemann (1994), are what we are now left with. It is worthwhile focusing attention on Schönemann’s (1994) list of reasons as to why mathematics seems to be ineffective in rendering itself a tool to the social sciences:¹²

- A sound body of factual knowledge was built up within physics long before axioms were constructed as point of departure. In other words proceeding from axioms to observations is not always the method followed within some areas of physics
- Only in a very select number of physics and mathematics fields has post-hoc axiomatisation been achieved
- Some areas within physics have proved quite successful despite having no axioms with which to guide them
- Scientific progress is more often made via induction based on observation and has not proceeded from deductions from axioms
- Induction is not a formal mechanical process but takes into consideration much else besides a narrowly construed set of determinations and lastly
- Science has often accepted results which have been contrary to traditional beliefs

Mathematics cannot necessarily be construed as the equivalent of the study of numbers as is evidenced within the area of geometry for instance and to say that mathematics is only quantification would be false (Russell, 1983). However one cannot ignore the measurement link between mathematics and its import into the realm of psychological assessment. Logical and mathematical “truth” such as it may be, cannot be equated with certainty, as Mill argued, such truths are merely based on a very large number of instances from which inductive inferences can be concluded (Ayer, 1983). What may seem to be logical, given the extremely large base of data on which to formulate such logical statements, is once again subject to finitist strategies to which humans are limited and included in this is the rate of computational power increase in the future. Once again there is no proven algorithm which would state whether such a system would maintain its logicism right up to the infinite end.¹³ The answer may well lie within the Platonist (theoretical) realm, as yet inaccessible. Mill’s view is synonymous with the fact that mathematical certainty can only be achieved via sensory observations (Lehman, 1979) and is severely contrasted to those propounding an *a priori* approach to mathematical certainty and truth.

On the other hand, hypothesis testing and the experimental method so much adhered to within psychology is tentatively related to the levels of chance and probability of certain outcomes and traces its roots to early experimental statistics. The latter was a branch of mathematics not given as much regard as the more abstract mathematical subjects then under study during the early twentieth century (Dalmedico, 1997) but was nevertheless seen as a branch of applied mathematics (Fisher, 1958). This section is not concerned so much with the fact of mathematics per se, but more so with its application within psychology and assessment which, when considering the overwhelming influence of statistics and measurement in psychology, pales into insignificance as a contributor to a perhaps ill-directed trajectory that the discipline is traversing. Measurement theory assumes statistical viability and validity and statistical techniques consequently have as their base of manipulation basic arithmetical procedures. This really is “the science that elaborates the abstract structures that all progressions have in common merely in virtue of being progressions” (Benacerraf, 1983a, p.291). Before measurement theory and statistical practice is looked at within intelligence research and allied areas, the mathematical underpinnings of these manipulations must be more closely scrutinised. In order to query the foundations of mathematics it is necessary to review the philosophy behind these foundations, an exercise some (Putnam, 1983a) might argue to be unnecessary. Unnecessary when one comes to identifying the course of mathematics and its practical application yes, but such an inward turn to basic epistemology and ontology is necessary in this particular instance as mathematical theories carry with them ontological assumptions (Lehman, 1979).

The enterprise of mathematics, as vast as it is, can trace its course back to the notion of numeracy and its roots are derived from the seemingly simple task of counting¹⁴ and the necessity of measuring land, counting produce and people (Boyer & Merzbach 1991; Bronowski, 1974; Clapham, 1996; Eves & Newsom, 1965; Ifrah, 2001; Livio, 2003; Martí, 1996; Moring, 2002; Omnès, 2005; Pascoe, 1992; Porter, 1997; Roberts, 1995; Sardar, Ravetz & Van Loon, 2000; Stewart, 1995, 1998; Williams, 1997). Counting of frequencies is one of the most utilised practices in psychological quantification (Michell, 1999) and precedes measurement (Wright, 1999). Mathematics’ role is paramount in measurement and although measurement did not originate with mathematics, it has given to the discipline of measurement its tools and thus “provide(s) the ultimate foundation for better

¹² Schönemann (1994) is also at pains to question the utility of mathematics in the natural sciences as well!

¹³ These are metaphysical arguments and one can safely proceed with work in the real world as we know it. Yet these types of questions always remain lurking in the background.

¹⁴ As history illustrates, what is seemingly simple today was in the past considered unfathomable and the strides made were immense at the time. How could anatomists fail to see that blood circulated in the body for instance (Zimmer, 2005)? Of course many factors play in on any one situation within historical contexts, contexts which one is not always privy to and hindsight is in fact quite conceited (Dawkins, 2005). How could we fail to negotiate our way around the issues of assessment?



practice and the final logic by which useful measurement evolves and thrives" (Wright, 1999, p.73). As is to be expected, geometry and trigonometry, albeit not in any formally disciplined activity as such and having been utilised by the Mesopotamians, Egyptians and Greeks, preceded algebra by a good two thousand years. This method was originally employed by the Islamic countries (Moring, 2002; Russell, 1983). The rudiments of counting, order and magnitude are not unique to human beings (Boyer & Merzbach, 1991) and just as there is speculation as to a language centre in the brain so too is there debate circling the issue of a similar innate mathematical centre in the brain or "number module" (Butterworth, 2000) or at the very least innate neural circuitry already present in infants (Livio, 2003), a speculative adaptive mechanism (Demetriou & Valanides, 1998).

Our mathematical rendering of reality could be construed as the equivalent of how we think (Heyting, 1983b).¹⁵ Counting and the use of mathematical concepts can also be framed within the cultural context in which they occur (Ernest, 1994; Hersh, 1994) but that it is an exclusively human innate propensity can be argued. The need to propose order on a system from without is a reflection of an almost obsessive need to arrange, order, manage and make manifest the underlying natures of things and as Brouwer succinctly states "the results of counting and measuring take so important a place, that a large number of natural laws introduced by science treat only of the mutual relations between the results of counting and measuring" (1983, p.77). Is this then not the *prima facie* case for psychological assessment? The need to propose order by fiat of measuring and counting? Is there perhaps not an alternative yet equally scientifically feasible manner in which to propose order?¹⁶

Regarding the twentieth century, the major highlights within mathematics are perhaps the most daunting in terms of their overwhelmingly large influence within areas other than mathematics. Events prior to 1931 were an effort to mechanise the reasoning processes involved in mathematics (Hofstadter, 1980). France, Germany, Britain and Italy were the foremost mathematical countries toward the end of the nineteenth century having been displaced somewhat by the United States during the latter half of the twentieth century, predominantly due to emigration during the second World War (Dalmedico, 1997). Prominent mathematicians during the late nineteenth and early twentieth centuries included Henri Poincaré (1854-1912) who, among many other contributions, introduced the concept of the group and utilised analogies emanating from non-Euclidian geometry as well as applying qualitative techniques to celestial mechanics (after work pioneered by Newton, Lagrange and Laplace¹⁷) and laid the foundations for topology (Levenson, 1997; Murray, 2004). David Hilbert (1862-1943) who, like Poincaré, was wide-ranging in various mathematical areas saw among other things the formation of a systematic axiomatic method for mathematics in 1899 (Chaitin, 2006; Coveney & Highfield, 1995; Dalmedico, 1997; Hofstadter, 1980; Maddox, 1998; Rucker, 1987; Stewart & Golubitsky, 1993). He did this by deriving a formal axiomatic model for Euclid's geometry and hoped to create for mathematics "certitude" of its methods (Hilbert, 1983; Murray, 2004) and created the new discipline of metamathematics or the theory of proof which was of course to be reanalysed by Kurt Gödel (Maddox, 1998).

Hilbert's foundationalist approaches towards mathematics, which melded in well with the positivism then reigning (and can therefore be contrasted with the Platonist foundation) (Kreisel, 1983), resulted in the point of departure for modern algebra as well as a school of logic and was firmly entrenched in set theory and axiomatics and algebra.¹⁸ This resulted in various offspring among them empirical formalism whose dictates maintain that an objective subject matter does exist for the enterprise of mathematics stripped of all but the most rudimentary philosophy (Curry, 1983). Chaitin (2006) however offers a convincing counter-argument to the "fact" that mathematics is concerned with empiricism and states that, like physics, mathematics can be classified as quasi-empirical and that in order for mathematics not to become isolationist, the discipline should make axiomatic leaps whether or not it is open to provability. The notion of mathematics as abstract science which can relinquish the system of empirical checks is the most often presented front (Ellis, 1966; Woods, 2003) for its propositions are presented *a priori* and changes in the physical world have no bearing on its abstract nature in the ethereal realm; they are timeless propositions (Ellis, 1966). Whereas Euclid's work was constructed, Hilbert's axioms existed from the start (Bernays, 1983). Hilbert's work spread into areas as diverse as relativity, quantum mechanics, matrix algebra, group theory and theoretical physics (Dalmedico, 1997). What was at the time considered an almost perverse turn-around for mathematics came in the form of an affront to codified Euclidian geometry and Aristotelian syllogisms as many centuries were to pass before a renewed look at these axiomatic foundations took place (Hofstadter, 1980; Kreisel, 1983). The idea of non-Euclidian geometry was a shock to many as it highlighted the fact that mathematics as a tool was not only a utility for real world studies but was also a tool for more abstract and esoteric areas of concern; all encompassed within what was once considered reality; the true version of reality. The Kantian notion of intuitive derivation of reality was thus turned on its head (Eves & Newsom, 1965; Mays, 1966). Along with the discovery

¹⁵ In keeping with the intuitionists' approaches towards mathematics (see below) mathematics is our mathematics, a squid's mathematics is theirs and a Zoggian's mathematics is theirs too (wherever and whenever Zoggians's happen to be). Which is the true mathematics? Is there a true mathematics? These are arguments and questions often raised in relation to the Platonists. Yet again, we are confronted with the eternal dichotomy, forever standing in the way of thought.

¹⁶ Are our brains so hard-wired that to conceive of such an equally valid proposal lies forever beyond us?

¹⁷ Himself playing a role in the development of population statistics and thus one of many forerunners of statistics in the social sciences (Lazersfeld, 1977).

¹⁸ The mathematics taught in high school can be largely attributed to the Hilbert school of foundationalist mathematics.

of non-Euclidian geometry, the recognition of the existence of algebraic structure played a role in further cementing the development and acceptance of the axiomatic method in mathematical research (Eves & Newsom, 1965). What reality was at this time was now called into question. Along with an effort to formalise mathematics and to establish for it logical and axiomatic foundations, flaws were evidenced and although not destroying the neat foundations which had been laid, forever altered the course of how mathematics was to be thought about. Three major “crises” in mathematical foundations (or at least the philosophical interpretations emanating from these supposed crises) can be highlighted: the discovery of non-Euclidian geometry, the non-existence of a consistency proof for mathematics and no universally agreed upon solution to various opposing views within set theory (Putnam, 1983a).

4.2.1 The philosophical implications of mathematics¹⁹

In order to critically appraise the influence that mathematics has had on psychology and assessment in particular, philosophical issues, once again, need to be turned to. This ever-present philosophical concern which has permeated the discussions so far, attest to the need for a “return to roots” within psychological assessment. As with any topic within philosophy, mathematical philosophy is no less a daunting area of research and debates often overlap the boundaries between theoretical physics and mathematics. The application of mathematics to various areas of physics was doubtless necessary for the advancement of physics and in order to proceed with these areas of investigation, the “language of mathematics” had to be written and systematised as with any discipline calling itself a science (Gribbin, 2003). The universe was, as Galileo stated, written in the language of mathematics (Damasio, 2003) and physics is currently written in and described most accurately by mathematical language (Gardner, 2003), a language which is perhaps the most sophisticated of the sciences (Bronowski, 1974).²⁰ However, this summarised discussion will merely touch on the philosophical issues pertinent to mathematics as it pertains to fundamental issues of assessment and what it is means to measure.

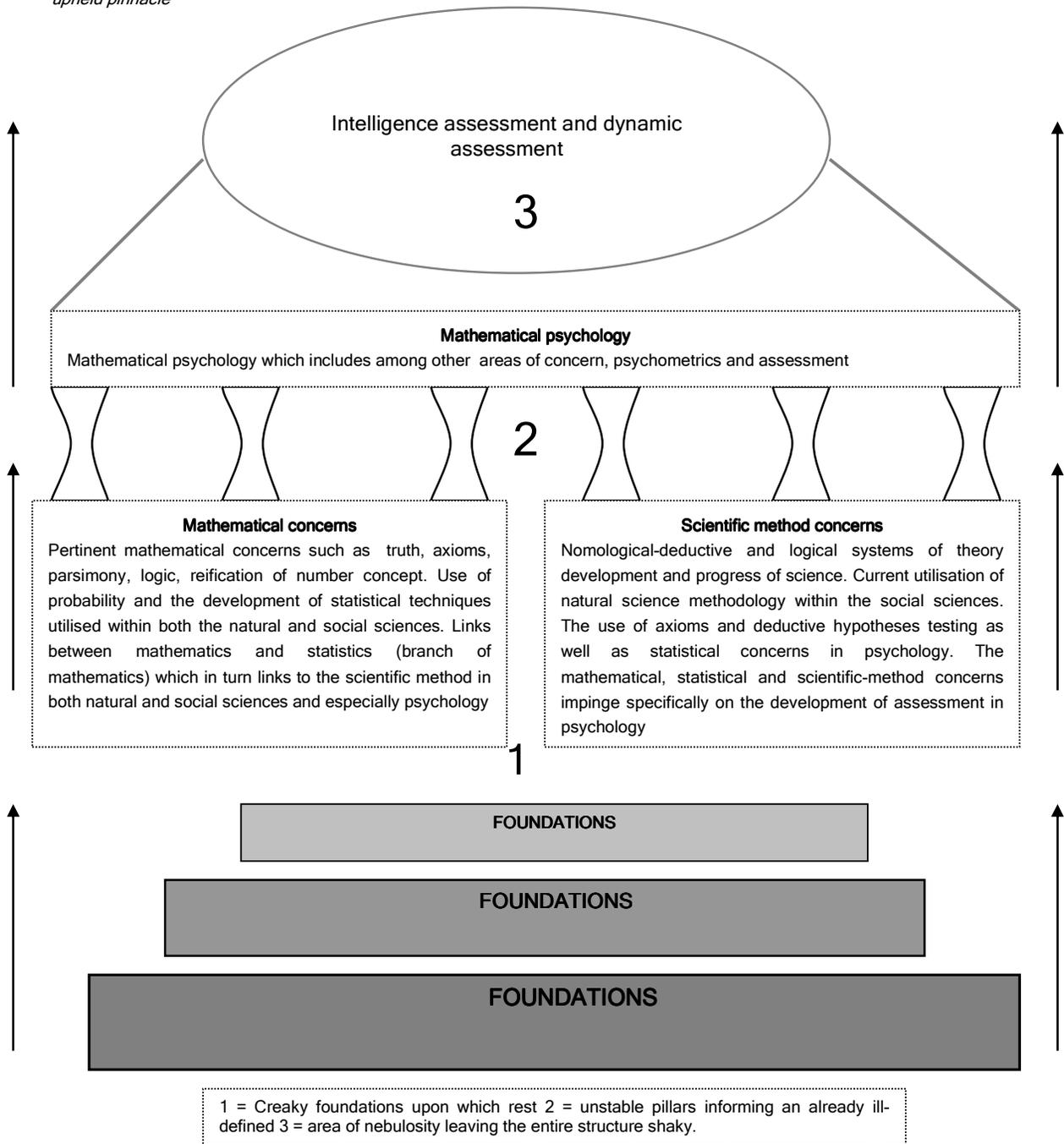
Broadly speaking, the study of mathematics has highlighted the role played by philosophical inclinations and how proponents of various schools of thought have in fact conceptualised the place of mathematics and the type of answers it can expect from its varied areas of endeavour. As with prior discussions on consciousness, nature/nurture, intelligence and realism/relativism; epistemology and ontology are core to these various mathematical affiliations; what is the area of mathematics? What can be studied in a mathematical manner? What is “true” in a mathematical conjecture and how can something be proved conclusively? The notion of truth in mathematical reasoning especially, dates back to the early Greeks (Penrose, 1999). What is or is not amenable to mathematical rendering? Are there limits to which mathematics can aim? What is the proper area of mathematical concern and most fundamental of all what is mathematics? These are of central concern to mathematical philosophers which might have an unnerving effect on mathematical psychologists for instance, for they are the utilisers of a method which is itself still at odds as to its own nature. Mathematical psychologists along with statisticians employ sophisticated techniques in order to deploy such methods on the terrain of psychological assessment, an area itself replete with conjectures and lack of finality. Assessment then, is itself utilised in an area even more undefined and nebulous, namely intelligence research. In sum, the foundation structures are creaky, the pillars are unstable and the ideas upheld by these shaky foundations are not really understood.²¹ Figure 57 details the overlapping areas of concern which is shared by philosophical mathematics and the scientific method and how they play forth into the area of psychological assessment concerns.

¹⁹ Mathematics as formal discipline carries on just as surely as if there were no philosophical issues entangled in its history and has applications in everyday life, the likes of which we take for granted. So although the thesis posited here is that a necessary re-evaluation of fundamentals needs to take place, one also cannot deny that the practice is alive and well and it is often the case that many mathematicians may themselves have no particular viewpoint on the matter of foundational philosophies (Penrose, 1994).

²⁰ Although he refers to humans’ need to impose order on the world in the context of the study of calendrics, Gould (1998) nevertheless cautions the reader not to be “oversold on nature’s mathematical regularity”. The image of mathematics being the universal language is to be tempered by what, why and how it is utilised by humans.

²¹ How disconcerting then to think that most lay-people (those in fact seeking to benefit from assessment) are the most likely to think that mathematics is a “pure” discipline, that assessment is based on hardy and robust mathematical psychological techniques and that intelligence is a defined construct. None of which is true. This is not denying the incredible strides made in any one area, however the rate at which these areas are touted as being valid and reliable is misleading and at times unethical.

Figure 57 Three realms of shared concern; the base and pillars of which may not be as entirely pure as seemingly implied in the upheld pinnacle



Immanuel Kant's idea of the mathematical *a priori* was not entirely in keeping with Plato's (perhaps the first theoretician; Ayer, 1983; Livio, 2003) ideal realm (theoretical realm?) in which all things in their true and pure forms existed. Although one can see the allure and value of Plato's reasoning (Penrose, 1994, 2005; Scruton, 2004). Mathematical objects as such do not exist (Plato advocated that they did); in other words there is no transcendental realm in which mathematical objects are situated; rather our experience of empirical phenomena is grounded within the framework of *a priori* mathematical understandings (Kant) (Iffrah, 2001; Scruton, 2004).²² Mathematics conveys "rigorous truths which reason is capable of discovering without need of experiment and yet truths which may always be confirmed by experiment within such broad limits as experimentation requires" (Cournot in Iffrah, 2001, p.355). This is a fundamental issue worth pursuing here as it highlights the place that mathematical reasoning inhabits. The appealing idea behind Plato's transcendental realm is that truth or proof exists to prove mathematical entities but which are forever independent of our knowing them directly (Livio, 2003; Scruton, 2004). David Hilbert, though not subscribing in full to the Platonist ideal of separate existing realm of number can be crudely considered a Platonist (Scruton, 2004). Those favouring the idea that proof was all there was and nothing beyond it were more inclined to think about the mathematical realm as existing only as and in proof itself and hence it was constructed via the mechanism of proof (Livio, 2003; Scruton, 2004). This is in keeping with the ideal as depicted by Kant. However problems soon emanate from such a position due mainly to the idea from logicism which, in this instance would maintain that a confirmatory proof or proof negating a logical claim is all that could be afforded within the proof where surely confirmation or negation would be necessitated. The situation unravels when neither proof is the offering. This would lead to a system which is either meaningless or neither true nor false.

The three most widely agreed upon modern philosophical foundations to the study of mathematics are the logicist foundation exemplified by the British philosopher mathematician pair Bertrand Russell (1872-1970) and Alfred Whitehead (1861-1974), the mostly French intuitionist (or constructivist) foundation led originally by the Dutch mathematician Luitzen Brouwer²³ (1881-1966) and the mostly German formalist foundation developed principally by German mathematician David Hilbert (Benacerraf & Putnam, 1983; Brouwer, 1983a; Clapham, 1996; Curry, 1983; Dummet, 1983; Ernest, 1998; Eves & Newsom, 1965; Fuchs, 1967; Lehman, 1979; Penrose, 1999; Wang, 1974). There are of course numerous arguments against the absolutist vision of any the three foundationist schools (Ernest, 1998) but their supremacy, value and role within the enterprise of mathematics cannot be understated. Mathematical exactitude reside on paper for the formalist, in the mental realm for the intuitionist and in the independent theoretical realm for the Platonist (Brouwer, 1983a, 1983b).

Regarding the origins of the logicist movement there are numerous mathematicians and their works which converge in on the logicist foundation such as Gottlob Frege's (1848-1925) first proposed system of the derivation of mathematics from a logical foundation (1884, 1893, 1903) who in turn had been influenced by Leibniz and Peano's attempts at application of flexible symbolism within mathematics (Ernest, 1998; Gödel, 1983a; Harnish, 2002; Woods, 2003).²⁴ This preceded Russell and Whitehead's independent opus on the very same issue and Frege took as support of this view Hume's assertion of the analyticity of mathematics (Carnap, 1983a; Frege, 1983, Mays, 1966; Scruton, 2004). At the time though, Frege contended that mathematics and logic were indistinct,²⁵ but this was largely ignored until the arrival of Russell and Whitehead's treatise (Boyer & Merzbach, 1991),²⁶ and even in Frege's description of the concept of number he was able to derive a definition utilising letters of the alphabet (Frege, 1983). Leibniz maintained the notion of logic as a science and the actual reduction of mathematics to logic was maintained by Dedekind and Frege. Mathematical theorems were later stated in logical symbols by Peano (Eves & Newsom, 1965). Russell and Whitehead's treatise utilised Georg Cantor's (1845-1918) idea of a one-to-one correspondence illustrating the idea of equinumerosity²⁷ (Cantor established set theory and developed the idea of infinite sets among other accomplishments) (Boyer & Merzbach, 1991; Gödel, 1983b; Scruton, 2004; Stewart, 1990; Wang, 1983), the work of whom was considered as sublime by Hilbert. Together, the ideas of Cantor, Hilbert and Poincaré made manifest the centrality of proof (Penrose, 1999) within mathematics.

²² Although Kant's theory was largely overshadowed by the discovery of the exact feature which had made Kant's case: the axiom of parallel lines which had been turned in on itself to reveal another axiom in keeping with non-Euclidian geometry (Brouwer, 1983a). Ayer (1983) maintains that by stating that something is *a priori* is to level at it the notion of its being tautologous.

²³ This mode of reasoning goes back as far as Descartes and Kant and Brouwer and his followers can be considered as modern proponents of this school. Similarly, Aristotle can be viewed as the originator of the logicist approach (Beth & Piaget, 1966; Eves & Newsom, 1965).

²⁴ Frege maintained that Socratic rhetoric was just for show or "colour and shading" and that each linguistic sentence could in reality be reasserted as objective content (Ernest, 1998). When one thinks about this, it is not difficult to understand how and why Frege may have come to such a conclusion. One can, after all, proceed in error in such rhetoric and come out looking positively victorious. Perhaps Frege had in mind a way in which logical deduction would be the ultimate arbitrator in such cases. Looking at arguments logically and then reducing them to symbolic notation is reminiscent of the hypothesis procedure within psychology. The link between such endeavours becomes ever more manifest. It would seem that quantitative psychological methodology has quite a history behind it, one that is perhaps not often thought of.

²⁵ Until Frege's as well as Russell and Whitehead's works, mathematics had traditionally been aligned with science and logic aligned with Greek (Russell, 1983).

²⁶ Incidentally Gottfried Leibniz (1646-1716) had already employed logical means of proofs but had erred in this regard.

²⁷ "The concept *F* is equinumerous with the concept *G*" and in so doing Frege was able to show, in a logically argued case, that "we have reduced one-to-one correlations to purely logical terms and can now offer [this] definition" (Frege, 1983, p.143). Is this a foreshadowing of rules of correspondence utilised within representationalist measurement?

The logicist school did not promulgate the foundations of mathematics as consisting entirely of logical predicates, but reduced mathematics to logic in addition to theories of sets and properties (Benacerraf & Putnam, 1983; Quine, 1983a). In so doing the bridge between mathematics and logic was highlighted; a gulf which had until the works of Frege, Russell and Whitehead's treatise been considered unbridgeable. Although it often appears that the logicist school did in fact streamline mathematics to logical predicates (Carnap, 1983a). The main differentiating feature between mathematics and logic is the finite predicates of the latter and infinite predicates of the former (Heyting, 1983b). The Italian mathematician and leader of the modern Italian formalist school Giuseppe Peano (1858-1932) whose symbolic logic had as a consequence the development of important notation (Brouwer, 1983a; Clapham, 1996) was also responsible for the axioms of integers and his work towards the axiomatisation of mathematics in general (Eves & Newsom, 1965) and all that was left now was to define the fundamental concepts utilised within these axioms (Russell, 1983; Scruton, 2004).²⁸

The logicists' entrenchment of logic as foundation of mathematics was flawed, however, due primarily to the fact that set membership proved to be the foundation of mathematics as well and not only logic which had among other repercussions certain insoluble paradoxes, the most well-known of these paradoxes known as Russell's paradox. Such seemingly self-referential paradoxes (there were other paradoxes) were the irritations for which attempts at banishment culminated in Russell and Whitehead's *Principia Mathematica* (Hofstadter, 1980). Russell's paradox had been anticipated as a paradox by Cantor (Penrose, 1999), an attempt for which resolutions were offered (Clapham, 1996; Scruton, 2004), which leads on to the set theory. The logician Frege's original ideas in set theory (1884) in part inspired the works of two mathematicians who helped lay down seven axioms for set theory and which are known today as Zermelo-Fraenkel set theory (after Ernst Zermelo; 1871-1953 and his contemporary Abraham Fraenkel; 1891-1965). These axioms are not logical truths but rest on intuitive foundations in which sets themselves are the primitives of mathematics which is *a priori* (Scruton, 2004). Set theory posits that numbers themselves are sets (Carnap, 1983b) in contradiction to the nominalist idea of number, where numbers were not objective things but only meaningful in terms of what they expressed or the context in which the numbers were placed (Minsky, 1988). The number four "4" is not anything in particular other than a representation of the set of "fourness", this conception of a set of four is really a Platonist idealisation of what four really is or at least what it means (Parsons, 1983; Wang, 1983). The set containing the set of empty elements is itself a set of one set and hence is not zero but one (Penrose, 1999). This particular rendition of number is of course just one such derivation for the concept of number (another scheme includes Alonzo Church's lambda calculus and there is also another scheme which envisages numbers intuitively without the need to capitulate to the concept of set at all; Benacerraf, 1983a). Frege's (1983) definition of number was that, as an "objective object" it only made sense once it was contextualised within a sentence of some meaningful setting (Ernest, 1998), a sentiment echoed by Russell (1983). He added to it by emphasising the human tendency to continually define primitives in terms of yet more primitives and so on ad infinitum, but knowing that our capabilities are finite we are unable to continue upon a path of infinite definitions and so stop at the most logical point, hence the emphasis on logicist leanings. An interesting feature within measurement theory in psychology is the use of intensive measures as extensive measures via conjoint measurement. In a manner mathematics' primitives find a measurement theory counterpart in non-derived extensive measures.

Scruton (2004) furthermore adds that one of the set theory foundation axioms, namely the "foundation axiom" states that there are no ungrounded sets²⁹ which Scruton levels as a partial attempt at the solution of Hempel's paradox of confirmation, so named because there is nothing illogical about the manner in which laws are confirmed. That is, laws can be inductively confirmed by instances in which something is said to be as well as something which is said not to be.³⁰ Von Neuman, as mentioned above, also provided an alternative to the Zermelo-Fraenkel set theory and so numerous were the branches of mathematics and applied mathematics that Von Neuman and Norbert Wiener (1894-1964) were both involved in many varied aspects of it continuous growth, which was starting to overlap more and more into the social sciences (Boyer & Merzbach, 1991). Hilbert's formalist school of mathematics (programme) as well as the logicist school could not hold up to scrutiny and perhaps the largest blow to the formalist school was Gödel's incompleteness theorem (Gardner, 2003; Penrose, 1999).

In essence, Hilbert's formalist agenda was to locate for any area of mathematics a number of axioms with defined rules and procedures allowing any defined reasoning in that area to be incorporated. Any mathematical proposition in this system would be consistent and complete and established for non-finitary mathematics a finitary construction (Von Neumann, 1983). What is

²⁸ Five axioms from which all arithmetic can be derived. The first three axioms present three primitives (or numbers) and so allows the fifth postulate to prove theorems about all numbers by considering only these three (Hempel, 1983). This led to the logicists' attempts to define the three primitives and to thus illustrate that the postulates are derivable via logic from the definitions, which is precisely what Frege and Russell independently set out to do utilising Cantor's one to one correspondence idea (Scruton, 2004).

²⁹ Which means that there are no sets which contain members which contain members which contain members and so on ad infinitum.

³⁰ To make this clearer: "all post-boxes are red" would be inductively confirmed if every street corner evidenced a post-box which was red. The logical equivalent to "all post-boxes are red" is a sentence stating "all non-post-boxes are not red". But this leads to an absurd state of affairs for one could quite happily spend the rest of one's life confirming that post-boxes are red by making completely inane comments. To get rid of this annoyance within set-theory then, the foundation axiom states that no such statements would be allowed. Here then is a link between the foundations of mathematics and the logic of scientific method (Scruton, 2004) which was discussed at length in chapter 3.

formalistic about this agenda is the fact that any prepositional meaning per se is irrelevant as it merely represents symbols.³¹ Gödel showed that no specifically defined proposition (or its negation) could be proved within any formal system³² and hence the notion of complete truth as understood by the formalist school is in fact incomplete. The formalist approach towards mathematics was anathema to the view endorsed by more Platonic-minded mathematicians in which, as stated above, an ethereal realm of mathematics existed (Penrose, 1999) (although in a manner of speaking, one can support the view that this be viewed as the realm of theory and was perhaps just framed in this manner by Plato and his cave depiction). Penrose (1994), as a self-confessed Platonist, highlights the need to understand the implications of Gödel's theorem (Gödel himself a Platonist), which does not result in “unknowable” truths lurking forever out of the grasp of human thought, but that solutions to problems can be known, they are simply not dependent on formal rules and contingent axioms thought up by human beings (or any other organism for that matter) (Coveney & Highfield, 1995; Ernest, 1998).³³ In other words, there does exist a Platonic realm of sorts, not a physical place, but rather a state of truth or knowingness for which solutions for insoluble and intractable problems do exist but for which solutions await discovery (Benacerraf & Putnam, 1983). The Platonic realm makes its presence felt within measurement theory more specifically within classical test theory's positing of a true score, which as a notion or idea is yet to be fully understood within its own theory (Borsboom, 2005; Borsboom & Mellenbergh, 2002). Is the true score in actual fact the construct score? The issue of hypothetical and substantive construct via its numericised score is a recurrent theme within this thesis and is found repeatedly in themes discussed in various dynamic assessment models in chapter 5 (sections 5.2.1 and 5.2.9).

This is in contradistinction to the idea of mathematics conceptualised by the constructivists (intuitionists) in which mathematics is a creative activity for any undecidable proposition, in terms of human understanding of rules governing any proposition. Its truth too is undetermined and because there is nothing else which is relevant to the issue, truth is simply not available in the current system and hence there is no theoretical realm in which the truth resides. Rules and systems and any truths governing these systems are bound to time and place and there is no truth independent of human thought; no transcendental realm (Heyting, 1983a). The Platonists disagree and maintain that any correspondence between propositions and their mathematical counterparts are timeless and not context-bound. In other words just because something happens to be insoluble during a specific time and place it does not necessarily mean that a solution does not exist (Benacerraf & Putnam, 1983).³⁴ Could one not equate the constructivist programme with a relativist one and the Platonist programme with an objective one? In fact the similarities and contrasts between the mathematical foundationalists and early philosophers of science hinge around very much the same type of issue. Namely the level of reality to be studied and the realm of the unknowable (empiricist vs. Platonist mathematicians or philosophers, which in similar vein to the behaviourists in psychology, allowed no place for the existence of abstract entities, i.e. sense data was the only viable type of data) (Ayer, 1983; Carnap, 1983b). In a sense then, the affiliation towards the Platonist way of viewing mathematical reality could be construed as being at odds with the positivist programme in general and thus inconsistent with the affiliations stated in chapters 2 and 3. How is the author to defend this inconsistency?

Other than offering an account of attitudes towards the reality of mathematical entities on the one hand and looking at positivist tendencies within the global practice of science on the other these views are perhaps inconsistent. Others would disagree with the very need for the foundations of mathematics to be studied in the first place, as there really is no need for it as the discipline can get along without it all the same (Bernays, 1983; Putnam, 1983a). Putnam (1983b) maintains that a balance be struck between metaphysical realism (intuitively grasping at the Platonic realm) and scientific verificationism (empiricism) in an attempt to “solve” issues between varying philosophical positions within foundational mathematics. Hence, if this is construed as inconsistent then it either needs more thought or is trifling enough not to be of concern. Figure 58 illustrates the interwovenness of the three realms and figure 59 illustrates the case with the real life construct “IQ”.

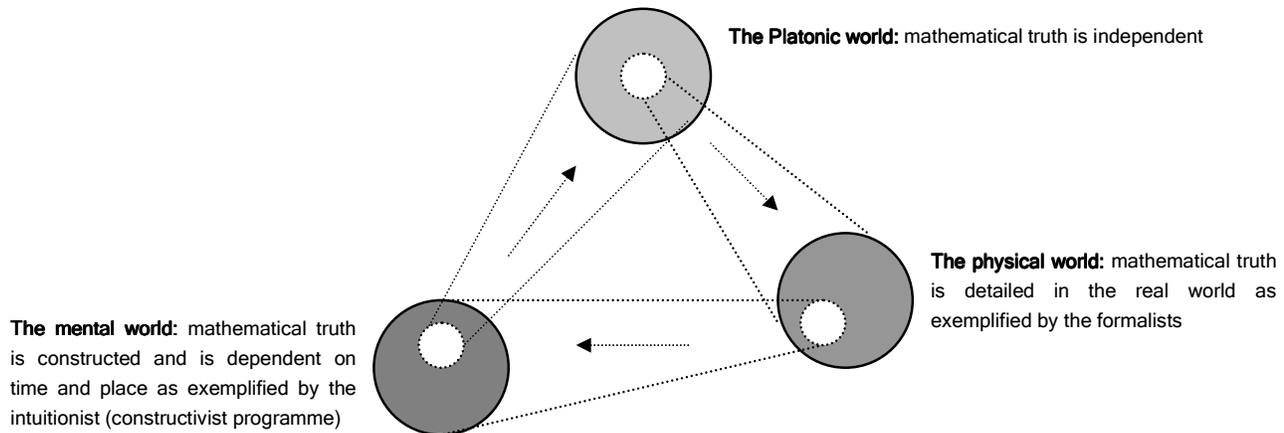
³¹ This idea is reminiscent of the idea behind Ramsey sentences (see chapter 3) in which redundant terms are done away with in order for theories to proceed unencumbered. This is also reminiscent of Frege's attempt at logical sentences stripped of all but the bare logical essentials.

³² The author is reluctant to enter into mathematical details, the meaning of which will by and large be non-sensical to a non-mathematician, such as the author! Penrose (1994) is of the opinion that this is just lazy and one way of taking the easy way out. This is conceded.

³³ The Platonic concept makes its appearance again in the section discussing measurement below.

³⁴ This Platonist argument can be taken further. There may very well exist solutions to problems about which humans are not cognisant. There could be potentially other types of space-time continuums of which we are simply unaware for which there exists mathematical uncertainties about which we are unaware and for which there exist solutions, to which we are also unaware. So, in a manner, the Platonist “take” on mathematical certainty and philosophy is quite compelling. The constructivist rendition of mathematical solubility, it would seem, is too dependent on time and place. Once again a relativism of sorts raises its head.

Figure 58 The relation between various worlds and how each world takes a small part of its predecessor with it into the next world with attenuations as to the mathematical foundations (Penrose, 1994, p.414)



The Platonic realm includes proofs and solutions (as well as proofs to the contrary) about any known as well as unknowable mathematical topics, even those not humanly knowable for they exist independently of our knowing them or even of our not knowing them. In fact it is entirely plausible that if any such truth were evidenced we would not necessarily be capable of recognising this truth as having obtained “or of getting ourselves into a position in which we can so recognise it” (Dummet, 1983, p.105). Yet the mental world can imagine a good deal that was once unknowable and now knowable but can only begin to fathom the depths which the Platonic or theoretical realm encompasses. The physical world is a given and the mental is to a large extent dependable and an extension of this physical world, yet conjectures in the mental world can play back forth into the physical world as is evidenced with theoretical physics. The idea or notion of a Platonic world is really just a metaphor to aid in our understanding that neither the formalist nor intuitionist renditions of what mathematics is, can come to any sort of “truth” when confined to the definitions of their own respective programmes.

Paradoxically within empirical science, mathematical truth is not as easily proven as other disciplines within science have shown yet its application within empirical science is without question (Hempel, 1983). Traditionally the route followed by any science in determining for it the truth or falsity of an hypothesis was to prove the validity or lack thereof of the hypothesis, this was discussed at length in chapter 3. Mathematically speaking though, the parallel to verificationism within the scientific method is the method of proof (Gardner, 2003; Putnam, 1983b). Mathematics as empirical science however does not avail itself of quite the same characteristic. Empirically validating that $6 + 2 = 8$ is a complicated affair if the task is to be undertaken in a similar manner as utilised in physics for instance. Stating an hypothesis such as $6 + 2 = 8$ and relying on empirical observations to confirm such a relation is almost impossible. The nature of the *a priori* in mathematics can be accepted on the basis of a tacit understanding of the meaning behind what is attributed to the concepts of the numbers “6” and “2” and “8” as well as the operator “+”. What is necessitated by these figures is clearly understood by those utilising them.³⁵ Students will understand that when given the figure 6 it will be necessary to add to this another figure, namely, 2 and to describe the resultant relation emanating from having put the two together. This is usually taken for granted, but the implicit understanding can be explicitly complex, especially if proofs have to worked out. The price that is paid for utilising this *a priori* analytical truth is that the statement says nothing about facts. It is not a factual system at least not in the real world of factual information. We do not inhabit the Platonic realm and do not possess all theoretical truths (Hempel, 1983). Numbers are abstract “measures” as stated at the start of this chapter (Ifrah, 2001); counting live stock, measuring the angles of structures and employing techniques to aid in the construction of buildings as well as understanding the movement of stars in the sky.

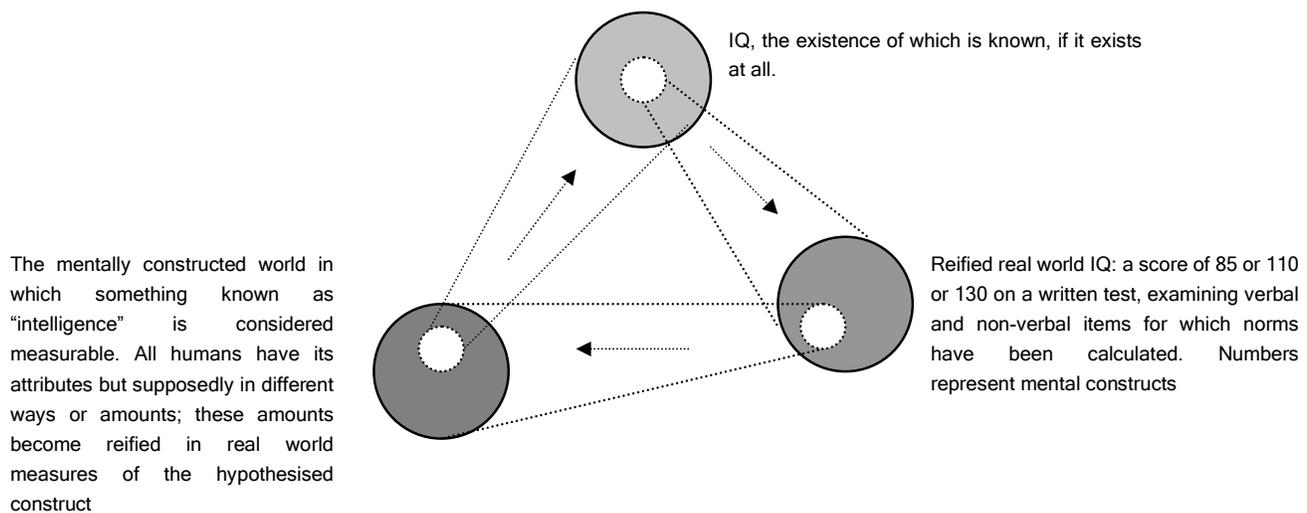
The move from measurement of objective things (lines) or entities (angles in a triangle) to the measurement of unknowable entities (intelligence) is a move perhaps not deeply considered as problematic. Things are attributed meaning by means of words which are later substituted by numbers representing these words, which are in turn substituted by more abstract symbols until such time that symbols are employed for regions within the mathematical realm for which no conceivable physicality exists (Ifrah, 2001). The level of abstraction reached within any area of mathematics does not impel the need to study mathematics for the sole sake of enhancing the mind, as was stated by Karl Jacobi but because mathematics can aid in the understanding of the world in which we live (Fuchs, 1967). That it has proven to be precisely this is telling of modern conceptions of the interplay

³⁵ The closest the author has come to describing the strange nature of symbols is to refer to Stewart’s (1995) analogy in which he states that “mathematicians are forced to resort to written symbols and pictures to describe their world - even to each other. But the symbols are no more that world than musical notation is music” (p.ix).

between measurement and psyche. “ $6 + 2 = 8$ ” is hanging up in the air somewhere, lodged in the ethereal realm of truths yet to be discovered; the statement is not anything in particular, is neither exists here nor there; it is simply a statement describing the relations between two things. What things? It is meaningless to ask. It is in fact quite amusing to ponder on this very issue for it is something we take for granted in our everyday dealings with mathematical and numerical entities. It is very abstract and extremely vexing as a discipline with which to try to come to grips. Have psychologists misapplied the services that mathematics has to offer or have they misinterpreted the results from mathematical manipulations of psychological data? The error does not lie with the enterprise of mathematics but with the skewed interpretation by psychologists’ own understandings of measurement.

The assessment of intelligence from neuroscience applications (measurement of neural conductivity) is one method of assessing what is purported to be intelligence; the assessment of intelligence by measuring the results on a test supposedly assessing for intelligence is another method of assessment; statistically deriving results from sets of answers on a test is yet another method of adjudicating the level of intelligence. Psychology has employed all these methods in what many consider to be successful methods. The interwovenness of biological understandings of the brain with the measurement accuracies from mathematics and the statistical refining of results has jettisoned the enterprise of assessment into a new era. Yet we are no further along the path of what it is that constitutes intelligence in the first place. Are the applications at fault? Surely not. Are the techniques at fault? Surely not. Then what is the problem? As Hempel (1983, p.291) puts it: “in the establishment of empirical knowledge, mathematics has the function of a theoretical juice extractor: the techniques of mathematical and logical theory can produce no more juice of factual information than is contained in the assumptions to which they are applied”. One can likewise extrapolate these sentiments to the arena of assessment.

Figure 59 The relation between the three worlds utilising IQ as reified yet ill-defined measurable construct



Similar sentiments govern figure 59 as they do figure 58. Although not intended by Penrose (1994) to be utilised as a description of psychological reality, the case for assessment rests firmly on the assumptions predicated by mathematical measurability and reification. The intuitionist school (which broke upon the mathematical scene in 1924) can be considered a break away from formalism in which the rules governing the system of sets and membership of such sets are more emphasised than the sets themselves which are not necessarily believed to be in existence as such (in terms of the Platonic meaning of the entity existing). Followers of this intuitionism (entitled such due to its supposed mirroring of the human mind; Penrose, 1999) could trace their philosophical roots back to Aristotle who had deviated somewhat from the ideals proposed by Plato. Aristotelian syllogisms cannot however serve as foundation for the enterprise of mathematical science because it does not provide a complete analysis of reasoning within mathematics (Beth & Piaget, 1966; Poincaré, 1983). A mathematical idea was said to exist only once the mental counterpart of the same idea became manifest, hence one would construct the idea which led then to the alternative label for this approach.

One of the main contentions of the intuitionist school was the rejection of the Aristotelian law of the excluded middle³⁶ which briefly states that the negation of something is logically equivalent to the affirmation of its opposite which forms the basis of the

³⁶ Perceived by Hilbert as a heresy of sorts (Coveney & Highfield, 1995).



reductio ad absurdum rule often deployed in pointing out egregious errors of the logical type;³⁷ something either is or is not (Boolos, 1983; Coveney & Highfield, 1995). This school also denied the existence of a certainty until such said certainty could be proven to be true or false. For example, until the proof to Fermat's last theorem was concluded the issue of the theorem being true or not was denied an existence, thus making mathematics contingent upon time and place; a dependence which some found illogical and unappealing (Penrose, 1999). One can clearly see the stark disparity between the intuitionist and Platonic modes of reckoning about the reality of mathematical entities.³⁸ Aristotelian conceptions of methodology emanated from three postulates; deduction, self-evidence and reality. The third postulate is particularly emphasised with regard to the opposing intuitive view as espoused by Descartes and Kant where there is a discernable shift from the domain of the real to the domain of the constructed (Mays, 1966).

4.2.2 The link made manifest

Mathematics^{39a} (*verb*) the study of number, form, arrangement, and associated relationships, using rigorously defined literal, numerical, and operational symbols'' (Reader's Digest Dictionary, 1988). Such is the unpacked definition of what it means when one practises mathematics of sorts. Notwithstanding the enormous range over which this definition currently extends, for these purposes, this definition is ripe with key words pertinent to psychological assessment. How successful such a venture has been within the domain of psychology is questionable (in terms of the use of number, form and arrangement in reference to non-numerical aspects such as IQ, colour and liking/disliking of a painting for instance) and as dynamic assessment very much forms part of the repertoire of assessment tools, it too is not exempt from criticisms levelled at it.

The reality of an object and its counterpart or shadow in the mathematical realm is easily stated in mathematical terms, such that two apples as readily seen by the eye, is rendered mathematical as $2x$ (x representing an apple). The method utilised to establish this transfer from the real physical realm to one of corresponding mathematical reality is deceiving at first glance, as the transfer back into the realm of the physical from one of mathematics cannot always proceed as easily (if at all) (Brown, 2001a). Measuring an object (an apple) and measuring its properties (colour or taste) are not synonymous concepts nor ventures; likewise measuring a rank ordered statement is not necessarily the same as measuring its properties (what the statement is saying, or what it means to a person). Using the aforementioned example of rank-ordered statements: quantifying *rank* (its nominal order) and quantifying *meaning* (falling outside Steven's original conceptions of measurement scales) reside within very different realms; realms which are either not delineated correctly or adequately enough, or, which are simply not taken cognisance of at the outset. Are psychological assessors to abandon all known methods of assessment which fall within the realm of the physical as exemplified in the scales of measurement? And if this is done, what exactly is going to replace this method? Or is the mere fact that such a question need be asked already indicative of our reliance on an outmoded appraisal of what it means to assess within the social sciences and specifically within psychology and dynamic assessment?

Dynamic assessment is currently and has for many years been burdened with the unresolved issue of not being able to bring meaning to quantifiable change (quantifiable as understood within the physical scales of measurement) and numerous attempts at ordering or bringing meaning to such change has merely resulted in the re-issue of the same physical mode of quantification to the problem. Various angles of explanation are considered when trying to explain away what it means to have undergone change (typically from a pretest to posttest scenario) but the angle of explanation is not what should be highlighted. It is the fact that any angle of tentative explanation is proceeding from an incorrect stance from the outset. The application of tools within the physical reality's realm is simply not applicable to change occurring in a totally unique realm (the above-mentioned differences between physical and mathematical reality and how they are measured). Phenomenological tools of change assessment are themselves inadequate in discriminating the nature of change, for although not as sharply described as are more rigorous tools (techniques having as their foundations the scales of measurement) phenomenological tools are themselves apt to lean towards meanings residing on scales of sorts (although they can never be charged with such explicit liaison with scales of measurement as can more quantifiable tools).

³⁷ One can see a thread linking the logical/mathematical with the varied philosophies of scientific method, proof, argumentation, deduction and how something is considered to be true or not. The link then, although tenuous upon first observation has turned out to be more than just tentative. For what is considered logically deductive (logic and mathematics) can be transplanted into the realm of science (scientific method) and transferred almost wholesale into the social sciences where as its last stop it is considered pure, complete and valid. Three notions, which as it has turned out, are not quite what they seem to be. The establishment is currently saddled with a monumental historical edifice of repute, yes, but ill-defined and uncomfortably positioned within psychological assessment. The first thing to do then, in an attempt to re-evaluate this issue, is as mentioned a number of times already, to revert back to the origins of assessment. This thread is convoluted, knotty and spread thinly in some regions.

³⁸ And the entire edifice of mathematical psychology is predicated upon such a lower order edifice. Granted, once the ascent towards assessment is begun, such shaky foundations and their consequences become less marked. What is the point then of illustrating mathematics' less than pure foundations? It is inconceivable to the author at any rate that these consequences would not have some sort of ramifications somewhere at the top.

³⁹ The assertion may well be made that language too is a notational system representing reality as it is filtered through our brains and if this is the case mathematics is merely another notational system. However, a possible retort to this would be that just as with language, mathematics too is limited in its representational capacity to reflect a true reality (no subtle leanings are made towards relativism here).



Logical positivists and empiricists would not condone forays into metaphysical realms, realms which cannot withstand the scrutiny of physical reality as transposed into mathematical reality (yet it is undeniable that many metaphysical research programmes did just this; and which have subsequently led to many discoveries later on such as the germ theory of disease, the wave-particle nature of light and so on; Nickles, 2001). Blending methodology from neuroscience into psychological assessment requires that use be made of the real and physical realm and so instantiating the need for scales of measurement. Is this argument/debate thus rendered null and void? Perhaps the right questions are not being asked or at the very least are being ill-phrased. A crude analogy can for the moment be employed to further examine the type of frustration being dealt with here: it is often stated that “words are not enough” by which one can extract the meaning that language is incomplete or too limited in its vocabulary and/or manner of expression to do justice to what is being felt by the person stating it. Akin to this analogy is also the disputatious contention that thinking cannot co-occur without language (cognisance is taken of instinctive actions which are behavioural manifestations of hard-wired actions and intentions).⁴⁰

Just as charged particles were originally considered mathematical entities and not field points, so too can IQ points be considered mathematical entities but not real “brain stuff” (Brown, 2001a), for intelligence is not IQ and IQ is not intelligence (Grigorenko, 2004a). In fact one is far closer by stating that an IQ score may represent phenotypic values and not, as envisioned in the early twentieth century, as a genetic value. Emanating from incorrect conclusions from heritability studies the perception that this paradigm “measures” “intelligence” is incorrect and in keeping with the theoretical-substantive criterion or differentiation, “the heritability paradigm refers only to studies decomposing observed phenotypic variance in *human traits* into its *estimable* (not measured!) components” (original emphasis) (Grigorenko, 2004a, p.54). The fact that atoms are today discernable⁴¹ as physical objects (Shao-Horn, Croguennec, Delmas, Nelson & O’Keefe, 2003) does not necessarily mean that such a path is in store for the ever-elusive IQ score. Two major philosophical issues which impinge heavily on change measurement within dynamic assessment’s pre-posttest score scenario concern the nature of *representation* and the nature of *empiricism* (observable and non-observable entities) an issue touched on in chapter 2.

Concerning *representation*: When objects are represented mathematically, the representation may extend only as far as the object itself and not to its properties as discussed above. Is an item on a psychometric test measuring the item’s “objectness” or is it measuring what it represents? Numerical assignment may imply one of many things, among others;

- quantitative reality as is seen in real life scenarios, for instance five key strokes */////* may reflect the number of sticks on the ground taking on a structural feel to what it means to represent a number (Brown, 2001a, p.263); a more or less one-to-one correspondence with reality. Here the mathematics is descriptive of reality
- quantitative description of the reality, for instance “the green book has 232 sentences” here describes what colour the book is as well as the numerical representation of the number of sentences; nowhere does “232” appear in the book, it merely represents a framework for interpreting the properties of the book (Brown, 2001a);⁴²a second-order correspondence with reality; here the mathematics is representative of reality

Concerning *empiricism*: This attends to what is and is not directly observable⁴³ (Trout, 2001a). The weight of an object, the elasticity of an object and the particular phase in which a material may happen to be are all directly observable, measurable and quantifiable. Any mathematical modelling imposed on these objects which seek to determine behaviour may proceed along the lines of description. The change in attitude towards the liking or disliking of a presidential candidate is readily quantifiable in terms of a first-order level descriptive exercise. However, the second-order representational mode⁴⁴ of quantification is elusive. Quantifying the answers to questions concerning a political candidate (when a scale of one to five is used for instance) is straightforward. The crucial and most often ignored part of this exercise is the transference of a behavioural concept (in this instance attitude) to a number. A behavioural phenomena and a numeric quantity are not synonymous. The fact that these two “notions” were ever brought together has lead to the particular situation with which researchers are currently saddled. The original and at-the-time inspiring idea to numerically represent on any scale of measurement any behavioural aspect (Francis Galton largely initiating such an enterprise) has perhaps long since outgrown its usefulness and should have matured beyond a

⁴⁰ The word ‘intention’ however is laden with meanings of forethought, for to be intent upon something, one can surely not be thoughtless for it is an intent that something occurred as planned.

⁴¹ Although the answer as to what it is that brings mass to particles (the elusive Higgs particle or the god particle) will have to wait for results from experiments conducted with the Large Hadron Collider (Muir, 2005).

⁴² Just as numerics represent abstract or real objects, varying radix systems of numerics may too well be used. For instance the hexadecimal numeric for the decimal number 232, is equal to E8. The point emphasised here is that the actual numeric representation may be arbitrary, the fact however that it imbues specific meaning is not quite as trifling a matter.

⁴³ Note that atoms are not directly observable by the human eye, but of course this does not mean that they are not directly observable; they are however *not* inferred and this is the distinguishing feature between observable and non-observables.

⁴⁴ Halford and Wilson (1980) elaborate on what they refer to as second-order isomorphism which is the representation of representations. Here the link with the discussion on cognition in chapter 2 is highlighted. The problems concerning representation pervade both the fields of measurement theory (see section on measurement theory below) and cognition alike. They are both “concerned with providing a valid model of the world, or of some segment of it which happens to be important for a particular purpose” (p.359).

simple one-to-one correspondence of behaviour to number.⁴⁵ This quantification quandary of twentieth century positivism was a direct spin-off from logical positivism evidenced during the early twentieth century (Mouton, 1993a).

The issue of concern here is not the need or lack thereof for mathematics (for mathematics' role in aiding communication or in fact *being* the communication channel within the realms of physics for instance is not in question) but its use within a different realm. This is the psychological realm which does not facilitate the use of representation *as well as* the realm of physics which is an issue that needs to be addressed. That it *is* used and *can* be used is not in question; that it can be better used is however debateable. The origins of the practise of "psychophysics" as it was known in the late nineteenth century were firmly entrenched in the mathematical norms of the day in which physics and other aligned natural sciences had found a rich repository (although itself still developing) of techniques deemed helpful in elucidating physical concepts. Why not impose upon the behavioural realm such a vast array of techniques which would no doubt aid in securing for psychology a firmer and more acceptable platform in accordance with the strict objectivity of the day? (Savage & Ehrlich, 1992). This is indeed what played forth in the development of psychophysical measurement (Gould, 1997a; Leahy, 2000; Nunnally, 1978; Sahakian, 1981).⁴⁶ Even if it was conceded that psychological constructs be measured on the continuum of nominal-ratio scales, it still stands that many behavioural sciences concepts do not avail of ratio scale attributes (Smit, 1996) nor interval scale attributes although test developers strive to attain such levels, interval level measures often yield scores with ordinal properties⁴⁷ (Cliff, 1991b; Murphy & Davidshofer, 1998; Rigdon, 1998). Rasch (1980) was keenly aware of this when he stated that the ordering of responses on a test does not mean that abilities are being measured on a ratio scale.⁴⁸ Person ability and item difficulty are conjointly solved allowing for ratio-level analysis, hence a stochastic measurement model attempting to avoid the "foolish practice" of ambiguous and contradictory results (Wright, 1997a, p.114).

In his original probabilistic model, Rasch stipulates that ability and difficulty are solved simultaneously and that the probability is described as a ratio of both properties and not either on its own. They are thus calibrated on a common scale (Woodcock, 1999). In fact Rasch (1980) utilises an apt example from physics to make clear his exposition on this point. Maxwell's analysis of the relation between force and mass/acceleration is studied and compared to parameters of a psychological model (Rasch, 1980). In ascertaining the law of force being equivalent to the product of mass and acceleration, the idea of force is formalised simultaneously (mass x acceleration).⁴⁹ Parameters follow laws, not the observed entities (behaviour) (Rasch, 1980). Lidz (2003) makes use of an example from age equivalence scores stating that these types of scores are neither ratio nor interval and that blanket comparisons of such scores across ages cannot be done, very much the old adage of a difference of 6 months at age 3 is not quite the equivalent of this same difference at age 18. Added to this is the crucial difference between an object and its properties which seems to have been denied, overlooked or simply not considered an issue worth much attention otherwise early proponents of psychological measurement would not have been so easily lead to assume that all is measurable. The current critical theme is not one of deciding the truth or falsity of intensive or extensive⁵⁰ scales of measurement but whether measurement using these scales is indeed applicable at all. Notwithstanding the fact that the entire edifice of psychological measurement is now predicated on these scales of measurement (Kerlinger, 1981), it is contended that this fact is now playing a role in the stale-mate reached within dynamic assessment research.

4.2.3 Summary

The enterprise of mathematical discourse is a flourishing one and is not touted here as being in any manner responsible for what may at times be considered misguided efforts on the psychological measurement front. However, the need to impress upon psychological measurement researchers the foundationalist concerns within mathematics is considered timely. As may often be the case, mathematics is viewed with awe and due acknowledgement of its seemingly unbiased, logical and abstract nature and the whole machinery is often viewed as pristine. In many instances this view is correctly espoused especially in areas which are ideally suited to its methods and results such as theoretical physics. Applied in various real-world contexts it remains, doubtless, unsurpassed in its abilities to bring answers to questions heretofore unanswerable. The context of psychological assessment is an area of concern within the larger arena of subjects to which mathematics can be appointed as problem-solver. The point of concern does not, however, lie with the mathematical method but with how the method is utilised and deployed within the field of human assessment. The foundations of mathematics are riddled with puzzles and conundrums and as with many disciplines it is

⁴⁵ Or "numerals" if nominalists are to be kept happy.

⁴⁶ This is not to say that quantification of behavioural phenomena cannot edify to a certain extent when practised in other contexts less influential in terms of affecting people's lives; such as the quantified approach that Murray (2004) utilises in his recent tome on "Human Accomplishment". Here, quantification of attitudinal and other qualities may not seem as out of step as at first glance may appear.

⁴⁷ Presumably due to the paucity of ordinal statistics and an ordinal-based psychometrics (Cliff, 1991b).

⁴⁸ There are models though which question the need to be so reliant on interval and ratio-scaled measures (Chin, 1998).

⁴⁹ Note that the extensive measures of force and mass are themselves compound ratios (Asimov, 1993).

⁵⁰ Extensive measures (interval/ratio level measure) concatenate units of measures and thus make possible assertions the nature of which include statements such as "*a* is twice the length of *b*"; whereas intensive measures (ordinal level measure but excluding categorical or nominal measures) simply state that "*a* is larger than *b*". Exclusion of intensive measures resulted from the belief that only empirical concatenation resulted from measurable entities or objects (Savage & Ehrlich, 1992).



still to resolve a number of key issues. The veneer of sublime abstraction is a surface which can be easily scratched to reveal cracks and creaks of its own. Mathematics' creaks of course, is not the main consideration within this argument though but plays forth into how it is accepted within other related areas of concern. The last bastion of purity (as is often considered by naïve social scientists) is not as pure and unadulterated as may originally be the perceived case and the maintenance of such a view will only hamper any progress to be made within the quantified fields in social science disciplines. The next section discusses the role of statistics within assessment.

Much early pioneering work within psychology was made functional due to certain mathematically informed predicates upheld by psychophysicists in the fields of vision, learning and information processing. Mathematical psychology espoused certainty and accuracy to a defined level not often seen elsewhere within the psychological domain. Undoubtedly this lent credence to the discipline in the guise of its adorned scientific mantle. Mathematics as a formal discipline had been couched within science and grew markedly in the eighteenth to early twentieth centuries. Coincident to this development was the parallel development of many formal scientific disciplines largely confined to the scientific framework in which theory developed according to certain notions of accepted scientific practice. Psychology as burgeoning science was among these formal developments and so partook in the many accepted notions of the day of what was meant by scientific progress and development on certain fronts. The importation of mathematical rigour was to be expected. That the utilisation of this "pure" technique was considered to be relatively faultless (in comparison to many other disciplines) attested to its acceptance within social domains. It is hardly surprising then that mathematical psychology played forth into various early sub-disciplines within psychology.

The development of formal systems of theory development can be considered as going hand-in-hand with the subsequent development and solidifying system of rigorous proof within mathematics. In an attempt to codify the enterprise of mathematics, strict formalism was seen to be the answer to the proper and objective progress of mathematical science. It is not surprising that the laying down of formal tenets for the development of scientific disciplines was also echoed in the similar progression within mathematics and in onto the area of psychology, specifically psychological assessment. However, this pure and abstract nature of mathematics was to be dealt severe blows in the early twentieth century in which it was evidenced that mathematics, like any formal discipline, has its own share of shortcomings and unresolved issues. None has been more striking than the crushing deflation of the formal axiomatised system of mathematics in the mid-twentieth century. The direct link between mathematics and assessment is not an obvious one but an established link is evident between the two realms and it is considered an important link due to the problematic issues with which psychological assessment is currently saddled. More pertinent to this study's contention is the philosophy undergirding mathematical discourse. Ideas stretch as far back as Plato, Aristotle, Euclid and much later through to Kant and although considerably revised, much of the core concepts still remain within current debates. The link between positivism in science practice and logical positivists can clearly be seen. The co-evolution or co-development of both the enterprise of science and mathematics has perhaps been underplayed by historians of science but becomes manifest and is jettisoned to the foreground when viewing the historical development of the psychological discipline as science. It is not surprising then, that mathematical certainty has worked its way into the efforts and areas of interest which currently hold sway within the social sciences.

The lack of the basis of certainty, the calling into question of the foundationalist assumptions of mathematics and the flagrant attack on certain basic assumptions within mathematics has had repercussions within the discipline, perhaps more so on a philosophical than applied front. The basis upon which measurement resides is then not as solid as once supposed. Are psychologists aware of these issues? Can anything be done to amend the present trajectory of psychological assessment which is predicated on outmoded notions of accuracy and solidity? This brief discussion on mathematics as one element within assessment has attempted to bring to the fore critical issues pertinent to the future development of assessment. By viewing and understanding the background of one of its bases, perhaps assessment can re-assess its course to come.

4.3 Statistical foundation

Prelude

The brief discussion on the reality of pure mathematics ties in closely with the similarly brief rendezvous with statistical endeavours in psychological assessment. Social science's rationalisation of the continued utilisation of statistical techniques in order to buffer and support notions of scientific credibility may start to wear thin if the promised goods are simply not being delivered (accurate identification of potential and prediction of success within psychological testing). As was the case with the view taken at the outset on mathematics above, it cannot be denied that the multitude of statistical techniques developed specifically for use within psychology have been works of wonderment. As with the development of mathematics within science so too did social and behavioural statistics develop within a scientific framework (Stigler, 1999) once again underlining the importance of a need to view the development of science within the social sphere (see chapter 3). How far can quantitative techniques propel psychological science though? It is perhaps not so much that the techniques are in any way flawed as much as it is the application of the techniques which are brought into question (Brown, Hendrix & Williams, 2003; Chow, 1998b; Huberty, 1993). Brown, Hendrix and Williams (2003) question the utility of inference within psychology and ponder the resultant

use of descriptive statistics as sole statistical agent. The difference between statistical hypotheses and the related theoretical hypotheses is also an area which causes confusion among researchers attempting to “prove” theories (Abelson, 1997b) based on the significance of (i) the statistical hypothesis and (ii) maintaining that the result indicates future replicability as well as (iii) making claims on the theory via only one test. Dynamic assessment, as with its intelligence assessment counterpart, are dual testing strategies facing the same statistical treatment but due to dynamic assessment’s change-based philosophy and reliance on more qualitative methodology it need not necessarily maintain its position alongside mainstream intelligence assessment. Traditional statistical inferential testing did not set about to prove the probability of obtaining a result as much as it provided a method for ruling out chance (Borsboom, Mellenbergh & Van Heerden, 2003; Harlow, 1997; Howell, 1987). It set out to detect enough evidence to suggest that the effect being tested for is in the direction hypothesised (Harris, 1997). The main contention within the social science literature as it pertains to the utilisation of statistics, is that hypothesis testing has in some measurable way become synonymous with statistics (Nester, 1996). This trend is hardly surprising given the amount of attention that is paid to significance testing and subsequent hypothesis generating research designs within psychology. The major focus of this perennial interest and debate is the emphasis placed on hypothesis testing and not the mathematical statistical argumentation behind numerous and decidedly beneficial statistical techniques. As Nester (1996) states in connection with the continued flawed use of hypothesis testing by professional statisticians “continued association with hypothesis testing is not in our own best interest” (p.408).

As argued in chapter 2, the need for cross-disciplinary approaches towards the study of intelligence may be one of a few feasible methods to be pursued in the twenty-first century. A similar situation regarding the increased use and sophistication of robust statistical and methodological techniques in the area of intelligence and potential might also boost increased access to information about this constructs. Chapter 3 highlighted the trajectory followed by science and social science theory within a scientific framework and when considering the edifice upon which the whole enterprise of science is built, it is hardly worth contesting the role played by statistics and statistical methodology in pushing psychological assessment even further to the fronts of knowledge acquisition. However, the misperception that significance testing is the only manner of hypothesis testing (or Fisher’s equating of null hypothesis significance testing with scientific hypothesis testing) and moreover that it is the only true way in which a science can progress, is erroneous (Schmidt & Hunter, 1997). There are other methods (discussed below) which can determine significance of hypotheses other than the traditionally accepted one of null hypothesis significance testing.

When utilised with sound understanding and judgement, many statistical techniques are decidedly helpful within the social science (Abelson, 1997a, 1997b; Chow, 1998b; Harlow, 1997; Nickerson, 2000) and will most likely be around for some time to come (Huysamen, 2005). Why then the need to assess its role in this section? The main reason is its present incompatibility with certain perceptions of what dynamic assessment is and how this manner of assessment seeks to assess change. Various attenuations of statistical methods exist to address these issues (see the section on measurement below) but the nagging question of the need to statistically quantify change persists, not just in the area of dynamic assessment but also within the social sciences as a whole. “It seems safe to assume that many [research psychologists have not had] a lot of exposure to the mathematics on which NHST [null hypothesis significance testing] is built” (Nickerson, 2000, p.246), a sentiment shared by Estes (1997). The general lack-lustre repertoire of superficial robust results within the social sciences can be traced back to the pressure placed upon these subjects to do justice to their existence as independent valid and reliable disciplines. The blame of NHST proliferation has been laid squarely on the “pernicious” hypothetico-deductive method of scientific inference (Rozeboom, 1997, p.336). What role has and does social science statistics play in assessment within the social sciences? This issue is now looked at.

4.3.1 Statistical issues

The core issue relating to the statistical foundation level of this analysis as pertaining to psychological assessment and experiment is the two-fold problem of statistical significance as exemplified by the p -value approach of Ronald Fisher (1890-1962) and hypothesis testing as exemplified by the fixed-alpha level approach of Jerzy Neyman (1894-1981) and Egon Pearson (1895-1980) (Huberty, 1993). Also included is the possibly underutilised and undervalued approach of probability theories, among them Bayesian probability, first introduced in the social and behavioural sciences in 1963 (Rupp, Dey & Zumbo, 2004), which is a culmination of other probability methods and techniques among them the binomial probability distribution, after Jacob Bernoulli who applied the binomial distribution as an inverse probability for understanding the observed event⁵¹ (Wright, 1997b). Bernoulli’s 1713 treatise on the theory of probability was the first comprehensive account of probability (Boyer, 1991). Inverse probability allowed raw observations to be cast as consequence of the particular stochastic process within stable formulation. But in order to make the step from raw observation to inference one needs to identify the underlying stochastic process through which an inverse probability can be defined and Bernoulli’s distribution was the easiest and most widely available tool to do just this (Wright, 1997b). Simeon Poisson’s distribution, like Bernoulli’s is also a discrete distribution but the outcome is not limited to

⁵¹ The Bernoulli distribution is a discrete probability distribution which assumes a value of 1 (success) with a concomitant probability p and a value of 0 (failure) with its probability $q = 1 - p$ (http://en.wikipedia.org/wiki/Bernoulli_distribution; Everitt & Wykes, 1999).



a choice of 1 or 0; moreover it is a limiting instance of the binomial distribution (Boyer, 1991). Poisson's compound distribution is the "parent of all useful measuring distributions" (Wright, 1997b, p.34) and the natural parameter for the Poisson distribution is a ratio of an object's location and the measurement unit of the instrument used. This formulation preserves concatenation and divisibility resulting in different units implying the same location (Wright, 1997b). As will be discussed below, Fisher took another step in 1920 by developing the likelihood version of inverse probability in an attempt to construct maximum likelihood estimation (utilised within IRT models) (Wright, 1997b). Is it interesting to note the development within probability statistics throughout the seventeenth to twenty-first centuries and how researchers have plied their tools upon the trade making it feasible to use numbers in a manner aiding inference. Nevertheless, continuous debates ranging over four decades within the psychological literature pertaining to statistical inference⁵² and hypotheses testing led to the resultant Task Force on Statistical Inference (TFSI) being instituted by the Board of Scientific Affairs. This was convened under the American Psychological Association (APA) to deliver a framework offering guidelines as how best to pursue various contentious issues within statistical psychological reports (Krantz, 1999; Wilkinson & TFSI, 1999). At the outset, it must be made clear that the historical development of statistical techniques within the social sciences is not a straightforward history and that issues such as the utilisation of hypothesis and significance testing is a perennial debate. Dynamic assessment, as scientific method of change assessment is lodged between various approaches and philosophies towards such assessment. On the one hand a wholly qualitative and clinical approach towards change inducement and assessment is at odds with the more robust quantitative approaches of change assessment. Both seek to bring about measurable change within the individual in the assessment situation which can last from an hour to many years.

Measurable change is accounted for in terms of numerical differences between pretest and posttest scores and also by clinical assessments of individuals over a lengthy period of time. The role played by statistics within all of this can vary from purely descriptive statistics enabling the researcher to gauge the starting level of competence through to the level at which skills have either been improved or not; to inferential statistics which enable the researcher to infer and possibly predict outcomes of various interventions efforts and future performance or behaviour within any one domain or various domains. Typically, as with much of the literature on psychological assessment, research designs employ hypotheses as points of departure leveraging conclusions on design and method which can vary across studies. Questions which can be posed about dynamic assessment stretch far beyond what the method does and how it accomplishes what it sets out to achieve. In order to address the statistical influences on dynamic assessment as method of change assessment fundamental issues need to be addressed. Such core issues have received attention in chapters 2-4.

Accusations lodged at any particular approach of psychological assessment need to be very mindful of the development of the method within the greater realm of psychology, for it alone is not wholly responsible for what perhaps may seem to be misguided efforts on a number of fronts. Issues pertaining to how the brain and mind is viewed by psychology and other related disciplines has been discussed, along with issues involving the topic of environmental and genetic contributions towards behaviour. Core philosophies are what mould the approach. The approach then is housed within a context of scientific growth and change and is dominated by the settings of scientific discourse. Many issues within this context are yet to be resolved. These issues were highlighted in chapter 3. Dynamic assessment, as manner of change evaluation, is not alone in terms of being ineluctably drawn into and being overwhelmingly influenced by historical trends - it is by nature (as with many disciplines) a contingent sub-discipline. Issues that plague dynamic assessment are issues which plague other disciplines. There is no quick and simple answer as how best to approach such myriad problems. Scrapping statistical inferential techniques along with the traditional practice of hypotheses testing is hardly a path to be followed. The point of building a case in chapters 2-4 is to illustrate the entanglement of dynamic assessment as method of change evaluation and to highlight the need to pry apart the various layers of historical impingements. Dynamic assessment is firmly grounded in solid scientific technique and will not be easily extricated from its foundations.

The TFSI broached the debate on the applications of significance testing and stimulated discussion of the topic in their 1999 article in *American Psychologist*, concluding that the APA revise their statistical sections in their publication manuals (Wilkinson & TFSI, 1999). The summarised guidelines which are relevant to this section are set forth below:

- clarity of study aims is sought by refraining from "cloaking" the study in a guise that does not apply to it - are the studies hypotheses generating or hypotheses testing?
- the measurement variables described need to remain consistent throughout the study. Measurement precision is not, however, synonymous with operationalisation

⁵² The forerunner of statistical inference is the probable error (PE) of a mean which was extended from the PE of test scores and was first utilised in 1910 (Huberty, 1993) but Nester (1996) maintains that it was used as early as 1840 in the area of biology. Probability per se, though was evident in work by 1713 (Wright, 1997b). Probability's role is paramount within the social sciences specifically.

- utilise effect size estimates when reporting p values and depending on the nature of the units of measurement, the effects sizes can be either standardised or not⁵³
- interval estimates of effect sizes should be stipulated

The article, although addressing issues of concern to the psychological research community did not really offer guidelines radically different from guidelines offered in the past but concluded that “statistical methods should guide and discipline our thinking but should not determine it” (Wilkinson & TFSI, 1999, p.603). This sentiment can likewise be lauded as the prevailing sentiment regarding dynamic assessment and its acceptance within the broader assessment arena.

4.3.1.1 Historical contingency

Dynamic assessment, as with many sub-disciplines within psychological assessment and psychology as a whole, is bound by traditionally accepted modes of operation, research design and results-reporting and deviations from these norms of scientific practice may often be judged as flaws. The rigid application of NHST can be seen to be ritualistic (Huysamen, 2005; Nester, 1996) and even a bad habit (Shrout, 1997). The struggle that dynamic assessment in its clinical mode faces is one common to other areas within psychology which too are not necessarily amenable to statistical inference as this is simply not the right mode for their existence as mode of knowledge-acquisition. The prevailing context of scientific progress needs to be considered (as has been in chapter 3) as well as fundamental philosophies pervading the history and origins of certain key points of departures (as evidenced in chapter 2). Why has psychology “progressed” to this point in time in which statistical technique plays a greater role than it perhaps should? Once again, not much can be said to discredit many statistical techniques, and this has never been the aim of this defensive treatise on the subject but a common theme thus far established in this study is that the application of these tools into an area not amenable to such tools is perhaps ill-conceived.

Much of social sciences statistics emanates from issues surrounding probability and chance and until the advent of the twentieth century, there were no formally directed research endeavours to develop the methodology of significance testing beyond Laplace’s ‘probability of causes’ which relied on Bayes’ rule (Fisher, 1958; Gigerenzer, Swijtink, Porter, Daston, Beatty & Krüger, 1990; Mulaik, Raju & Harshman, 1997). Incidentally, Laplace was instrumental in rescuing the work on Bayes’ inverse probability and it is due to Laplace that the theory of probability grew as it did (Boyer, 1991). The theory of errors, unlike the probability of causes was a more developed branch of statistics (Fisher, 1958) and found its utility in the estimation of errors in areas like astronomy (Quetelet, as mentioned in chapter 2 brought over ideas from astronomy into the social sciences). The determination of errors in observations in applied areas of work as well as the minimisation of these errors played a large role in the development of the statistical techniques which psychologists currently employ (Gigerenzer et al., 1990) and many statistical techniques were originally developed for deployment within the social sciences (Porter, 1997).

The general aim of evaluating hypotheses based on data is quite an old one, is associated with the name of Bayes and dates to 1710 (Huberty, 1993; Nester, 1996; Neyman & Pearson, 1933). It was epitomised in the early twentieth century by the combined efforts of Karl Pearson, Ronald Fisher and W.S. Gosset. Fisher’s⁵⁴ work in the fields of agronomy,⁵⁵ physiology and medicine; Gosset’s work in the Guinness brewery in Dublin and Pearson’s work in biometry (and thus founding the basis of psychometrics; Porter, 1997). These efforts are telling of the need to impose order and a mechanised framework on data collected in the field (Fisher, 1958). Hypothesis testing was, however, sporadic during the nineteenth century and only came into its own after the publication of Gosset and Fisher’s various works (specifically the t test, the Chi square test and F test; exact distributions which were developed) (Lehmann, 1992). Moreover, Lehmann (1992) maintains that Fisher drew attention to hypothesis testing only as incidental. Kendall (1978) dates the utilisation of statistics, as is understood today, to 1660 and thus excludes an already lengthy history of enumeration for purposes other than for estimation and prediction; the terms in which he ensconces statistics. Kendall’s (1978) main emphasis shared by Lazarsfeld (1977) illustrates what prompted the development of statistics concerns;

⁵³ Appendix 1 utilises standardised effect sizes as the measured variables differed in most of the studies which were meta-analysed.

⁵⁴ Fisher, like Galton was a polymath and there aren’t too many subjects on which he did not lay his hands (Stigler, 1999). Fisher’s driving motivation behind some of his statistical work was eugenics, which was also Galton’s favourite. Egon Pearson, Karl’s son worked at the Galton laboratory in London. Small world indeed. Prominent figures cannot however be divorced from the times in which they lived and worked. Mendelian inheritance, evolutionary theory, the eugenics movement, the development of intelligence assessments and utilisation of newly developed statistical techniques must have been terribly exciting in the early days. Especially when pioneers started to weave together all the threads. Are there any such pioneers today? The world has become too complicated for polymaths to flourish quite as successfully as they once did. The history of intelligence testing, the progression of psychological science and statistical technique and methodology can all be traced to a number of individuals working in-and-around similar issues of the day. The Galton-Darwin-K.Pearson-Fisher-Neyman-E.Pearson link is particularly evident within this subject area. Contrast this historical line to early works within dynamic assessment (see chapter 5). No wonder such a chasm is evident; core philosophies cannot be more widely dispersed! Moreover, the basis of mathematical statistics was closely associated with research in the areas of biological evolution, heredity and the eugenics movement (Porter, 1997).

⁵⁵ No pun intended. Just as an aside, it is quite humorous reading through Fisher’s books (1956, 1958) in terms of some of the examples chosen for his explanations of statistical techniques. For instance he cites examples such as “comparison of relative growth rate of two cultures of an alga” (1958, p.140); “effect of nitrogenous fertilizers in maintaining yield”(1958, p.136) and particularly pertinent to this thesis “frequency of criminality among the twin brothers or sisters of criminals”(1958, p.94).



namely politics, which fuelled the need for arithmetic which could aid in the most judicious use of resources and as a means of controlling taxes. Huberty (1993) dates the first statistical approach to the English scholar John Arbuthnot in 1710 in which the gender ratio of births was the main concern. Stigler (1999) maintains that the history of statistics, at least since the seventeenth century, is one of a collection of ad hoc and miscellaneous mathematical tools dealing with data, the process being governed by an evolutionary pruning of sorts within the tradition of scientific enquiry.

Fisher also helped introduce the idea of the experiment which, until 1910, had not been associated with the use of statistics (Gigerenzer et al., 1990). The combination, then, of experimental design and statistical inference is a relatively new one and was of course adopted very early in the formal development of psychology as scientific discipline. As chapter 3 highlighted, the growth of psychology as science as formalised discipline is everywhere evident, none more so than in its methodology. Rigid deployment of Fisherian statistics, some might advocate, has served to hamper the development and progress of a psychology equally relevant to individuals as to groups. Early twentieth century social science statistics, it must be recalled, centered on detection of error and probability of distributions of groups; groups involving crops, diseases, measurements, stars and stout among other areas of concern.⁵⁶ The development of positivist influenced behaviourism (the philosophies of the Vienna Circle were at their heights; Oakes, 1986) with the emphasis on sensory data in the 1940-1950's; the synonymy of randomised experimentation (Fisher) with treatment and control groups and the "institutionalization of inferential statistics in American experimental psychology" (Gigerenzer, 1991, p.255; Stigler, 1999) pushed psychological testing on a course of robust scientific technique even though significance testing as technique within psychology was critiqued as early as 1955 (Schmidt & Hunter, 1997). Prior to the 1940's very few psychological articles published statistical reports (Kline, 2004). This was also the time during which Stevens published his work on the scales of measurement which have become the mainstay of measurement scaling in psychology (Stevens, 1946, 1951b). Stevens was considered a "middleman", voicing indirectly, opinions from the logical positivists, especially those of Carnap (Michell, 2003). Pure statistical techniques were never developed for application in areas as diverse as psychology, education and economics (Porter, 1997) but have inevitably been applied and misapplied in these varied contexts (Rossi, 1997). In other words these ingredients were what made and makes psychology a science. The Fisher-Neyman-Pearson school of statistics became the orthodox approach to follow (Oakes, 1986). Unfortunately not all sub-disciplines within this subject avail themselves of such mechanisation. Michell (2001) adds that "if no attempt is made to test a hypothesis, then there is no adequate scientific reason to accept it. Thus, mainstream psychologists adopted the rhetoric of measurement because of the political advantages to be gained for the discipline, not for any adequate scientific reason" (p.214). The author's already stated opinions on philosophical affiliations in chapter 2 highlighted the leaning towards positivist influenced methodology.

This may seem contrary to the critique to be levelled at null hypothesis significance testing. However, it is not the method per se that is being criticised but the incorrect utilisation of the method in an area of psychological research which does not necessarily avail itself of the method's techniques.⁵⁷ However, during the 1960's a noticeable decline in positivist strategising and the upsurge in Bayesian statistics heralded a new era in technique (Oakes, 1986) and has been utilised within the cognitive domain, where the need for less time-intensive methods calls for more efficient models (Neufeld, 2002). Strides in modern test theory were also being made during the 1950's in contrast to techniques utilised in traditional classical test theory, even though the requisite computational power was not yet at the disposal of statisticians (McDonald, 1999). As with statistics, being an applied branch of mathematics, test theory is also considered a branch of applied mathematics and for this reason the co-occurrence of trends within statistics and test theory is not surprising. Simultaneous developments in mathematical statistics flowed over into both realms.

⁵⁶ Early descriptive statistics (although not formally recognised as such) concentrated on the counting of people within countries and can be related to the need, as discussed above in the section on mathematics, to count and measure. As early as 1532 the number of deaths in London was counted on a weekly basis (Donnelly, 2004), although tallies of deaths, citizen numbers and accounts for actuarial reasons go back as far as ancient Rome (Kendall, 1978). It seems that this propensity to count, measure, statistically infer and mathematically manipulate are fundamental needs in the world in which we live. Melding the inclination to measure and count along with the rigour of science discourse makes it seem almost natural that assessment should have proceeded along the course it has; however ill-fitting this course has proven to be.

⁵⁷ Once again, the author reiterates the original stance promoted at the outset of this study. Psychology has to decide where in the spectrum of formal science it should be placed. "Soft" areas of concern are fine. "Hard" areas of concern are also fine. However, the constant overlapping of methodologies and statistical choice of data analyses is pressurising these vastly different areas of concern by moulding them into structures that are anathema to their core philosophies. Soft psychology (clinical and counselling among others) can function entirely on their own in their own methodological manner without necessitating the use of hard core psychological technique. The issues need to be separated. There is nothing wrong with such a separation but holding the whole of the discipline hostage to a specified approach is suicidal. Unification via subject matter or separatism via estrangement of method and technique are two ways in which this can be broached.

*Fisher vs. Neyman-Pearson - prelude to null hypothesis significance testing*⁵⁸

Both the Fisherian and Neyman-Pearson (after Jerzy Neyman and Egon Pearson, son of Karl Pearson)⁵⁹ schools are frequentist or classical school and the Bayesian school (after the Reverend Thomas Bayes; 1702-1761) is subjectivist and is often referred to as the common sense approach to statistics, preceding work done in classical statistics⁶⁰ (Edwards, Lindman & Savage, 1963; Harlow, 1997; Kline, 2004; Mulaik, Raju & Harshman, 1997; Pruzek, 1997). However, Reichardt and Gollob (1997) are reluctant to equate Bayes theorem with the subjectivist approach as one need not be a subjectivist to utilise Bayes. Fisher's approach is synonymous with statistical testing whereas Neyman-Pearson's approach is synonymous with hypothesis testing (Reichardt & Gollob, 1997). Crudely then, the two issues have, throughout the years, become melded into one erroneous method of statistical inferential logic. The Bayesian approach has not been as influential in social science statistics as the former, even though Karl Pearson had occasionally made use of Bayesian assumptions (Gigerenzer et al., 1990). The Neyman-Pearson school has overshadowed the Bayesians even though the latter regularly rears its head (Lindley, 1992). In essence, Bayes' theorem relies on the availability of prior distributed probabilities which conventional significance testing does not (Oakes, 1986). These priors vary over a number of hypotheses which is perceived to be its biggest flaw (DuMouchel, 1992; Moore, 1997) as initial priors are estimated,⁶¹ leading to less exact measures and based largely on belief (usually expert or well informed belief) (Kline, 2004; Neufeld, 2002; Rindskopf, 1997; Trafimow, 2003) and are thus unconditional (Killeen, 2005). The theorem is an instance of normative decision theory which necessitates base rate information which is combined in the probability of future events; failure of which leads to the use of heuristics as guiding factors⁶² (Kleiter, 1994). Due to the probable nature of Bayesian statistics, the conditional priors are set to a limit of 1. This entails the setting up of a parameter which necessitates certainty on the part of the assessor. However, being certain about such a parameter also leaves room for being wrong about it resulting in what Borsboom, Mellenbergh and Van Heerden (2003, p.209) aptly state as being very difficult "to be wrong about something if one cannot be right about it". Parameter estimation⁶³ requires a true value, one which is not available even within prior conditionals. It is perhaps prudent to think about how dynamic assessment might avail itself to the use of Bayesian statistics in a pretest-posttest scenario in which prior probabilities of success or failure or even growth are taken into account when assessing the likelihood of improvement over mediatory interventions. Oakes (1986) points out that the historical antecedent to this personal belief in priors is in fact rational belief which is measured in an objective manner. Nevertheless, this is in contrast to conditional posterior probabilities in which unknown population parameters are considered random variables as opposed to their unknowable yet fixed status in the frequentist paradigm (Rupp, Dey & Zumbo, 2004). However, as more data is gathered the original subjectivist choice of prior probability is partialled out over successive iterations and the new data serves to update prior beliefs (Pruzek, 1997; Rupp et al., 2004) or revises the rule based on new information (Majumdar, 2004; Walliser & Zwirn, 2002). Surely this can in some manner attest to a happy medium in which dynamic assessment specifically can function? Probability can be interpreted in two main ways depending on the subscription to either frequentist or subjectivist approaches (Reichardt & Gollob, 1997). These differences are illustrated in table 11.

⁵⁸ It has been contended that mathematical statisticians do not see what all the fuss is about in terms of both approaches (Huberty, 1993). Mathematically speaking, the two techniques are similar and yield useful information but it is the misinterpretations stemming from misuse of the approaches that niggles some research psychologists.

⁵⁹ The relationship between Fisher and the duo Neyman and Pearson was strained to say the least. They did not personally like each other one bit (Kline, 2004).

⁶⁰ Bayesian inference is perhaps the newest technique added to its already extensive repertoire (Edwards, Lindman & Savage, 1963).

⁶¹ Very similar to face validity actually where expert opinion is called on to make certain judgements about items in a test or questionnaire (Anastasi, 1988; Murphy & Daivdshofer, 1998; Reber & Reber, 2001; Rust & Golombok, 1992; Smit, 1996). This is the first place to start; how "objective" is that? Bayesian prior estimates are usually doing the same thing! Reliance on expert opinion and committee members' experience play a decisive role in determining what is and is not included in a prior probability distribution (Rupp et al., 2004). The functional form of the prior probability distribution is of course taken into account, i.e. binomial probabilities for success will vary between 0 and 1. However, caution is attached to these decisions especially if the sample size is small. Larger sample sizes allow the data to dominate posterior distributions and previous personal opinion does not carry as much weight as it did before (Rupp et al., 2004). Perhaps this subjectivist prior approach is appealing in areas such as intelligence assessment in which prior expertise is used to determine intuitive concepts of what intelligence as a construct purports to measure (Rupp et al., 2004). As Dawkins (2006, p.106) states "Bayes' Theorem ... is a mathematical engine for combining many estimated likelihoods and coming up with a final verdict, which bears its own quantitative estimate of likelihood. But of course that final estimate can only be as good as the original numbers fed in ... the GIGO principle is applicable here...".

⁶² Cognitive heuristics are themselves very powerful and perform adequately in many instances; think of stereotyping (Augoustinos & Walker, 1995). On the face of it, treating individuals as if they were the mean stereotype is clearly absurd if not unethical; but this method of survival has performed an invaluable service! However, within psychometric testing, the use of heuristics is considered untenable due to its unscientific stature as decision-maker thus the need for verifiable estimates such as offered through Bayes (Martin & VanLehn, 1995).

⁶³ Parameter estimation and hypothesis testing are the main goals within a statistical rendering of psychological reality and one will never escape this it seems, for even advanced forms of mathematical modelling of responses to items on tests are preoccupied with such issues (Van der Linden, 1994).

Table 11 Differences in probability definitions within the frequentist and subjectivist approaches (Reichardt & Gollob, 1997)

Definition of probability	
Frequentist / Classical	Subjectivist / Bayesian ⁶⁴ / Inverse
<ul style="list-style-type: none"> Both agree that probability be associated with mathematical axioms of probability 	
<ul style="list-style-type: none"> Probability of a repeatable event is asymptotically relative to the frequency of its occurrence over unlimited repeats (under identical circumstance bar random variation) 	<ul style="list-style-type: none"> Probability is the prior belief in the occurrence of an event in nature. The state of nature is dependent on the observer
<ul style="list-style-type: none"> Probability of a finite occurring event is either 1 or 0 	<ul style="list-style-type: none"> No differentiation between repeatable and finite events being probable
<ul style="list-style-type: none"> Probability is independent of human cognition 	<ul style="list-style-type: none"> Probability is dependent on human cognition and this also varies across observers
<ul style="list-style-type: none"> Prior probability is determined by the population from which the sample is taken <ul style="list-style-type: none"> ◆ <i>Uniform prior</i> distribution yield parameters that are all equally likely (non-informative)⁶⁵ ◆ For the frequentist this would mean that the randomly chosen sample was drawn from a randomly chosen population (across populations) ◆ <i>Nonuniform prior</i> distribution (any prior distribution that is not uniform and informative; such as a <i>beta</i> prior which is suited to binomial estimates as well as the normal distribution) ◆ For the frequentist this would mean sampling from a randomly sampled population in which the population parameters themselves are normally distributed ◆ <i>No usable prior</i> distribution (when there is simply no information available or too little information is given). The frequentist would have no usable prior distribution if a random sample is drawn from a randomly drawn population from which the population parameter is unknown Underlies the Neyman-Pearson school of statistics. Inference to the parameter is made via analogous findings from sample statistics. Decisions are made prior to data collection and inference follows automatically once the data is collected 	<ul style="list-style-type: none"> Prior probability is determined on subjectivist yet expert beliefs <ul style="list-style-type: none"> ◆ <i>Uniform prior</i> distribution yields parameters that are all equally likely (non-informative) ◆ For the subjectivist this would mean that a random sample was drawn from a population in which all possible population parameters were considered as equally likely ◆ <i>Nonuniform prior</i> distribution (any prior distribution that is not uniform and informative) ◆ For the subjectivist this would mean randomly sampling from what is thought to be a normally distributed population of population parameters ◆ <i>No usable prior</i> distribution (when there is simply no information available or too little information is given). The subjectivist would have no usable prior distribution if the uniform distribution was thought to be incorrect but the shape of the nonuniform distribution could not be completely specified Underlies the Bayesian school of statistics. Inference to the parameter is made directly. Competing rival hypotheses are evaluated in tandem with the gathered data

The theorem is a method for taking into account the conditional probabilities when estimating the probability that an hypothesis is true (Reber & Reber, 2001) as well as determining the probability that the sample effects are worth further investigation (Harlow, 1997). In utilising prior information (unconditionals) with empirical data (data collected), posterior distributions are generated (likelihood of results) which are utilised as the basis for statistical inference (Pruzek, 1997). Bayes' theorem has been successfully used within economics, military and combat strategies as well as artificial intelligence for instance (Gigerenzer et al., 1990; Majumdar, 2004). Mathematically, the joint or combined probability of data and the hypothesis is the product of probability of the data and the conditional probability of the hypothesis, given the data (Anastasi, 1988). Reluctance to use Bayesian statistics may be partially due to its non-positivist tenets, which as has been discussed in chapter 3 does not make for good science progress, as perceived by positivist classical significance testing (Barlow, 1992; Edwards, Lindman & Savage, 1963).

⁶⁴ Recall that not all subjectivists ascribe to Bayes' theorem. The two are not synonymous but have both been included here for purposes of broad-based comparisons.

⁶⁵ Coined "the principle of insufficient reason" by Laplace and "the principle of indifference" by Keynes (Oakes, 1986). If there is no information as to the likelihood of an event given a spread of probabilities for events, one assigns equal probabilities to all such events.



Expressed in more compact fashion after Kline, 2004:

$$p(d \wedge h) = p(d) p(h|d) = p(h) p(d|h)^{66} \quad (i)$$

which if solved for the conditional probability allows for the expression to be written out as Bayes' theorem:

$$p(h|d) = p(h) p(d|h) / p(d)^{67} \quad (ii)$$

Equation (i):

- $p(d \wedge h)$ is the conjunction of the data and the hypothesis which is the product of
- $p(d)$ which is the probability of the data and
- $p(h|d)$ which is the likelihood⁶⁸ or posterior probability (and hence conditional)

Equation (ii):

- $p(h|d)$ is the posterior probability of the hypothesis given the data which is determined by the
- $p(h)$ and $p(d)$ which are the prior (unconditional) probabilities of the data (regardless of the truth of h) and the hypothesis (its probability before the data are collected) and
- $p(d|h)$ which is the likelihood or conditional probability of the data given the hypothesis (similar to the p value in NHST)

Fisher contended that knowledge of prior probabilities was rarely the case and argued against Bayesian probability (Geisser, 1992; Krueger, 2001; Mulaik, Raju & Harshman, 1997; Oakes, 1986) and maintained "that the theory of inverse probability is founded upon an error, and must be wholly rejected" (Fisher, 1958, p.9). As a frequentist, he preferred sampling from known existing populations and not relying on probabilities of these populations parameters, in other words he preferred certainties and subsequently Bayesian inference was largely ignored within early psychological research (Good, 1992). It resurfaced in 1937 in a paper by De Finetti and has since been intermittent in its influence within the social and behavioural sciences as well as statistics (Barlow, 1992). Fisher is known for his views on sufficient statistics which is the notion that a quantity x depends on an observable random variable x but not on an unobservable parameter θ . It is sufficient if that statistic captures all of the information in x that is relevant to the estimation of θ (http://en.wikipedia.org/wiki/Sufficiency_%28statistics%29). In other words, a statistic exhaustive of information pertaining to its modeled parameter (Wright, 1997b).

One can see the logical reasoning of Fisher given his advocacy of experimentation as research technique. Neyman and Pearson agreed with Fisher regarding the use of Bayesian assumptions but the former also believed that Fisher's work was not entirely deducible from first principles and was also not logically coherent (Oakes, 1986).⁶⁹ Neyman and Pearson added the alternative hypothesis to their theory incorporating the probability of accepting an hypothesis when it is false thus attempting to make more complete, practically applicable and consistent Fisher's original test of significance (Gigerenzer, Swijtink, Porter, Daston, Beatty & Krüger, 1990; Harlow, 1997; Howell, 1987; Huberty, 1993; Kline, 2004) but in so doing they were vehemently opposed by Fisher (Cohen, 1990). The Neyman-Pearson approach formalised the accept-reject theory of hypothesis testing (Fraser, 1992) which is currently the more generally accepted form (Nester, 1996). Optimum decision rules were to be enforced so as to allow for acceptance or rejection of hypotheses, which was contrary to Fisherian and Bayesian ideas of testing with the aim of informing belief (Oakes, 1986). To determine if there is a difference between the dynamic assessment performance of an experimental group who has received mediation and a control group who has received no mediation the null hypothesis is in fact tested, i.e. assuming that both groups are equal in terms of having gained something which might on the surface seem odd seeing as that is not what one is testing but rather the alternative hypothesis (Cohen, 1990, 1994). Moreover, significance testing is only inferential in as far as the associated probability is concerned and is decision-theoretic as far as its significance levels go (Oakes, 1986).

⁶⁶ For $p(h|d)$, read: "the probability of the hypothesis given the data".

⁶⁷ Making use of two random events A and B , the equation is exactly the same: $p(A|B) = p(A \text{ and } B) / p(B) = p(B|A) p(A) / p(B)$ (Rupp et al., 2004).

⁶⁸ Likelihood is proportional to probability. In other words if $p(A|B) = p(B|A) p(A) / p(B)$; then $p(B|A) / p(B)$ are the likelihood functions of $p(AB)$. The likelihood of hypothesis A given the data B is proportional to the probability of the data B given hypothesis A . Hence $L(A|B) \propto P(B|A)$ assuming $\alpha > 0$ (Edwards, Lindman & Savage, 1963; <http://en.wikipedia.org/wiki/Likelihood>; Oakes, 1986). $L(A|B) = p(B|A)$. But $p(A|B) \neq p(B|A)$. Likelihood cannot be interpreted in a probable manner. Miller (2004) states succinctly: probability = knowing parameters and hence predicting the outcome; likelihood = observation of data and hence estimating the parameters. Hence, "if the probability of an event X dependent on model parameters p is written $p(x|B)$ then we would talk about the likelihood $L(p|x)$ " (p.26).

⁶⁹ An echo perhaps of the axiomatisation of mathematics during the formalist period? It is interesting how similar the paths taken by different subject areas really are.



What researchers want to know is the probability that the null hypothesis is true given the observed data. NHST focuses on the given of a null hypothesis and then tries to determine the probability of obtaining the observed data, i.e. $p(d | H_0)$ (Cohen, 1994; Huysamen, 2005). Fisher maintained that H_0 was to be nullified and the alternative hypothesis to be debunked (Hunter, 1997; Porter, 1997) or at the very least to have judgement suspended (Howell, 1987) if there is insufficient evidence to reject the hypothesis. This is a scenario often playing out within the natural sciences (Rossi, 1997). Suspending judgement about the alternative hypothesis (which Fisher never employed in any event) does not mean that the null is true. Fisher chose, rather, to replicate his experiments (Gauch, 2006); an aspect neglected in the social sciences. The greater the number of replications the greater the likelihood of accuracy in findings because there is a concomitant decrease in chance variance, akin to power. Current erroneous logic has since led to thinking that H_0 is in fact zero; hence the “nil null hypothesis” (Cohen, 1994; Huysamen, 2005; Nickerson, 2000) even though it is merely nonsignificant (Schmidt & Hunter, 1997). Compounded by the tendency of researchers to equate observed differences to true differences (Nester, 1996) the tradition of NHST has been seriously questioned and its use as a tool of science has also been scrutinised.

Theoretically then, Fisher would suspend judgement based on an experiment which yielded a non significant result but Neyman-Pearson would make a definitive decision as to the outcome or conclusion (Howell, 1987) a situation which has persisted in psychological practice and is quite pervasive in introductory psychological statistics textbooks (Huberty, 1993). Fisher (1958) even states that some researchers and academics may have misread him in prior publications on this issue and reiterates his stance on the process of nullifying the hypothesis saying that this process is “part of the hypothesis to be tested” (p.125). This brings into conflict the whole idea of scientific progression based on refutation of the alternative or research hypothesis as advocated by Popper and the idea propounded by Fisher; that of inductive inference via rejection of the null hypothesis (Cohen, 1990; Meehl, 1978). Recall Meehl’s preoccupation with the progress of psychology as a scientific endeavour by seeking to value the affirmation of theoretically predicted research results from practice as opposed to merely stating groups’ statistical significant differences at a prior determined probability level (Ray, 2006). Citing Huberty (1993, p.318) the following step-wise discussion (table 12) highlights in brief the difference between the Fisher and Neyman-Pearson approaches to significance and hypothesis testing:

Table 12 Main difference between Fisher’s P-value based approach and Neyman-Pearson’s fixed alpha approach (Huberty, 1993, p.318; Kline, 2004)

Fisher significance testing or P-value based approach	Neyman-Pearson hypothesis testing or fixed-alpha approach
1. State H_0	1. State H_0 and H_A
2. Specify test statistic (T) and referent distribution	2. Specify test statistic (T) and referent distribution
3. Collect data and calculate value of T	3. Specify α value and determine rejection region (R)
4. Determine P value	4. Collect data and calculate value of T
5. Reject H_0 if P value is small; otherwise retain H_0 (but suspend judgement on H_A ; Howell, 1987)	5. Reject H_0 in favour of H_A (in other words accept H_A ; Howell, 1987) if T value is in the rejection region; otherwise retain H_0

The arbitrary choice of zero as a prior probability for H_0 cannot account for cumulated knowledge in the discipline (Killeen, 2005). How does Popper’s rationalisation of the attempts to falsify theory work within NHST? The predictive variable (one which upholds the theory) is set up as H_0 and attempts at rejection of H_0 leading to a challenge of the predictive theory. This is referred to as the “strong” form of testing (Cohen, 1994; Meehl, 1997) or the “acceptance-support” form (Kline, 2004; Nickerson, 2000). It is Popper’s philosophy in NHST action (Meehl, 1967) and what Chow (1998b) refers to as repeated attempts at falsification which results in a convergence of operations over a number of such attempts. The alternative hypothesis is what the researcher wants to try and disprove (Kline, 2004). Its opposite incantation is, however, decried as “weak” when theories are confirmed by the rejection of H_0 or the “rejection-support” contention (Huysamen, 2005; Meehl, 1997; Steiger & Fouladi, 1997)⁷⁰ where H_A is the researcher’s alternative hypothesis illustrating the chosen theory (Kline, 2004). The strong/weak issue is more an epistemological and logical issue as opposed to a statistical one (Meehl, 1997). This is what Trafimow (2003) refers to as researchers setting up obvious hypotheses to be rejected when in fact they should concentrate on testing nonobvious hypotheses (although this makes life at present difficult for those seeking article publication, as H_0 acceptance is not “good science” as far as current ideas on psychological science go; Hunter, 1997. See chapter 2 section 2.4.5.3.6 where this very issue is cited as militating against the growth and progress of psychology as a formal discipline). Meehl (1989) begrudgingly refers to this process as hypothesis concoction seeking “to preserve a theory from falsification” (p.36) and the “weakness of null hypothesis *refutation* as a means of corroborating psychological theories” (original emphasis) (Meehl, 1998, p.5, 1990). See Table 13 for a clearer exposition of the weak and strong renditions of NHST.

⁷⁰ Thompson (1988) in Nickerson (2000) asks the question of how many software packages in fact test for the alternative hypothesis as most concentrate on the “no difference” relationship.

Table 13 Alternative forms of NHST

NHST weak and strong forms (research designs predicated on both deduction and induction)	
Weak	Strong
<ul style="list-style-type: none"> ▪ Reject-support contention ▪ H_0 to be rejected ▪ H_A is the researcher's theory ▪ Researchers take care to make α as low as possible due to the greater likelihood of falsely rejecting H_0. Thus emphasising Type I errors and increasing the power of the test ▪ Sample size is important ▪ Over-reliance on point nil null estimates hypotheses. Unable in most cases to generate point hypotheses due to the weak form of ad-hoc theories ▪ Makes the goal easy to obtain ▪ Typifies the social sciences via a process of verificationism ▪ Seemingly objective veneer of science practice ▪ Sets the scene for rejecting an already likely null hypothesis thus offering nothing new in the way of information ▪ Uninformed social science researchers are either unwilling or unable to change this scenario for fear of non-publication and/or due to ignorance of the philosophy behind the weak form of NHST ▪ Should start to present findings which are non-significant. Meta-analysis goes some way in rectifying this situation by including the "file drawer" calculation 	<ul style="list-style-type: none"> ▪ Accept-support contention ▪ H_0 to be accepted ▪ H_A is not the researcher's theory ▪ Researchers should concentrate more on Type II errors as there is a greater likelihood of falsely accepting H_0. Hence α should not necessarily be set too low. Power is inextricably linked to NHST. Increasing power would necessitate increasing sample sizes but to obtain prior levels of power sample sizes become increasingly difficult to obtain ▪ Sample size can work against the research hypothesis ▪ Does not necessarily rely on point nil null estimates hypotheses but rather a range or composite; in many instances relying on point estimates thus making the job of acceptance-rejection more clear cut. This is possible only due to the theory's already strong status ▪ Makes the goal difficult to obtain ▪ Typifies the natural science approach towards disconfirming the research hypothesis; a trend of falsificationism. Note that natural sciences thrive despite the non use of NHST ▪ A more objective form of science practice than is offered by the weak form ▪ Sets the scene for a stringent application of nullifying the research hypothesis ▪ Informed researchers as well as editorial boards see the light in terms of assessing the progress of a discipline by means of setting up more stringent hurdles. Publication of negative results needs to be taken more seriously ▪ Should encourage further publication of negative or non significant results

Physics, for instance, proceeds by setting itself ever-increasing standards of severity (invoking the strong form of testing) whereas the behavioural and social sciences seek the way forward by relying on less severe although scientifically feasible standards (and hence invoking the weak form of testing) (Meehl, 1967). Recall in chapter 3 though that Meehl (1997) has stated that psychology is far from the stage of Popperian falsificationsim in terms of its mechanism of growth - it is too immature to proceed on the basis of refutation alone. It seems Meehl has, since 1967, tempered his opinions regarding the utilisation of this approach; in fact he as much admits his staunch Popperian views which have since 1967 mellowed (Meehl, 1998). Likewise falsificationist NHST (attempting to falsify H_A as opposed to H_0) is thus not quite yet viable nor feasible at this stage (Krueger, 2001). In addition, Huysamen (2005) mentions the ill-preparedness of statistical courses (specifically in South Africa) in dealing with a change in NHST and other related areas such as introducing Bayesian statistics (Moore, 1997), due to the unfamiliarity of the technique within the social sciences even though it was introduced into psychology in the 1960's (Kline, 2004). The author is not convinced of this retort as it does not even attempt to rectify the situation. It is logical that people are unfamiliar with some aspects, but most people are unfamiliar about most things until they learn them!

The p value is an exact level of significance statistic from Fisher (but recall that he only utilised H_0); fixed levels of α across studies is the Neyman-Pearson approach which represents values for the probability of a Type I error and β the probability of making a Type II error (Kline, 2004; Oakes, 1986). This hybridised approach towards statistics utilising Fisherian statistics (historically preceding Neyman-Pearson by approximately ten years; Huberty, 1993) along with the Neyman-Pearson school of statistics is interesting as they are both quite different in approach (Gigerenzer et al., 1990; Huysamen, 2005; Kline, 2004; Loftus, 1996; Mulaik, Raju & Harshman, 1997). The resulting "mishmash" of the hybridised approach has had the unfortunate consequence of being incorrectly perceived as statistical methodology within many statistics courses (Cohen, 1994). Neither frequentist school would have looked with favour upon the "forced marriage" of the approaches (Chow, 1998a; Gigerenzer et al., 1990, p.106, 1991). Are assessment specialists even aware of the controversy surrounding the historical origins and misapplied statistical techniques which are used everyday? Do practitioners ever question the fundamental philosophies behind their



reasoning? Is this one reason why dynamic assessment fails at times to deliver the necessary goods in terms of robust science? Dynamic assessment predicates are already tuned towards qualitative methodologies which do not necessitate the use of NHST in order to ensure its progress within assessment. The above-mentioned issues which plague mainstream intelligence assessment issues need not bear upon dynamic assessment in the first place. Of note are the models assessed in chapter 5 which evidence strong testability according to Madsen's HQ ratio but do not, as a rule, necessitate NHST in advocating their science. Even though this type of statistic is used to "prove" the veracity of the method it is not employed in the original derivation of hypothesised novel constructs (meaning-making constructs which will be discussed in chapter 5).

The afore-mentioned progenitor's statistical work as utilised in psychology today are sometimes at odds with what occurs in practice. Practical instances include:

- disregard for other indicators of potential leanings towards significance such as confidence intervals utilised in classical frequentist statistics, Bayesian inference⁷¹ (two techniques which have been in existence far longer than significance tests; Schmidt & Hunter, 1997), the plotting of data, the likelihood function
- the ensuing miscegenation of two approaches towards chance and certainty
- the simultaneous following of divergent recommendations such as the need to establish a prior level of significance before an experiment is conducted, thus following the dictates of Neyman-Pearson only to be followed by the suggestion of not commenting on non-significant results and thus following Fisher
- Neyman's behaviourist slant on data interpretation has been all but ignored
- Type I and Type II errors are regarded as philosophical issues
- which leads to statements extolling the rigour and accuracies with which null hypotheses have been rejected

Appendix 1's meta-analysis takes into account the number of studies needed to make insignificant a significant cumulated effect size. Although as Huysamen (2005) points out, a single study with a very large sample size could theoretically have the same NHST result as a meta-analysis utilising the same sample size even though it emanated from various smaller studies. However the NHST logic would be applied to the individual studies at the level of participants whereas the meta-analysis treats study as sample. All non-significant findings which have been filed away in drawers never to see the light of day are hence given weight. Will data manipulation, design and interpretation ever work back onto the reasoning behind certain psychological assumptions? Perhaps if we start to look at how we use and interpret data we will start to ask different questions about phenomena and not first work towards a research design. "Phenomenon - then - research design" and not "research design which is set in stone - then - phenomenon" could reflect a credo of sorts. It is perhaps timely to once again interject an authorial note here and state that however unequivocally the ideal of science is supported, as stated and maintained throughout chapters 2 and 3, the need to recognise the futility of some perennial efforts within a social science discipline is emphasised. It is not the question of maths, science, statistics, methodology, and overarching framework of ensconced theory growth that is being questioned but their subsequent ill-fitting use within psychology. Meehl (1978) critically appraises the resultant lack of integration of ideas behind statistical significance and a progression of science built on refutation and laments the unfortunate path that has been followed in the "soft sciences" headed by significance testing. It is a thesis that many psychologists trained today are simply not aware of original problems and are perhaps not even aware of prevalent issues surrounding nagging issues. This is evident in their current existence. NHST smacks of logic and mathematical certainty; after all the foundation of statistics is mathematics and logic and these have already been highlighted as being fraught with uncertainty and foundational cracks.

4.3.1.2 The hegemony of null-hypothesis significance testing

As with any contentious debate there are always two-sides to each story and the rendering of an objective picture surrounding this issue will result in a resolution of sorts in a shorter period of time. Many dynamic assessment research designs (specifically the South African literature; Murphy, 2002; Murphy & Maree, 2006) are specifically deployed to ensure that results point to either one of two scenarios: either an intervention is significant in terms of producing measurable change or it is not significant in bringing about change. The result cannot lurk about in no-man's land as this is simply not good science as it is not objective and does not adhere to positivist tenets of "progression". Statistical decision theory as exemplified by null hypothesis significance testing within psychology (Boniface, 1995) has served the positivist agenda well in terms of following its most basic mandates. However, as to whether it follows the development of science as exemplified by a number of theories within the philosophy of science is questionable (Brown, Hendrix & Williams, 2003). The usefulness of this type of accept-reject research design within psychology has been raised as a philosophical query (Krantz, 1999) befitting the style of this study. Krantz (1999) ties two issues pertinent to chapters 3 and 4 together in his question of whether the logic of null hypothesis significance testing matches the logic of scientific inferential practice specifically in psychology and dynamic assessment even though NHST is not all there is to scientific inference. This leans heavily on the question of how a science progresses from naive ideas concerning human

⁷¹ Bayesian "confidence intervals" are referred to as credible intervals (Oakes, 1986; Rindskopf, 1997).



behaviour, intelligence and learning potential towards the science of human behaviour and intelligence. NHST is often employed to bring the veneer of objectivity to psychology (Loftus, 1996). Often the decisions made on the basis of NHST findings are logical derivations but are not necessarily reflective of insight gained into any particular psychological phenomena. The “dubious epistemological claims” (Brown, Hendrix & Williams, 2003, p.78) of NHST has thus come under fire from a number of fronts as will be illustrated below. Recall that great names in psychology have founded schools based not on NHST but on sound observation, theoretical development and experimental endeavours and include Piaget, Skinner, Pavlov and Bartlett (Barrett, 2005).⁷²

A criticism often levelled at psychology in general is that it fails to replicate, the “field often spends a lot of time spinning its wheels without really making much progress” (Edwards, Lindman & Savage, 1963; Kline, 2004; Loftus, 1996, p.161; Lykken, 1991; McDonald, 1997), although it is conceded that replication is more difficult within the behavioural sciences as opposed to the natural sciences (Kline, 2004). What is it about NHST that is so contentious to some and not to others? One such issue is NHST’s reliance on inductive inference which can only propel a field forward if studies are replicated (Krueger, 2001) which psychology often does not do (Huysamen, 2005). As highlighted above, the link between scientific progress in keeping with the tenets of positivist leanings towards the growth and development of a subject considered scientific and the methods utilised in order to achieve and maintain such standings within the framework are not always manifest or clear. Perhaps they are simply taken for granted. In the hope of maintaining psychology’s reputation (such that it is) within the broader practise of science, research designs which look objective, formal and scientific are employed within varied domains and none has been carried out with such wide-spread enthusiasm as NHST (Nester, 1996; Nickerson, 2000). Ironically, natural science disciplines such as physics and chemistry perceive reliance on significance testing as unscientific (Schmidt & Hunter, 1997)! Note the emphasis, once again, on the ideal of formalism just as was discussed in the section on mathematical formalism. Loftus (1996) highlights the logic of NHST and why it is thus a universally employed technique within the social and behavioural sciences:

- i. the research question is suited to the design which is one way of trying to show that an independent variable will have an effect on a dependent variable (dynamic assessment interventions of varied sorts will effect posttest scores; especially on those whose resident potential has been severely underestimated by static or conventional assessments)
- ii. two or more groups are compared for differences which may or may not have been caused by the intervening variable (control and experimental groups’ scores are compared for differences after dynamic assessment interventions have taken place). The major task is to determine whether any differences are due to the intervention or if they are randomly occurring measurement errors
- iii. in order to accomplish this, a probability estimate (p) is necessitated which will show the probability of observing differences as great as those in fact observed. It is usually the case that the ‘no-difference between groups hypothesis’ (H_0) needs to be rejected (the probability of evidencing as large a difference as was evidenced is greater than chance alone; the control and experimental groups really did differ on posttest scores).
- iv. the researcher makes a binary decision based on the level chosen for acceptance of the probability. This level is known as alpha (α). If the p value is less than α then the null hypothesis is accepted; or as Loftus (1996) states, a strong decision is taken to reject H_0 . If the p value is greater than α then H_0 is accepted or a weak decision is taken to reject H_0 . As this is all a game of chance, it is possible to incorrectly reject or accept H_0 . If a true H_0 is incorrectly rejected then the probability of making this error is equal to α , referred to as Type I error. This is known. However, if the researcher incorrectly fails to reject a false H_0 then the probability of doing so is known as β which is not usually known as there is usually no information pertaining to populations means assuming that H_0 is false. Power is usually also not known because power is equal to $1-\beta$. Referred to as Type II error, but were it known, the power of the test then too would be conditional and not exact (Chow, 1991, 2002), an aspect highlighted by Chow as perhaps having escaped attention or being misunderstood by those touting the utilisation of power versus NHST as indicator of an experimental effect
- v. based on the decisions taken at each step, the researcher then has to cope, with at times, very complex data sets

Now that the scene is set for NHST interpretation, Loftus (1996) summarises six questionable aspects regarding the use of NHST within psychological settings and dynamic assessment is one such domain:

- H_0 is rarely true to begin with (Cohen, 1994; Harris, 1997; Hunter, 1997; Killeen, 2005; Krueger, 2001; McDonald, 1997; Meehl, 1978, 1989, 1990; Rindskopf, 1997; Schmidt & Hunter, 1997). It is usually the “default hypothesis” (Kline, 2004); the “silly null hypothesis” (Nester, 1996); the “notoriously non-powerful statistic” (Rust & Golombok, 1992) or “the usual dismal prediction” (Meehl, 1990) that is rejected at a later stage and can be dismissed without any data collection at all (Rindskopf, 1997). In this vein the two hypotheses are treated asymmetrically and very rarely is a prior probability

⁷² Of course it can be stated that these researchers’ work was far more amenable to these strategies than perhaps it is for researchers working in fields where constructs are not so easily defined.



(Bayesian) associated with H_0 other than zero⁷³ (Oakes, 1986). If H_0 is usually false, sufficient sample sizes will effectively deal with their consequent rejection (Kline, 2004) and of course this is not stringent science. As a point hypothesis⁷⁴ this makes sense as no two populations means' will ever be precisely the same. In other words there will always be a difference somewhere lurking in the means, the task is to determine the feasibility of stating how different the two means in fact are. Yet, classical significance testing relies on a point null hypothesis and a distributed or diffuse alternative hypothesis (Edwards, Lindman & Savage, 1963) and in a way, stacks the odds in favour of rejecting the null. Three groups receiving qualitatively different mediatory exposures will obviously differ on mean performance as they all receive differing amounts of mediation. H_0 is employed only as a means of illustrating the implausibility of a result being obtained given the data and as such "a significance test does not permit one to assign any specific degree of probability to the hypothesis" (Gigerenzer, Swijtink, Porter, Daston, Beatty & Krüger, 1990, p.93). Nester (1996) agrees with this sentiment as he states that even if treatments yielded similar effects the chances of this happening are zero! Significance testing was never meant to be universally applicable in the first place (Hunter, 1997). Logically, one cannot defend the reasoning behind why H_0 should be a hypothesis of no difference regardless of how large the sample size is and subsequent acceptance of H_0 due to an insignificant result cannot be justified (Krueger, 2001). If H_0 is almost always false then logically the rate of Type I errors is 0% and not 5% which results in Type II errors being made (Cohen, 1994). Testing for differences between all three groups' means, results in no new information as there will be differences regardless. "Rejecting a typical null hypothesis is like rejecting the proposition that the moon is made of green cheese. The appropriate response would be 'well, yes, okay but so what?'" (Loftus, 1996, p.163). Recall that a null hypothesis is really just a hypothetical model (Gigerenzer et al, 1990). P is not the probability that the null hypothesis is true and rejecting the null hypothesis does not necessarily lead to the vindication of the alternative hypothesis (Cohen, 1994; Nickerson, 2000); i.e. $p \neq$ value of H_0 and $1-p \neq$ the value of H_A

- attaining a level of statistical significance only indicates that H_0 is false and conveys no useful information as to the underlying pattern of the populations means' which is, after all, what researchers are after. Employing post-hoc tests are often riddled with their own problems of similar false rejections of H_0 when errors of the likes of number iv above, are made. Planned comparisons are able to overcome this problem but are rarely used in practice
- NHST rarely pays attention to power primarily because it cannot be accurately computed. Type II errors and hence power cannot usually be computed because there is simply no information available detailing quantitative hypotheses. With high power the researcher can conclude that there are small differences between group means and accepting H_0 may be justified. Low power implies that there may be large undetected differences between means. The use of confidence intervals (allowing a range of data including the parameter whose probability is known; Krishnamurty, Kasovia-Schmitt & Ostroff, 1995) is an alternative to the use of power, a recommendation shared by the TFSI (Nickerson, 2000; Wilkinson & TFSI, 1999). "A lack of power analyses often stems from the lack of quantifiable alternative hypotheses that characterizes the social sciences in general, and psychology in particular" (Loftus, 1996, p.163). Within some natural science disciplines, confidence intervals are not interpreted as significance tests (Schmidt & Hunter, 1997). Section 4.4.2.3 below includes an IRT model specifically developed to encompass change across testing situations which utilises confidence intervals as measure of estimability and modifiability estimates
- the mechanical, "formulaic mode of inquiry" (Robinson 2000 in Brown, Hendrix & Williams, 2003, p.78) and automatic need to dichotomise the two-valued decision of acceptance or rejection of H_0 is lamentable as very few social issues ever present in which decisions of the categorical yes/no variety obtain (Abelson, 1997a; Cohen, 1994; Edwards, Lindman & Savage, 1963; Gigerenzer, 1991; Huysamen, 2005; Kline, 2004; Krueger, 2001; Lehmann, 1992; Rossi, 1997; Schmidt & Hunter, 1997). To cite Cohen (1990) "there is no ontological basis for dichotomous decision making in psychological inquiry" (p.1311). Most judgements within the social sciences especially need to be mindful and this is very difficult to apply when the decision is based on a dichotomy (Harlow, 1997). Statistical inference especially within psychology should only allow for more or less support to be added to a conjecture and should not be utilised for the sole purposes decision-theoretic making based on arbitrary figures (Oakes, 1986). This is more an outgrowth of the Neyman-Pearson "decision-theoretic significance testing" approach as opposed to the Fisherian approach (Mulaik, Raju & Harshman, 1997, p.106; Oakes, 1986). Harris (1997) makes a convincing claim for the utilisation of the three-valued alternative as

⁷³ Recall Fisher's objection to this idea of subjective priors. But also bear in mind that his area of concern (agronomy) was better suited to nullification of H_0 .

⁷⁴ This is not to say that NHST as it cannot be functionally utilisable as stringent procedure of acceptance or rejection within certain set-ups (Kline, 2004; Meehl, 1990; Oakes, 1986). For instance, if one is determining the point at which a metal will start to collapse structurally given sufficient weight it is clearly best to signify a definite point at which this will happen. However, to play devil's advocate, the same philosophy should undergird health issues such as medical treatments for instance. At which point is it decided that a certain medication is indeed safe? Does one disregard outcomes evidencing p values over 0.05? What about trials evidencing p values of 0.06; should the medication not be given the go ahead merely because of the arbitrary choice of a p value? What about individual study results touting efficacy of certain treatments only to find that once cumulated in a meta-analysis such significance is partialled out and the subsequent effect sizes decrease? A call for Bayesian inference has often been made but due to the hesitancy of researchers to subjectively assign prior probabilities this method has been viewed with suspicion as it does not adhere to NHST's "objective" veneer (<http://www.cs.ucsd.edu/users/goguen/courses/275f00/stat.html>) (Nester, 1996). Bayesian inference typically utilises a range of values as opposed to point estimates (Rupp et al., 2004).

opposed to this traditional two-valued approach to NHST which in essence is the “simultaneous use of two directional alternative hypotheses one for and one against the research hypothesis” (Kline, 2004, p.84). Users of such an approach will be less likely to equate the nonrejection of H_0 with the acceptance of H_0 , a major point in the confusion in the interpretation of nonrejected hypotheses for the last sixty years (Howell, 1987). Conflation of the substantive theory (clinical or practical theory/hypothesis or causal theory depending on the context; Meehl, 1990; Nickerson, 2000 or practical reality; Danziger, 1990) with the statistical hypothesis (derived from the substantive; Oakes, 1986 or the constructed statistical reality; Danziger, 1990; Meehl, 1990) is often the reason that a number is the determining factor behind the acceptance or rejection of a hypothesis which does not necessarily lead to the acceptance or rejection of the theory behind the substantive hypothesis (Brown, Hendrix & Williams, 2003; Kline, 2004; Meehl, 1967; Nickerson, 2000; Wallis, 2004); i.e. rejection of the null hypothesis does not mean it has been disproved (Krueger, 2001). Meehl (1990, 1998) makes the valid point that the nullification of H_0 as an import from Fisherian statistics can hardly be blamed on Fisher but rather on those misapplying the rationale behind NHST. After all, agronomy experiments rarely presented with major differences between the theoretical and statistical hypotheses; the work involved in agronomy and the technique imported into psychological research are conceptually distant from each other. One cannot blame Fisher; he worked with crop yields and his methods aided him in so doing. One can however look upon social scientists with some disdain for having imported his techniques along with a mix of Neyman-Pearson into an area almost entirely ill-equipped to utilise those very techniques. To highlight the, at times, confusing role played by substantive theory (practical importance of theory) and statistical hypothesis Table 14 is illustrated below:

Table 14 Relation between substantive and statistical hypotheses and how sample size can impact on conclusions (Johnson, 1999; Nester, 1996)

Sates of mind of a null hypothesis tester and the interpretation of sample size as related to results of a statistical significance test		
<i>Practical</i> importance of observed difference	<i>Statistical</i> significance of difference and sample size status	
	Not significant	Significant
Not important	Happy / <i>n</i> okay	Annoyed / <i>n</i> too big**
Important	Very sad / <i>n</i> too small	Elated / <i>n</i> okay

** Related to Meehl's (1990, 1997)⁷⁵ “crud factor” (too much noise in the data) as at some level everything is related to everything (McDonald, 1997; cf. Nunnally; 1978, who delineates seven ways to fool oneself with factor analysis in similar vein). This is where expertise, experience and sound understanding of a field can bring into line the findings of NHST. NHST per se as statistical method is not flawed in terms of its method (Kline, 2004) but the unfortunate historical combination of hypothesis testing on the one hand and significance testing on the other within psychology is perhaps the cause of NHST's hegemony over the years. Researchers' blind reliance on statistical significance and the move away from the research underpinning it had already been critiqued by Boring in the 1920's (Danziger, 1990). This is in keeping with Brown et al's., (2003) contention that psychologists have confused numerical inference and conceptual inference. Once again the issue of a one-to-one mapping, an isomorphic relation between two conceptually different aspects is brought to the fore.

⁷⁵ Meehl (1998) in fact attributed this “crud factor” term to Lykken.

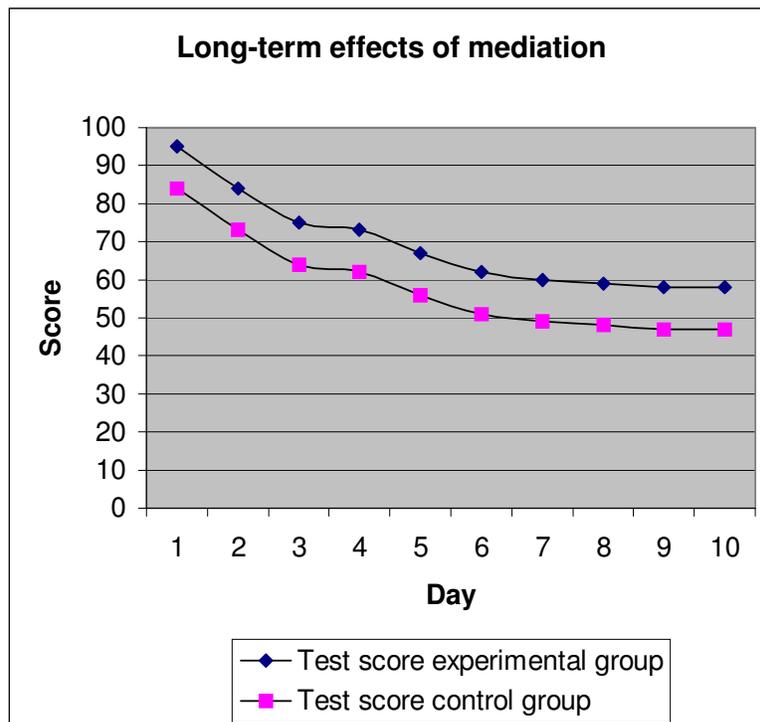
Chapter 2 discussed, among other things, the isomorphic mapping of neural-mind correlates. In this particular instance such a mapping is feasible as the one concept is reducible to the other. A similar mapping of number to reality however is not as tenable (Brown et al., 2003).

- However, in fairness to proponents of significance testing, Mulaik, Raju and Harshman (1997) highlight the fact that significance testing does not intend to disprove the substantive theory but only provide evidence for the validity of the statistical hypothesis, i.e. it is not the method which is to blame but the misinterpretation of its utility by researchers (Abelson, 1997b). Scientific significance and statistical significance are not synonymous. Recall that almost no single experiment within the behavioural sciences⁷⁶ is anywhere near conclusive even though the statistical hypothesis is quite clear on the issue. The leap from statistical and computational conclusions to inferential conclusions is an area fraught with problematic issues (Krueger, 2001). In fact Meehl (1967) points out with startling consequences that:
 - if there was increased precision of instrumentation and
 - even more sound logic flowing through experimental design and
 - increased sample sizes; one would in effect
 - increase the probability of corroborating the substantive theory via NHST “*even if the theory is totally without merit*” (original emphasis) (Meehl, 1967, p.111)
 - this manner of statistically navigating via means of significance testing might hold for statistical results but should not hold for substantive theory rejection or acceptance. This is, as already mentioned, the weak form of theory testing which allows for the further development of a theory when in fact is should not proceed in this fashion
- It is little wonder that confusion abounds when one considers the immense and vast array of incredibly accurate reportings of results in many areas within psychology and dynamic assessment but when taken as a whole no single substantive law-like conclusion is obtainable. Thousands of studies grace journals yielding impressive results but very few are able to make a substantive claim. Something is clearly wrong somewhere and the thread of this story began with chapter 2 and basic philosophical core issues followed by chapter 3 and basic understandings of the scientific method culminating in the discussion on prime considerations here. Nevertheless to continue with the present concern, Huysamen (2005) makes a case for the need that does perhaps arise where decisions of this nature need to be taken. In other words researchers are not necessarily interested in the size of or the degree to which something is or is not acceptable but are really interested in just knowing ordinal values (the answer is greater than zero as opposed to knowing by how much greater than zero it is). This human obsession with dichotomy has been addressed in chapters 2 and 3 as playing a decisive role within the scientific framework and has been mentioned in conjunction with the discussion on mathematics. It rears its head once again within the realm of statistical inference. NHST circles round the critical norm of accepting or rejecting hypotheses based on the 0.05 or 0.01 α level. This is an arbitrary number and not a proven fact chosen for the acceptance or rejection of an hypothesis originally deployed by Fisher (Gigerenzer et al., 1990; Huysamen, 2005; Kline, 2004) but not supported as sole determiner of the conclusion of an experiment (Huberty, 1993). An argument in support of this number is that a line has to be drawn somewhere and this just happens to the chosen point and in addition it does provide standardisation (Chow, 1998a; Huysamen, 2005). It also adds pragmatism to the method. The same information, namely, the probability of the data given the null hypothesis, will be extracted from the same data (Krueger, 2001). Granted this may be true but how valid is the application of this rule of thumb within many social science and behaviour contexts? Can dynamic assessment efficacy be reduced to an arbitrary figure? The assessment worked or it didn't according to the α level obtained. The acceptance of an hypothesis based on its 0.05 level or its rejection based on its 0.06 level is odd to say the least (Huysamen, 2005). Surely there is more to this method than this? Appendix 1's meta-analysis results does not make use of NHST in calculating the cumulated effect size of dynamic assessment studies (Chow, 1998a)

⁷⁶ “Almost no” as the author simply does not know. It is very doubtful if any experiment within any “soft” area ever comes close to approximating law-like results evidenced in some areas of the natural sciences. All the more reason to locate for the discipline an area amenable to either “non-science” and by this is meant a referral to science in terms of its concept discussed in chapter 3; or an area amenable to scientific methodology. The discipline as is cannot continue its existence somewhere in between, which is precisely where dynamic assessment is located. The author cannot help but insert this quote from Meehl (1967) regarding the practice of psychology; “... a zealous and clever investigator can slowly wend his way through a tenuous nomological network, performing long series of related experiments which appear to the uncritical reader as a fine example of ‘an integrated research programme’, *without ever once refuting or corroborating so much as a single stand of the network*. Some of the more horrible examples of this process would require the combined analytic and reconstructive efforts of Carnap, Hempel, and Popper to unscramble the logical relationships of theories and hypotheses to evidence. Meanwhile our eager-beaver researcher, undismayed by logic-of-science considerations and relying blissfully on the ‘exactitude’ of modern statistical hypothesis-testing, has produced a long publication list and been promoted to a full professorship. In terms of his contribution to the enduring body of psychological knowledge, he has done hardly anything” (original emphasis) (p.114). Traifmow (2003) is in agreement with this sentiment and it is also highlighted by Krueger (2001).

- Circular reasoning pervades theory and NHST. NHST posits that error is distributed in Gaussian fashion; the distributions are equally distributed over various conditions and adding up effects and error sources in a linear manner results in numerical indices. These assumptions are built into theory which biases against other types of reasoning within theories. Loftus (1996) employs a real-life example which will be attenuated to suit the following context-relevant example:
 - two groups, an experimental and control group are administered interventions. The control group is shown a video about how to study and the experimental group is given intensive mediatory assistance regarding study methods and techniques
 - over a period of time, the two groups are monitored for long-term changes in study behaviour. Unsurprisingly, the experimental group shows a higher level of acuity regarding study methods and does better in tests and exams than does the control group
 - the researcher concludes that a significant effect in test results was due to the intervention and illustrates this with the following figure

Figure 60 The theoretical significant difference between two groups either receiving or not receiving mediatory interventions couched within NHST technique (attenuated after Loftus, 1996, p. 165)



- standard data-analysis will conclude that the experimental group significantly outperformed the control group thus evidencing the superiority of the intervention. This, however, masks a true pattern which can be seen by the similar trends evidenced from both groups. They both decrease over a number of days with equal rates. The equality of rates of reduction, then, are testimony to the negligent effects of the intervention. The counter-argument would be that the experimental group nevertheless did significantly outperform the control group. The main aim though of the research was to document the long-term effects of mediation and based on the evidence did not seem to make any difference at all. NHST will miss this
- lastly but quite importantly is the logical error made when assuming that
 - H_0 should be rejected when
 - (i) $p(\text{observed data} \mid \text{null hypothesis}) < 0.05$
 - yet, rejecting H_0 implies that
 - (ii) $p(\text{null hypothesis} \mid \text{observed data})$ is small; which is similar to employing the logical argument that “P then Q” = “Q then P”, which is obviously false (Nickerson, 2000)



- (iii) but $p(d | H_0) \neq p(H_0 | d)$ ⁷⁷ which is what NHST logic in fact states (Cohen, 1994; Killeen, 2005; Mulaik, Rju & Harshman, 1997; Nickerson, 2000; Oakes, 1986; Trafimow, 2003). Probability and its inverse is not equal, to cite Carver (1978) in Suen (1990) "the probability that a person is dead given that the person was hanged is high. However, given that a person is dead, the probability that the person was hanged is quite low. In other words $p(\text{dead} | \text{hanged}) \neq p(\text{hanged} | \text{dead})$ (p.21). Bayesian statistics, however, does make use of inverse probabilities
- the second statement does not necessarily logically follow on from the first statement due mainly to the fact that the probability of the null hypothesis obtaining given the observed data could be anything at all (unless we have access to other information) (Chow, 2002). "When H_0 is rejected, it can be because of the falsity of any of the auxiliary theories about instrumentation or the nature of the psyche and not of the substantive theory that precipitated the research" (Cohen, 1994, p.999). Prior data on H_0 being true would be most useful which is precisely what a Bayesian technique would allow - it works with and re-integrates prior probability distributions in determining posterior probability distributions (Reber & Reber, 2001). In sum, the arbitrary level of 0.05 is just as meaningful as it is arbitrary! See chapter 2 section 2.4.5.3.6 for a reference to this in the context of the progression of science and instance confirmation which is based on conditional probabilities
 - Huysamen (2005) adds that rejection of H_0 does not logically imply that H_A is supported nor does it yield information about the degree to which the rejected hypothesis is false and lastly NHST is silent on its reliance on sample size

NHST yields $p(d | H_0)$ and does not yield $p(H_0 | d)$, $p(H_A | d)$, $p(\text{replication} | d)$ or $p(H_0 | \text{reject } H_0)$ which some researchers think it does (Kline, 2004). To link back to the ideas discussed in chapter 3 on the logical positivists and the ensuing development of science within such a framework, Cohen (1994) details the misinterpretation surrounding H_0 . If the logic of H_0 was phrased in syllogistic Aristotelian fashion the following would result:

- If H_0 is correct, then D (data) cannot occur (note the IF-THEN pattern)
- D occurs
- Therefore H_0 is incorrect

However, NHST (induction) is not couched in syllogistic reasoning (deduction) and the above *modus tollens* is not what usually results in NHST (Krueger, 2001; Nickerson, 2000) which is couched in probabilities such that you find the following:

A

- If H_0 is correct, then D is very unlikely
- D occurs
- Therefore H_0 is very unlikely

And by instituting probabilistic thinking the argument becomes invalid. In stating a false premise but making it appear palatable by dressing it in probabilistic clothing one is able to formally deduce incorrect conclusions. For instance:

B

- If the organism has legs it is a mammal
- the dolphin has no legs
- therefore it is not a mammal

The first premise is of course incorrect (lizards have legs for instance) however the syllogistic reasoning is correct leading to a correctly deduced conclusion (a Type III error, resulting from having asked the wrong question or in this instance, having posited the wrong statement; Harris, 1997; Killeen, 2005; Sehlapelo & Terre Blanche, 1996). If the statements were couched in probabilistic terms it would run as follows:

C

- If the organism has legs it is probably a mammal
- the dolphin has no legs

⁷⁷ $p(H_0 | d)$ is an inverse probability (Oakes, 1986). Neyman and Pearson maintained that the data might or might not support H_0 but given the data one cannot assume H_0 is upheld or not (Oakes, 1986).



- therefore it probably is not a mammal⁷⁸

The premise is probable, not absolute, thus making the premise more acceptable which is precisely the reasoning behind NHST. Following on from C, NHST would conclude the following:

D

- If H_0 is false, then the result (of statistical significance) would probably occur
- The result does not occur
- Therefore H_0 is probably true and hence valid

Gigerenzer (1991) refers to the tools of the trade within a science and distinguishes tools for data processing, namely, statistics; tools for nonempirical hypothesis testing such as logical consistency and tools for the original measurement for justifying what is being researched. The logic behind mathematics (chapter 4), the reasoning behind theory growth (chapter 3) and the need for quantification (chapter 4) tie in with this discussion on the logic of hypothesis testing along with the need for experimentation within psychology. Is it possible that, through the use of such tools, new theories can emerge and attest to Gigerenzer's (1991) "tools-to-theories" heuristic? This question is answered in the affirmative as often "the apparatus of statistics may be so closely tied to a regulatory function that the quantifiers help to generate or reshape the phenomena which they set out to describe" (Porter, 1997, p.102). Perhaps we should assess chapters 2- 4 in the light of tools having created for researchers new theories hitherto unavailable? Gigerenzer (1991) does make a compelling argument, one which needs to be taken seriously considering the leanings of this study which have questioned the techniques utilised for the study of psychology and assessment.

Oakes (1986), although now quite dated, obtained results from a variety of people (scholars, researchers and lecturers) on a number of questions concerning the proper interpretation of significance testing usage. The hypothetical set-up entailed a treatment given to an experimental group after which both the control and experimental groups' means were considered (20 in each sample). An independent means t test is administered and the statistic yielded is 2.7, df 18 at $p = 0.01$. The following table 15 details the most common misconceptions relating to NHST and the answer to each is an unequivocal 'no'!

Table 15 Common misconceptions pertaining to NHST

Usual interpretation of NHST	
i.	The null hypothesis is absolutely disproved
ii.	The probability of the null hypothesis has been found
iii.	The experimental hypothesis is absolutely proved
iv.	The probability of the experimental hypothesis can be deduced
v.	The probability that the decision taken is wrong is known
vi.	A replication has a 0.99 probability of being significant
vii.	The probability of the data given the null hypothesis is known

⁷⁸ Cohen (1994) includes syllogistic arguments which are inverted to the ones employed here. He makes use of the premise "If a person is an American, then he is probably not a member of Congress; the person is a member of Congress, therefore he is probably not an American". Cohen's example is more fitting because it denies the validity of H_0 and thus illustrates the whole logic much better than does the syllogism used by the author. What holds the one way will nevertheless hold vice versa.

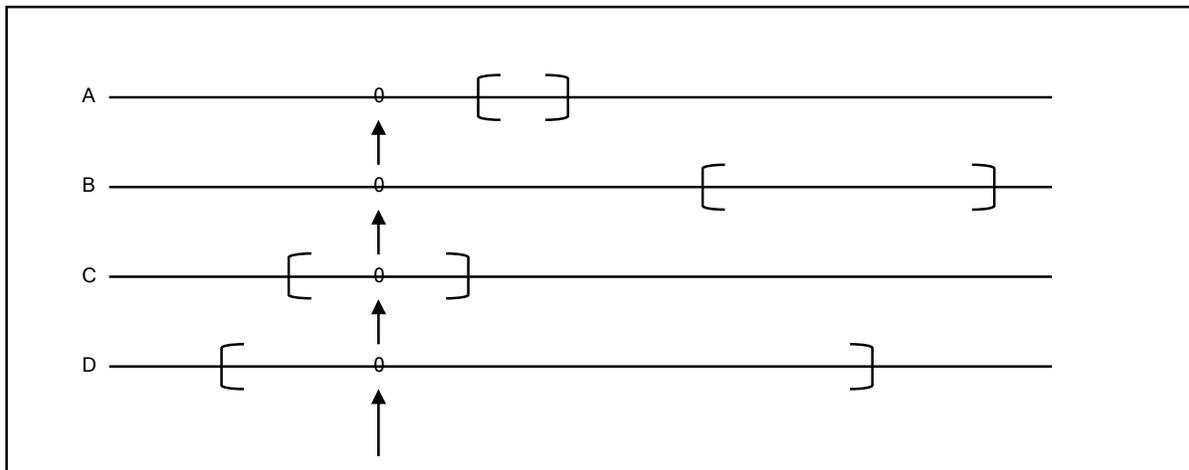
Table 16 Justification of the reasoning pertaining to NHST table 15

Reasoning
i. All that has been shown is the probability of equivalent means between the two groups. But then again, H_0 is rarely true to begin with! This is in and of itself no great feat. This same rationale underpins number iii.
ii. As with i, we only know that the probability of the H_0 is not due to chance.
iii. All that one can conclude is that there is a 99% probability that the results are <i>not</i> due to chance alone. This says absolutely nothing about the veracity of the alternative hypothesis' claim. We are no wiser about the substantive hypothesis than we were before the start of the experiment. A Neyman-Pearson confidence interval is not reflective of a parameter but an interval. This same rationale underpins number i.
iv. As with iii, we only know that the probability of the H_A is not due to chance.
v. The probability that the decision was based on chance is known and nothing more. Similar rationale underlies ii.
vi. Nothing can be said about the replicability of this experiment. The only way to further substantiate H_A is to conduct the study again. Unfortunately, replication in the social sciences is rare. The rule of thumb to follow if a claim is to be further supported would not be to increase power or sample size but to replicate studies more often. Associated probability is not power.
vii. Practically, this is the accepted interpretation of NHST; i.e. $p(d H_0)$. But it has been argued that this can never be! You only have one known aspect and that is the data with which you work. You cannot know absolutely what H_0 is otherwise you wouldn't need to test it in the first place. You do not know H_0 . You know the data. The data is given. Hence the probability of H_0 given the data is what you want to know. And as already pointed out, $p(d H_0)$ is not equivalent to $p(H_0 d)$.

Loftus (1996) suggests alternatives to NHST which, he states, are neither fancy nor esoteric. He suggests (as others have likewise suggested; Cohen, 1994; Hunter, 1997; Huysamen, 2005; Krueger, 2001; Meehl, 1997, 1998; Moore, 1997; Mulaik, Raju & Harshman, 1997; Nester, 1996; Nickerson, 2000; Steiger & Fouladi, 1997) that researchers provide confidence intervals which allow for the degree of statistical power to become evident; the range of probability estimates renders a much more plausible scenario than point hypotheses can (Reichardt & Gollub, 1997; Rindskopf, 1997). A significance test may well indicate significance or lack thereof, but yields no information as to the sizes of these differences between sample and population parameters. H_0 is rejected for instance and that is all that is known. Confidence intervals, however, are able to illustrate by how much the values were divergent from zero. Figure 61 illustrates this point. All four confidence intervals yield information beyond the confines of "H₀ is rejected".

- Line A contains an interval with a different range as well as a different length from zero to that of line B although statistically the two are equivalent in terms of H_0 being rejected. Line A's confidence interval is closer to zero than is line B which means that Line A's parameter values are closer to zero than Line B. This difference in size is an unknown in point estimate tests even though both are significant
- Lines C and D are both non significant. However, the parameter estimates are much closer to zero in Line C than they are in line D
- Statistically this makes no difference at all as NHST either accepts or rejects H_0 . However, the practical significance of this extra information makes interpretation of data much more amenable to discussion especially in an area fraught with difficulties in inferential conclusion. This reiterates the difference between the statistical and substantial hypotheses

Figure 61 Information provided by confidence intervals which are not provided by point estimate statistical tests (Reichardt & Gollub, 1997, p.273)





Steiger and Fouladi (1997) state four main reasons as to the reluctance of psychologists to employ confidence intervals with which Reichardt and Gollob (1997) are in agreement. Traditionally NHST is favoured due to the pragmatic concerns which override its utility. For instance, detailing narrow confidence intervals might have the resultant effect of supporting highly statistically significant findings but which are in fact practically trivial. Larger confidence intervals have the opposite effect of seemingly less accuracy. Most psychologists are unaware of confidence interval procedures and lastly, a few such innovative procedures are not included in major statistical packages. Relying on either one of the two methods of reporting significance would be less beneficial than perhaps reporting both (Reichardt & Gollob, 1997), as it has been noted that, at times, it is necessary to report results utilising stringent point estimates.

Loftus (1996) also maintains that due to some natural science disciplines' strong form of theory testing, attempting to falsify H_0 is usually beneficial and easier to do due to point hypothesis values that are assigned to the hypothesis. This is as a result of the stronger form of theory within the natural sciences (Meehl, 1998). Also, researchers should plot data as opposed to providing the information in the form of tables with F and p values which are more cumbersome to read. Huysamen (2005) does, however, present a cogent argument for the equally cautious use of confidence intervals and notes that although much is said against the designated arbitrary level of α not much is said about the equally arbitrary level of confidence intervals which are typically set at 95% and 99%. The reason for highlighting the NHST debate is not to castigate the method's utility value (especially in certain instances where its value is noted) but to highlight issues which have permeated through to areas such as dynamic assessment (among many other areas of social science research endeavour). Graphically displayed confidence intervals will better illustrate how closely the observed means of the sample data reflect the pattern of population means and hence confidence intervals serve as guides to the degree of acceptance of null hypotheses. Graphical illustrations of data in general is emphasised as an idea worth pursuing, prompted by the inadequacies of statistical decision theory (significance testing) to more accurately evidence what the data is saying (Brown, Hendrix & Williams, 2003). Confidence intervals attest to the chance of locating the true population means usually to within a 95% accuracy, so the larger the interval the less accurate the data and vice versa. NHST has a habit of imposing the "illusion of certainty on a domain that is inherently ambiguous" (Loftus, 1996, p.168) and by illustrating the data with accompanying confidence intervals one is better equipped to determine the degree of rigid acceptance or rejection of observed data. Harlow (1997), Kline (2004), Loftus (1996), Rossi (1997) as well as Schmidt and Hunter (1997) moreover advocate the use of meta-analysis as technique for eschewing traditional NHST over multiple studies (Kline, 2004), although Oakes (1986) has serious misgivings about the use of meta-analysis in this regard, maintaining that it perpetuates certain incorrectly held NHST notions.

Abelson (1997a, 1997b), Harlow (1997), Huysamen (2005), Nickerson (2000) as well as Steiger and Fouladi (1997) mention a few more options open to those wishing to peruse data from other alternatives available, such as by employing a model-fitting approach to data interpretation in which the goodness-of-fit is tested via the model of $data = model + residual (error)$. Errors are more easily picked up in this method as opposed to NHST. This mode of inference adopts both the falsificationist (strong form of theory testing; i.e. creating defeatable hypotheses; Harlow, 1997) option of NHST in addition to confidence intervals' parameter estimation, although this method is highly influenced by sample size. Bayesian statistics, as another complementary statistical technique is offered as means of data interpretation. As has been mentioned, the crux of the Bayesian method involves assigning a prior probability to the null hypothesis as being true and by adding the research data to the prior probability, the posterior probability of the null hypothesis being true is obtained. Immediately though, one can critically state that the choice of prior probability assignment is open to debate which is where the standardisation of NHST comes to the fore in its advantageous nature. Nevertheless, once the posterior probability of the null hypothesis is ascertained the result can be utilised as a prior probability in another study and over time and over a number of studies the procedure is self-correcting and is able to negate any "damage" incurred by arbitrary human choice of prior probabilities (which is the one main feature of Bayesian statistics against which criticisms have been lodged; Huber, 2002).

Huysamen (2005) not only offers alternatives to NHST but also offers recommendations as to complementing the technique as opposed to doing away with it entirely:

- Replace the nil null hypothesis with the non-nil hypothesis (a directional null hypothesis; thus including a range). In doing so, more information than just a rejection of a value of zero will be available, although if used in a sound manner, the point nil null at least brings a definitive probable point of rejection or acceptance (McDonald, 1997). The nil null hypothesis is often set up as a straw-man argument within the social sciences making the conclusion all the more scientifically useless (Kline, 2004; Oakes, 1986)
- Sample size determination via the analysis of power
 - such that effect sizes considered significant are detectable
 - yet to limit the amount of power which would detect effect sizes that are negligible
 - however, a prior decision needs to be made as to the size of α , the chosen level of power and the size of effects that will be detected by NHST. Tables stating these criteria are available

- At the very least if all else proves impossible, detail the effect sizes as well as NHST findings

Oakes (1986) summarises the fundamental characteristics of the various schools of statistical inferences and an attenuated table is reproduced in table 17 below.

Table 17 Comparison of schools of statistical inference (Oakes, 1986)

A comparison of schools of statistical inference							
School	Interpretation of probability	Emphasis on decision or inference	Emphasis on testing or estimation	Use of prior information	Initial or final precision	Location and nature of inference	Sensitivity to sampling procedure
Neyman-Pearson	Frequentist	Decision	Both	Informal	Initial	Probability statement on sample domain	Sensitive
Fisher	Fiducial ⁷⁹	Inference	Both	Informal	Final	Probability statement on parameter domain	Variable
Bayesian	Subjective	Inference	Estimation	Formal	Final	Probability statement on parameter domain	Insensitive

4.3.2 Summary

Psychology's over-reliance on statistical modelling and corrective techniques has long been considered by critics as testament to its perceived unsatisfactory status as science. This is erroneous thinking and has led to the, at times, unfortunate misuse of well thought out and mathematically sound statistical tools. Parallel to the developments within mathematics as formal abstract science arose statistical modelling which can be dated to early seventeenth century use. Utilisation of statistical techniques has co-evolved within psychological measurement and it is due, today, to the psychometric enterprise that many statistical models are used in other fields outside the domain of psychological testing. This is one of the very few occasions in psychology's life-span during which the discipline was able to give something to outside domains as opposed to continually taking from others. Statistical modelling has become suffused with the notions of probability and certainty, two fundamental criteria which it most certainly cannot avail of itself. This is an error and it has yet to be corrected although a substantial minority of advocates posit forth measures to correct for this error. Dynamic assessment is not blameless in its wholesale use of current statistical techniques which it uses to bolster its findings. It has been argued, however, that in order to remain relevant to the field and in order to help ensure its existence, it has had to resort to mainstream options and statistical utility is one such way. The most significant component of the discipline of statistics which grips psychological assessment is that of inference and the degree to which results are applicable to populations and how results are due to chance effects. This obsession with chance and population parameters is, interestingly enough, not a concern within the natural sciences.

To further the cause of statistical modelling and the requisite need to ensure correct levels of chance findings and its applicability to the broader population, hypothesis testing has grown and become almost the defining criteria for the existence of a statistical psychology. The discussion on statistical and mathematical underpinnings thus leveraging measurement in psychology cannot be divorced from reigning trends of science and philosophies of science. The use of statistics does not occur in a vacuum, and these impinging issues were highlighted in chapter 3. Although major bodies have looked into the question of ill-fitting statistics within the domain of psychology and other critics doing so for dynamic assessment, the full effects of these messages have yet to seep through and as with any discipline, psychological assessment is contingent upon historical contexts which in this case is peppered with a striving towards natural science models culminating in the use of null hypothesis significance testing (NHST). Originating in far flung fields such as agriculture and breweries, statistics was imported into domains and territories which, at the time, seemed to warrant their introduction. Psychology was a growing formal discipline barely twenty years old when great developments were being made within the field of statistics, so it is hardly surprising that the new techniques were set upon and factored into research designs.

⁷⁹ A measure of rational belief. Fisher does not outright reject or accept H_0 based on the data but suspends judgement.



Ronald Fisher, Egon Pearson and Jerzy Neyman are the names which punctuate the story of social science statistics particularly the concern with hypothesis testing. Due to historical events and the erroneous intertwinement of Fisherian experimentation and Neyman-Pearson hypothesis testing, much of psychology's later research and ensuing directionality was falsely skewed, mostly because of negligence of later psychologists' unwillingness or lack of knowledge in correcting the status of the discipline as it stood regarding hypothesis testing in the early 1940-1950's. Due to this negligence, a continuation of ill-fitting methods is still utilised, in areas of concern to this study, namely dynamic assessment. Alternatives to these trajectories have been explored, namely Bayesian inference, but little has come of it in mainstream assessment. The issue of probability and resultant inference is treated effectively yet differently within Bayesian inference and there is a call to utilise the method as opposed to the treatment of probability within the more classical approach as it has been argued, Bayesian inference is more in keeping with real-world psychological issues. The role of historical contingency plays out in the area of classical statistic vs. subjectivist statistics, after all Bayes wrote his works in the eighteenth century and thus could not champion his stance as could Fisher, at a time which was ripe for interference of this nature. In keeping with the tenets of natural science philosophy, both forms of NHST are assessed in terms of their weakness and strengths in determining how results are utilised within psychology, namely the weak and strong forms of NHST. Currently, the weak form is holding fast in a domain where it should relinquish its outmoded and long overdue hold on psychology, but this can only be done if researchers become aware of events. In furthering its own agenda dynamic assessment should move towards alternative measures of research design so that it no longer has to be accountable to a method which is seriously and philosophically flawed and largely misunderstood. It would in fact gain and not lose out by shifting its focus in this manner. Creating a false sense of certainty via statistical utilisation of flawed techniques is due to psychology's over-emphasis on striving towards a natural science orientation. But this can be corrected in ways which nevertheless remain robust yet different.

4.4 Measurement foundation

Prelude

Psychological measurement begins with theory extending from data and ends with inferences of stimuli and people from the data. Assessment has grown from measurement and the two are not synonymous (Meier, 1994). Ellis (1966) maintains that measurement is the link between mathematics and science, two aspects which have thus far been discussed. At least this is the formal understanding as given by Coombs (1967). If only the situation was that straightforward. Before measurement is even promulgated at the level of scaling, such activities are already a consequence of behavioural theory (Coombs, 1967). The fact that we assume constructs to be measurable is telling of two things: that behaviour is in some sense measurable (is it?) and /or the need to fulfil the accepted notion of science progress which inevitably entails measurement of some kind. Dynamic assessment's change-based philosophy of assessment strikes a disconsolate chord for those practitioners more inclined to view stable prediction as the pinnacle of test evolution (Ghiselli, Campbell, Zedeck, 1981). It is here, at the interface of educational assessment and prediction that the two aspects are conceived of as opposing rather than complementary forces (Bereiter, 1962; Biesheuvel, 1972). Changeability is the sought after behavioural aspect within dynamic assessment philosophy which, although at odds with mainstream edumetric assessment, seeks to make valid the idea of change within a robust psychometric framework. Mainstream intelligence assessment is informed from a framework encompassing a myriad of impinging variables or as Ceci, Rosenblum and Kumpf (1998, p.299) state "a galaxy of factors". Variable notions of environmental, biogenetic, ecological and behavioural factors are considered as contributory aspects constituting the holistic concept of what is popularly understood to be intelligence. One influential model of intelligence which emphasises a general underlying structure is of particular importance in this discussion on the measurement foundations upon which rest dynamic assessment encompassed within the broader arena of intelligence assessment. The notion of a general intelligence factor, g , underscores the stability model of unchanging scores in tests of mental ability.

Dynamic assessment's change based outlook is contrary to this very model of classical stability. To further deliberate on this issue a brief digression into the validity of practice effects is warranted. Although practice effects can hardly be equated with directed mediatory interventions of the sort espoused within dynamic assessment, the degree to which test scores can be altered due to practice effects may be of concern primarily because the underlying purported unchangeable g is not variable (Reeve & Lam, 2005). Practice effects seem to superficially alter scores within repeated measures designs, such as evidenced by Reeve and Lam (2005). Employing a host of statistical techniques including multi-group confirmatory factor analysis to test for measurement invariance as well as scalar invariance and item uniqueness across testing situations, the authors illustrated that g -based variances did not vary across tests. Reeve and Lam (2005) convincingly argued that the non-cognitive aspects which were being enhanced via practice sessions were not related to the general factor intelligence. Also no psychometric property of the instrument was being changed in any way due to these practice effects. What would be useful in this author's opinion would be to conduct this study utilising dynamic assessment as intervention variable (akin to the practice effect) and to subsequently analyse the output in the manner described by Reeve and Lam (2005). How would researchers argue lack of g change? Or would g change? If g changed this would entail psychometric property variance which would result in an overhaul of basic classical test theory upon which many such tests reside (Reeve & Lam, 2005). The issue of classical test theory and it's more modern counterpart, modern test theory, will be discussed below in conjunction with dynamic assessment's placement within the



two approaches. Cronbach and Furby's (1970) difference score can perhaps be equated with the Reeve and Lam's (2005) practice effect as can the former's change score be related to true growth and maturation over tests. Although this still brings into question the notion of the variability of g as it has been statistically shown that test-retest score changes are not g -loaded (Coyle, 2006) which may not be such good news for dynamic assessment initiatives. Such entanglements are indeed unresolved knotty problems which need to be taken apart and studied from as many angles as possible, hence the necessity of including chapter 2's discussion on the physiological contributions to the study of g and related issues.

Making manifest latent traits is the rationale underlying the need for measurement (Ghiselli et al., 1981) but the determination of what exactly is meant by latent trait very much hinges on the core philosophy attending such conceptualisation (Borsboom, Mellenbergh & Van Heerden, 2003). These authors contend that latent traits can only be considered within a realist framework as this interpretation is the only one that will suffice in terms of accounting for a causal fit between formal-theoretical and operational-empirical concepts of latent constructs. This issue is highlighted in the ACFS battery for children (Lidz, 2000b) in section 5.2.10. Moreover, Borsboom, Mellenbergh and Van Heerden (2004) are not in agreement with Cronbach and Meehl's (1955) conceptualisation of construct validity as being dependent on the nomological network in which it supposedly occurs. It either exists or it does not. Validity is more often than not assumed to refer to what the test measures but is in fact a reflection of the test scores themselves or the subsequent test score interpretations; constructs are not representative of test scores, rather the construct becomes manifest through interpretation of test scores and then perhaps only partially so (Borsboom, Mellenbergh & Van Heerden, 2004; Borsboom, Van Heerden & Mellenbergh, 2003; Suen, 1990). As there are a variety of interpretations surrounding test theory, there is also no consensus on what validity is. Semantically, validity will mean different things within different test theories. IQ results are the process of summing scores on multidimensional scales which are erroneously carried out on unidimensional scales. This results in a misunderstanding of what is meant by IQ. Moreover, whether or not IQ is related to these measures is also questioned. The tools of the trade are perhaps not to blame when assessments fail to accurately predict scores or resultant behaviour. As has been discussed above in the sections dealing with the mathematical and statistical foundations, the tools themselves cannot necessarily always be held accountable for errors that may at times be present within conclusions. Mathematical modelling of so-called lawful phenomena within psychometrics is an erroneous position from which to build more foundation. There is nothing suspect about developing mathematical models to aid in the understanding of data (Coombs, Dawes & Tversky, 1970) but the issues of lawfulness is in question. The concern running throughout chapter 4 is the original rationale for performing tests in the first place. Deriving statistical and mathematical models from first principles and providing what can at times be considered as proofs is not to be shunned or looked upon as trifling contributions to the science of psychological assessment and progress of the discipline. The question is the very need to do so in the first place. Are the very psychological latent constructs themselves amenable to quantification at all (Borsboom & Mellenbergh, 2004; Michell, 2001, 2004)?

Whether or not the construct "intelligence" exists is in any event not a matter for psychometric modelling or technique but is a question for substantive theory and will not be solved by psychometrics alone (Borsboom, Mellenbergh & Van Heerden, 2004). Theories of mental testing accommodate developments within cognitive and learning theory (Dillon, 1997) as is evidenced with newer IRT models encompassing change in their structure. Operationalising concepts is hardly a means of magically transforming latent traits into manifest quantity. The measurement foundation section will look at basic history and philosophy of measurement as it pertains to the psychosocial sphere even though it has been almost wholly informed from natural science rigour (Savage & Ehrlich, 1992). It will highlight the plight of dynamic assessment as method of assessment which, due to historical contingency, had to devise and uphold a measurement strategy in keeping with mainstream requirements in terms of reliability, validity and change score stability. Clearly proved and impeccably well thought out mathematical models of various test theories along with statistical techniques with which to manipulate, constrain and free the data can unfortunately do nothing to first principles which state that aspects are quantifiable when there is clearly no proof that this is the case (Michell, 2001).

Dynamic assessment currently dances to the tune of mainstream intelligence assessment models and current psychometric theory. Mainstream perception also dictates that the only future in which it can adequately serve a function, is one in which progress will be made along a continuum which it is currently following. Why does dynamic assessment have to fit in with mainstream psychometric theory in the first place and secondly why does it have to adhere to mainstream intelligence assessment models and theory? Why is it necessary that it should envision for itself a place within a hierarchy on intelligence measurement at all? Can it not do the following and still maintain its scientific credibility, after all, science is not measurement! Measurement is one of many facets endearing the method of science to its followers. Measurement does not make or break scientific method and hence neither should it do so for psychology, which unfortunately is precisely what is has been doing since its formal inception into the domain of science:

- Dynamic assessment already possesses its own unique theories and models of intelligence
- It should refrain from selling itself short as relevant model of change-assessment model
- Current intelligence research cannot even avail of its own track record a definition which is usable from one model to the next much less adapt from one testing situation to the next
- The logic is:



- If intelligence assessment and measurement is well nigh hopeless in its current state, then
- Why should dynamic assessment which has its own philosophy and history seek to follow in the wake of intelligence measurement and assessment as currently practiced?
- Dynamic assessment should carve out for itself a new vision and path to follow, one which encompasses what intelligence assessment has been unable (or unwilling) to do:
 - Follow on from its own repertoire of knowledge gathered since its inception
 - Predicate its trajectory on change-based assessment
 - Cease to work with the outmoded and weakly defined intelligence measurement procedures
 - Employ non-quantitative measures which nevertheless adhere to science as commonly understood by and within the community of philosophers and practitioners of science. Merely construing a construct as measurable does not necessarily mean that the construct is being measured, assuming that it exists as construed⁸⁰
 - Employ techniques such as conjoint measurement in such a manner as to allow for the utility of intensive measures which parallel the extensive measures utilised within the natural sciences
 - Divorce itself from current haggings in intelligence assessment in terms of construct validity, issues which, as history has so eloquently illustrated, is no nearer resolution that it was over a hundred years ago⁸¹
 - Predicating constructs on supposed correlated constructs is circular reasoning which dynamic assessment should not seek to replicate (added to this malaise is the very real concern of knowing very little of the original construct (intelligence) in the first place; Oberhauer, 2005)⁸²
 - The time has come to forge a new path and to leave the one well-trodden in its wake

The following will substantiate and motivate the impassioned reasoning in the above bulleted concerns.

4.4.1 Elements of measurement

Figures 34 and 54 in sections 3.4.2.1 and 4.1.1 respectively are instances where rules of translation are necessitated if transformations from one realm are to be made into another associated realm. Figure 34 utilises what was referred to as correspondence rules or bridge principles which served as translation-transformation functions allowing for information at one level (a concept of learning potential for instance) to be encoded at another level (conductivity of neural speed). Likewise, figure 54 illustrated the concept of isomorphism for a two-layered system of physical attribute-to-calculus transformation via a process of coordinating rules and rules of translation. Similarly, measures of psychological attributes are designated rules of transformation via scales of sorts like those epitomised by Stevens (1946) which is essentially a representationalist theory of test measures (Borsboom, 2005). Stevens' scales of measurement were not arbitrary in the sense of lack of insight into the characteristics inherent in the scales. It is the seeming lack of critical forethought into the issue of quantifiable construct which is made to map onto his scales that is arbitrary. As discussed in chapter 3, psychology's foundation as formal science has ineluctably fostered a rigorous natural science framework for its future development and has engendered a philosophy of quantification as core to this enterprise. Stevens wrote much of his work in the 1930's - 1960's and one cannot hold him responsible for being entrenched in a time and place where psychology's future as "scientific" was an almost given; a time which was, before Luce and Tukey's 1964 article, dominated by the conformity of numbers to scales which itself was representative of supposed relations between observables and numbers (Cliff, 1992). However, fifty years later, the need to re-look philosophical issues of prime consideration is necessary. It is ironic to note that measurement theories, competing with the received view, were developing in parallel and that work was published in various areas in an effort to give to psychometrics what was lacking in more traditional views of measurement.

There are a number of reasons as to why these parallel developments did not take hold as firmly as had the traditional views which by the early 1940's were rooted in mainstream discourse. Among other reasons cited are lack of computational power with which to carry out large scale and power-hungry sub-routines; the somewhat bare and abstract nature of axiomatic measurement theory; the lack of a cadre of psychologists who were mathematically and statistically able to follow the logic of what was being propounded in various avenues of measurement theory (Kline, 1998; McDonald, 1999; Schönemann, 1994)

⁸⁰ The more one ponders this situation the more strikingly absurd and silly it becomes. Can it be that the whole enterprise is based on such silly notions and ill-conceived logic? Upon this, we build ever more grandiose and sophisticated statistical models which we use to help ourselves out of theoretically questionable findings. Do we blame the statisticians or the psychologists who continue blindly with such methods? We use the tools "because they work"; but never seem to question their foundations.

⁸¹ This is not to say that there is no consensus at all when it comes to defining constructs such as intelligence (Owen, 1998). But the leap from informal agreement to formally posited unanimous agreement has yet to be made.

⁸² "The only way out is going through this circle again and again, each time refining the measurement instruments in light of more precise theoretical formulations and refining the theory in light of experience with current measurement instruments" (Oberhauer, 2005, p.393). Here it is patently evident that practice informs theory which informs practice, a theme resonating from chapter 3. Yet the limitations are pervasive and keep us within the system of known theory and measurement instruments. It is a necessary task to move away from these constraints into new territory but this is far easier said than done.

especially mathematically more complex item response theory (Sijtsma, 1993a) and the concomitant lack of appreciation for other means of obtaining information in a scientific manner yet remaining true to lived experiences (i.e. quantifiable yet non-numeric). This notion of quantifiable non-numeric measure can be seen in many dynamic assessment models where an attempt is often made to secure quantifiable scores for qualitative behaviour. Ultimately, such scores are numericised according to the strictures of normative testing, but there is an uneasy feeling surrounding the need to quantify for the sake of allowing the model a right to exist within mainstream assessment.

Assigning numbers or numerals or at least quantifying objects, events or notions by way of numerics presupposes that such objects, events and notions are quantifiable in the first place. Once such objects, events and notions are numerically quantifiable the deployment of various scales of measurement is, at their own rules of engagement, not an illogical path to follow. However, such numerical assignments and the operations carried out are bound by various rules expanded upon at great length in many texts dealing with measurement theory. A number of axioms are upheld when seeking to manipulate quantified objects, which, if adhered to, allow for mathematically acceptable notions of quantification and subsequent manipulation of these notions. Measurement theory as propounded and followed in the natural sciences is coherently defined and utilised, but its use within a domain which clearly does not stand in the same realm as that of the natural sciences needs to be questioned and critically assessed as its veracity as tool of correspondence may be significantly different. Rules of measurement ensure that such measures are adequately entitled to serve as measurement representations. Coombs, Dawes & Tversky (1970) and Pawlowski (1980) discuss four problems which present within measurement theory which take to task key assumptions within such a theory, namely:

- The representation problem
 - The questions
 - What can be measured? It is indeed pointed irony that intelligence is being measured without a universally defined understanding of what “it” is
 - What conditions will suffice for measurement to take place?
 - What rules are employed to ensure consistency throughout measures?
 - The discussion
 - Extensive attributes are measurable and possess additive structure. Intensive non-quantifiable structures present problems as they cannot be concatenated
 - A formal system in which correspondence rules allow for assumptions to be made regarding the assignment of numbers to events can be logically derived and deductively deduced. This set-up will follow a relational structure in which formal properties determine the type of relation occurring between the elements of a set. “The process of modeling and measurement are described as representations of empirical systems by formal ones” (Coombs, Dawes & Tversky, 1970, p.11). The authors continue to state the relationship between two systems as such:
 - ⇒ A system $\alpha = \langle A, R \rangle$ can be represented by another system $\beta = \langle B, S \rangle$ if there exists a function f from A into B (which assigns to each x in A a unique $f(x)$ in B) such that for all x, y in A :

$$x R y \text{ implies } f(x) S f(y)$$
 - ⇒ In essence, the relation that holds between x and y via R can be mapped onto the relation that holds between $f(x)$ and $f(y)$ via S . If the two systems are representative of each other such that α and β map onto one another then the two systems are said to be isomorphic (this is delineated below in section 4.4.2)
 - ⇒ If the model imposed is numerical then the process can be considered as measurement but numerical assignment to objects does not necessarily mean that the system is measurable. One needs to be able to prove that such representation follows rules in accordance with the above. For instance assigning numbers arbitrarily to people based on how early they arrive at a film cannot be included as measures
 - ⇒ Measures have to be transitive in order to be concatenated and not all measures (such as attitude or preference) are transitive. A person prefers x to y and y to z and z to x . Clearly the transitivity of this relation has been violated in the strict sense of number-to-object relation
 - ⇒ The interaction between formal and empirical analysis is the hallmark of measurement in science
 - ⇒ The question is whether psychological constructs are isomorphic to systems purporting to represent them
- The uniqueness problem
 - The questions



- How much freedom is there when assigning numbers to entities (assuming of course these entities can be measured)?
- Is the choice of number assignment arbitrary or is it dictated by the measurement process itself?
- The discussion
 - Three types of uniqueness problems present: the mapping of empirical structure into a unique numerical structure; how suitable various numerical models are for such mapping of the same empirical structure and lastly how to estimate response probabilities and the concomitant issue of how the unique estimation is dependent on a finite response set (Irtel, 1994)
 - In assigning numbers to objects utilising various scales, certain aspects need to be kept in mind
 - Ordinal measures are order preserving and the assignment of numbers is arbitrary in the sense that as long as the order-preserving function is maintained the scale is usable (this is a mapping from one empirical reality to another representational reality; Maxwell & Delaney, 1985; Narens & Luce, 1986). Scales are monotonically transformed if this order is preserved and any two scales which are so preserved can be said to be related. For instance:
 - ⇒ If a person prefers x to y and y to z and x to z then any number can be assigned to the values of preferences provided they preserve the ordering (note that these preferences are transitive and are thus amenable to this type of numerical assignment). We can, to all intents and purposes, assign values of 7.5 to x ; 0.0054 to y and 0.0023 to z or 1365 to x , 154 to y and 0.25 to z . Both scenarios preserve the order of $x > y > z$
 - The above example cannot hold for intervals measures however, as this scale concerns itself with the added feature of preserving the interval between successive numbers. Such scales are possible up to positively linear transformations. For instance:
 - ⇒ If preference order as well as interval magnitude is to be retained in transformations then, using the above example for $x > y > z$, it should be shown that $x - y = y - z$ assuming that the intervals are equidistant. Also, $2y = x + z$ should hold. Transforming to interval scale the notation becomes $u(x) > u(y) > u(z)$, it should be shown that $u(x) - u(y) = u(y) - u(z)$ assuming that the intervals are equidistant. Also, $2u(y) = u(x) + u(z)$ should hold. Transforming into another scale maintaining these two considerations, $u(x) = a$ and $u(y) = b$ for instance, hence $u(z) = 2b - a$
 - Absolute scales do not allow for transformations and an instance of this would be counting when viewed as measurement. For instance $3 \neq 4$ and cannot be transformed in any way without violating its inherent meaning
 - The scale type is determined by the admissible transformations. Note that latent trait theory does not avail of additive representation in the strict sense of representing response probabilities directly. Latent trait theory utilises mapping functions which restrict the range to an interval of (0, 1) an example being Rasch's logistic function (Irtel, 1994)
- The meaningfulness problem
 - The questions
 - What are the different inferences that can be made when utilising different scales of measures?
 - What decisions can be made following on from the measures?
 - The discussion
 - Inferences based on scales should be invariant across admissible transformations. If worthwhile information is to be sought from scales, limits inherent in scale properties have to be taken into account. The truth or falsity embedded within measured events need to remain invariant across transformations otherwise these values change, which is pointless. Events amenable to transformations yet preservative of their truth status necessitates that the event and scale are congruent. For instance, ordering fruit according to the time at which they are placed on the table is an arbitrary number-assigning process. Each piece of fruit is assigned a number based on its order of placement. It is nonsensical to state that the "pear is twice the apple" because the pear happened to be assigned the number 4 while the apple was assigned the number 2. The interval between them is not transitive
 - As Coombs, Dawes and Tversky (1970, p.17) state "a more difficult problem arises with respect to a statement involving numerical values for which no explicit measurement model exists. The measurement of intelligence is a case in point". The question then posed is: how can measurement be possible if number assignment does not follow any of the rules described above? The answer to this is three-fold and entails the following:



- ⇒ Prediction - in which the elusive dependent variable cannot be identified unless it is predicted for by concurrent measures yielding statistical (and hence substantive⁸³) correlations. It is partialled or factored out as existing
- ⇒ Description - such statistics are only informative so far as they elucidate or infer the underlying construct which for obvious reasons is very limited in descriptive statistics
- ⇒ Direct assignment - most often utilised in psychological research are scales where direct assignment of numbers occurs. Rating, category and magnitude discrimination are such scales which make no use of any rules as described above as there are no measurement models as such which have been developed (however see section 4.4.2.3 on IRT test theory). They are usually treated as nominal and ordinal scales
- The scaling problem
 - The questions
 - What is involved in developing a scale?
 - How many types of scales are there?
 - How is error dealt with in these various scales?
 - The discussion
 - Campbell's (1920) treatise on measurement stipulated the lack of progress of psychological measurement due to the fact that only extensive measures are amenable to mapping onto interval scales and since only intensive measures exist for psychological constructs, no psychological constructs are interval scalable. However measurement in the form of derived additive conjoint measurement makes it possible (Brogden, 1977; Luce & Tukey, 1964; Michell, 2003a; Perline, Wright & Wainer, 1979) to measure psychological constructs derived from axiomatic principles (Velleman & Wilkinson, 1994) (see section on conjoint measurement below). Scales, other than those advocated and discussed by Stevens (1946) exist (Velleman & Wilkinson, 1994). Scaling techniques and the manner in which error is treated will be discussed separately in the measurement models below

4.4.1.1 Extensive and intensive measures

The natural sciences pride their disciplinary success on primitive or extensive measures which are amenable to concatenation (they possess additive structure); hence quantification is paramount (Kline, 1998). Michell (2001) and Tyron (1996) echo Lazarsfeld (1977) in their opinions regarding the necessity and utility value of quantifying units of "things" or "stuff". The usual idea of measurement progression stems from the seemingly intuitive idea of the flow of number to quantity to measurement (Michell, 2001). Basic philosophy underpinning the growth of a psychological science meant that in order for a secure place to be found for psychology as robust discipline nothing short of physical attainment of concepts would be accepted. Lykkens' (1991) citing of Richard Feynman's "cargo cult"⁸⁴ syndrome comes to the fore in this particular instance. Psychology, as formal discipline, might look right, might have everything in place and be set to go. Alas, the various endeavours never seem to get off the ground at the conceptualisation stage. Attempts to salvage for the discipline some respectability and thereby engaging in some serious damage control in terms of its image results in natural science methodology being employed (Berka, 1992). This turn of events has subsequently only served to further entrench the discipline into a quagmire of unease. However unpalatable an equating of natural science concepts to social concepts is, a concerted effort in this regard is proffered by Tyron (1996) for instance who envisions a future of numerically assigned units of measures worthy of natural science respect.

Let us, for the moment regard just such a future in which concepts, as borrowed from the natural sciences are indeed implemented in the social science sphere. Taking the lead from primitive extensives (fundamental) from which all other ratios are derived, Tyron (1996) seeks to build for psychology a knowledge hierarchy similar to that which is in place for the natural sciences. Two fundamental extensive theoretical quantities in the natural sciences upon which most other quantities are derived are length and time (Barrett, 1998; Ellis, 1992) defined exclusively in terms of the number of wavelengths emitted from krypton-86 (defining the metre) and the number of transitions between two energy levels of cesium-133 atom (time). Units are assigned to these measures and find unanimous agreement throughout the physics world. This is not to say that length and time could not be defined in other terms and had physics developed in some other manner this might well have been the case (Michell, 2001). Derived from these two extensive measures are intensive measures or ratios expressing these primitives in terms of other measures such as mass, density, volume and area which are expressed as the products of two or more extensives (Domoter,

⁸³ Recall that the statistical and substantive hypothesis or construct is not necessarily one and the same thing!

⁸⁴ Feynman's "cargo cult" syndrome is an apt description of the mammoth task psychology faces in progressing beyond its current situation. Briefly, the South Seas cargo cult peoples during the second world war witnessed an unprecedented number of airplane landings which delivered favoured goods. In an attempt to bring the scenario to life once the war had ended they reconstructed the exact series of events such as would allow for the planes to land again. But to no avail, no matter how perfectly the scene had been laid, no planes ever landed. Psychology, it seems, is suffering the very same delusions. Everything is in place. So why is nothing happening?

1992). Derived concepts are predicated on primitives which have been clearly defined and as psychology has no purported extensive measures, perhaps as a formal discipline it can predicate itself on derived measures (or so the logic proceeds) (Schönemann, 1994).

Length and time are definable due to explicit agreement on objective criteria of measurement. The fact that these concepts can be measured is indeed perhaps the most noticeable aspect of the measurement enterprise, at least within the world of physics but the same cannot necessarily be said of psychology for which measured constructs do not always exist (Maraun, 1998). Numerical assignation does not entail that measurement cannot take place however and it also does not preclude quantitative measurement (nonnumerical quantitative objects or events), as Euclid's geometric axioms clearly show; these are utilised in nonnumerical ratios (Burgess, 1992; Koslow, 1992; Savage & Ehrlich, 1992). The assignment of numbers to events is only one such manner of measurement and is unfortunately the one dominant trend within measurement due almost entirely to the persuasive arguments detailed by Stevens (see below) (Cliff, 1992). Originating with physicist Fechner's (1860) publication on the elements of psychophysics and the research into reaction times, ca. 1862; Ebbinghaus' (1885) work on learning as well as Galton's work in individual differences; Helmholtz's (1887) treatise on counting and measuring followed by Hölder's (1901) treatise on axiomatic quantity and mass and Campbell's (1920) work on fundamental extensive measurement; Bridgman's 1931 discourse on dimensional analysis as well as Steven's (1946) treatment of the scales of measurement it was only in the 1960's that other views espousing measurement made their presence felt. For instance Luce and Tukey (1964) (Boring, 1961; Ellis, 1966; Krantz, 1964; Luce & Tukey, 1964; McGrath, 2005; Michell, 1999; Narens & Luce, 1986). The latter's approach towards measurement as well as the under appreciated works of Luce, Krantz, Suppes and Tversky (1971, 1990) has not taken as firm a hold on the measurement community as has the formers' representative works (Balzer, 1992; Cliff, 1992; Kyburg, 1992; Savage & Ehrlich, 1992; Schönemann, 1994).

Tyron's (1996) clearly physically inclined psychological knowledge hierarchy (which he refers to as behavioural physics) encompasses units of measures such as latency measured in time; duration measured in time units; countability measured in cycles; frequency measured in cycles per unit time; celeration⁸⁵ measured in cycles per unit time per unit time (an interesting squared notion) and inter-response times measured in time per cycle. Such a knowledge hierarchy could perhaps be instituted for a quantified realm such as that offered in figure 51 below. Unfortunately there is as yet no agreed upon unit of measurement resulting in numeric ratios of magnitudes on various scales of measurement being in error to an unknown degree (Barrett, 2000). The fundamentals discussed in chapter 3 are pivotal to the discussions taking place in this chapter. Psychology as formal discipline should engage in either a natural science flavoured approach to the study of aspects which are amenable to such renderings or it should not. Construing a nonquantifiable enterprise as pre-scientific (Michell, 2001) is naïve but in order for this view to be bolstered it will have to define smartly what it means by various nonquantifiable aspects. This can be done and moreover can be conducted in an enlightened scientific manner (Michell, 1999). To return to the notion of predicating psychometrics on derived measures (which strictly speaking is not the natural science model at all), the advent of conjoint measurement was brought in to "plug the hole" (Schönemann, 1994). Conjoint measurement would replace the need to derive measures, as it would itself become a fundamental measure. Before turning attention towards conjoint measurement, a brief tour of concatenation is necessitated.

4.4.1.2 Rules of concatenation – the necessary additive structure

Proposing to measure an attribute assumes that the attribute can be assigned a quantity, usually via a number. Defining quantity would then be a logical step to take in allowing the magnitude of quantities to be represented numerically. This allows for both a definition of the kinds of attributes which are quantifiable as well as what measurement is (Michell, 1999). Thirdly, hypothesising an attribute's quantitative nature via observational means enlightens the process of how to quantify. The proven track record of quantification within the natural sciences attests to the necessary concatenation structures inherent within their quantitative constructs but whether this feature of additive constructs can be said to include constructs such as intelligence is questionable (Schönemann, 1994). This discussion on quantification, as explicated by Michell (1999) is indebted to the writings of the German Otto Hölder who set forth his axioms of quantity and theory of mass (Coombs, Dawes & Tversky, 1970). Michell (1999) firmly states that regardless of the mathematics involved in the axioms, the logic of quantification is very much a branch of philosophy and not mathematics and must be treated accordingly. Quantifiable attributes, for instance, may possess length. The fact that an object can be said to be x metres long assumes a numerical relation; one that can stand in comparison to another object of y metres long. In order for this statement to ring true, numerical relations must exist. They do not exist for nonquantifiable objects (and thus, according to Michell (1999) a quantifiable psychology in terms of numerical relations does not exist. He does not state that the science of psychology cannot exist, only that as understood by measurement theory, it is nonquantifiable in most instances). The following exposition is taken from Michell (1999) who has relied on Hölder's axioms for measurement. Hölder's (1901) work was, apart from mathematicians' interest, effectively ignored for over fifty years and was resuscitated by the

⁸⁵ The word used in Tyron's (1996) article is "celeration". The author assumes that he meant something akin to "acceleration" but is not entirely sure of this.



subsequent works of Suppes (1951) and Nagel (1931) (Michell, 1999). Stevens's work, however was not ignored even though he had not explicitly set forth any guidance as to what to look for in measurement (Michell, 2002). Michell (1999, 2002) employs length as an example but any attribute Q can be substituted (Ross, 1964 uses a balance pan to illustrate his examples pertaining to the axioms of additivity). Numerical relations require additivity or an additive structure. Two objects of x and y metres stand in an additive relation to each other provided that:

1. for any lengths, x and y , one and only one of the following is true;
 - a. $x = y$
 - b. there exists z such that $x = y + z$
 - c. there exists z such that $y = x + z$
2. for any lengths z and y , $z + y > z$
3. for any lengths z and y , $z + y = y + z$
4. for any lengths x , y , and z , $x + (y + z) = (x + y) + z$
5. for any length x , there is another y , such that $y < x$
6. for any pair of lengths, x and y , there is another z , such that $z = x + y$
7. for every non-empty class of lengths having an upper bound, there is a least upper bound

The additive relation is a permanent property of these lengths and is independent of what is done to any quantified object. The first condition stipulates that both lengths are identical and, if not, that the difference between them is made up of another length. The second condition stipulates that for two lengths when added together will always result in a summation which is larger than either of the two separate lengths. The third condition stipulates the irrelevance of order of additivity. Likewise, the fourth condition stipulates the irrelevance of the order of the compound additivity. If all lengths have the structure imposed by conditions 1 - 4, then the lengths are additive. However, can all possible measurable lengths be considered? In other words, can other lengths that are not necessarily a part of those admitting to the first four conditions be included? If conditions 5 -7 are upheld then the answer is an affirmative one. The fifth condition stipulates that there is no smallest length as smaller lengths merely keep getting smaller. However condition two stipulates a lower bounded level which is not zero or smaller. Likewise, condition six stipulates that there is no upper bound, as lengths merely increase. Unlike condition five though, where there is a lower boundary, there is no upper boundary for lengths. This is tempered by condition seven which stipulates that all possible lengths are continuous.⁸⁶This concludes the necessary structure if lengths are to be considered measurable; but Borsboom (2005) gets to the heart of the matter and states: "additivity is often desirable because it is simple, but it is only desirable if substantive theory suggests that additivity should hold. Substantive theory *may* prescribe an additive model, but I do not see why it *must* be so" (original emphasis p.116).

Wholesale import of physical notions of measurement might not hold in the psychological realm. The need for additivity in psychometric construct delineation is expressed most cogently in latent trait models of change across measures and the criterion of necessary additivity is illustrated best in the discussion on multidimensional Rasch models for learning and change in section 4.4.2.3 below. Comparing relations between magnitudes, such that $x > y$ for instance, does not yield information directly bearing on the objects per se, but is an arrangement bearing on the relationship of the magnitudes of the two objects. This is perhaps the key point when detailing what can and cannot be identified as quantifiable measure. If a psychological construct such as intelligence is assigned magnitude (and this is not an agreed upon scenario) then a score $x >$ score y only holds in so far as the relation between the magnitudes holds true. This has to be translated back into the empirical realm where $x' > y'$ may not hold as an isomorphic relation. It has commonly and lamentably been accepted that the transformation from empirical notion (x' or y') is isomorphic to measured x and y . This is a gross and unproven assumption. Moreover, these conditions are upheld as magnitudes which need not necessarily lead onto quantification. Frege's conceptualisation of natural numbers as classes of similar classes and of real numbers as ratios of magnitudes is telling of the early understanding of considering ratios of magnitude independent of the quantity or magnitude assigned to individual objects. Whitehead and Russell also followed Frege in this regard. Recall Frege's logicist foundations and his use of letters of the alphabet to define number (see section 4.2.1 above). Narens and Luce (1986) provide three very short yet comprehensive expositions on structure preserving concepts, conjoint structures and concatenation structure (pp.179-180).

4.4.1.3 Conjoint measurement as fundamental extensive

Psychological traits were originally considered immeasurable (cf. Campbell, 1920) and in answer to this criticism conjoint measurement, as nonextensive structure interval-scalable method was offered as partial solution (Narens & Luce, 1986; Perline, Wright & Wainer, 1979) and has been referred to as deep measurement (Narens & Luce, 1986). In order for psychological traits

⁸⁶ The elaboration on the necessity of condition 7 is given in Michell (1999, p.51). See also Ross (1964, chapter 2) for more detailed descriptions of the elements of a philosophy of physical measurement and specifically pp.52-62 for elaborations on the theories of the additive type.



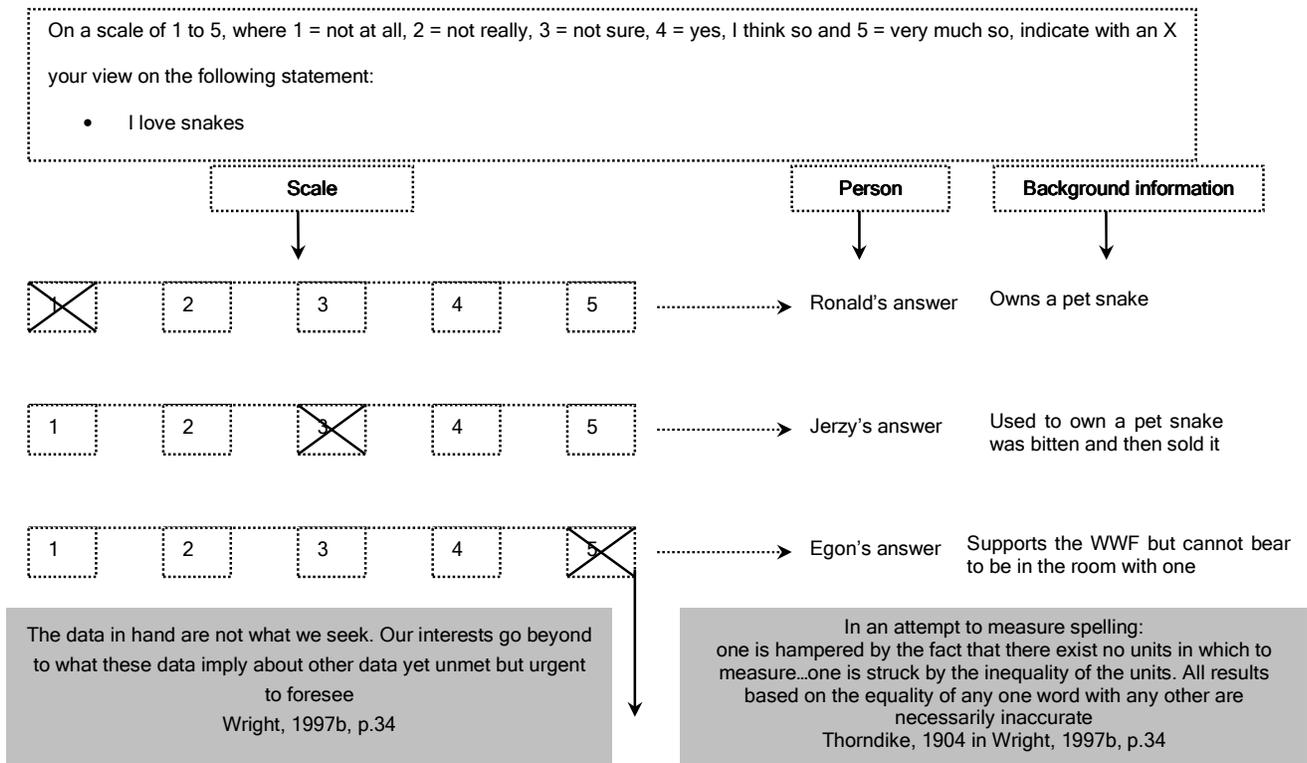
to be measures they would need to avail themselves of more than just ordinal measurement (Michell, 2003a) a scale on which quantitative measurement is suspect. Ordinance denotes order not magnitude and as such the two are hardly commensurate with each other. Recall the above discussion on numerical relations which require additivity or an additive structure. For instance, Thurstone's "law of comparative judgement" in which it is stipulated that an attitude being endorsed more so in one person than in another is evidence of a quantifiable measure is incorrect (Michell, 2003). There is simply no self-evident proof here evidencing that this is the case. That there may be a geometric distance between two points on a line is not contested, but making the leap between different emphasis of endorsing statements and equating those to points on a line is not only unscientific, but simply not thought through.⁸⁷Paul Levy's (1937) proof of indivisibility, which is logarithmically equivalent to conjoint additivity illustrated that stable laws could be constructed even when the decision as to what to count is an arbitrary one, but this requires infinitely divisible parameters (in Wright, 1997b). Rasch (1980) was later to apply the divisibility requirement for stability. An attribute is a simple order only if the following three manifest in terms of the levels' ordering (assuming a continuous variable) (Michell, 2003a):

8. transitive IFF for any a, b, and c, if $a \geq b$ and $b \geq c$, then $a \geq c$
9. antisymmetric IFF for any levels, a and b, if $a \geq b$ and $b \geq a$, then $a = b$ and
10. strongly connected IFF for any levels, a and b, either $a \geq b$ or $b \geq a$

Any continuous quantity is a simple order but a simply ordered attribute is only a quantity if the relation $a+b=c$ exists and satisfies the above-stipulated conditions 1-5. An attribute can fulfil conditions 8-10 but this does not suffice as additive, hence, an ordered attribute is not necessarily an additive one. Psychological measures (ordinal measures) are therefore not quantitative (Michell, 2003a) and without interval scales meaningful statements about differential and developmental psychology will be difficult to make (Jensen, 2005). Thorndike, as far back as 1904, had already recognised this fact in addition to the non-linear nature of raw scores (Wright, 1997b). Figure 62 attempts to illustrate this.

⁸⁷ Making available a line segment with markings of equal intervals in a questionnaire does not solve the problem either! Who is to say that the arbitrarily determined intervals should be so designated?

Figure 62 Attitude measures are not geometrical measures and even if they were they are not quantitative by default



- Clearly, it can be seen that Egon favours snakes to a much greater extent than Ronald. But Ronald owns a pet snake. This already indicates that Ronald's idea of "not at all" is not quite the same as Egon's. That's the first problem: raw scores are not measures (Wright, 1997b). Ordinal measures cannot assume nominally equal intervals on the scale (Cliff, 1991b)
- Egon cannot therefore be said to like snakes five times as much as Ronald purely by looking at the geometric display or units on a line. That's the second problem: additivity as understood from raw scores cannot be
- Jerzy seems to be pretty neutral but he in fact once owned a snake which would make one think that he actually likes them a lot more than he is showing
- Egon in fact is a nature supporter and thus endorses the "highest" level but as for snakes in particular he cannot stand them. He should really have endorsed number 1 or 2 had there been a common metric or scale and that's the third problem
- In essence: comparisons cannot be made because no common metric has been defined and assuming we relax this assumption for the time being what can be said of the additivity of the endorsements? They might be ordinal (but here they are not even that!) but there is nothing to suggest that the results are additive

A related aside - worthy of some consideration.....

Michell's (2003) insights are particularly illuminating. Paraphrasing him, but taking almost directly from his message the following is asserted:

- The distance between two points in multidimensional space is a function of the differences between the points on each dimension
- The simplest scenario would involve two dimensions. In order to ascertain the difference between the two measures, one needs to calculate the difference between the two points on each dimension, square it, add the two together and utilise the positive square root of the sum
- As the number of dimensions increase so too do the number of squared components
- This process is a Euclidian one (see the discussion above on the importance of non-Euclidian space and what it meant to the philosophers of mathematics in terms of the importance and solidity of presumed timeless axioms)
- Here is the crux: Euclidian dimensions work very well indeed for physical objects in such space but there is nothing to substantiate a wholesale import of this methodology into the psychological realm. Who is to say that psychological space is equivalent to Euclidian space? Scientists (mostly mathematical psychologists) seemed to have jumped a chasm without building a bridge to serve as a foundation for allowing them to do so in the first place

- Euclidian distance belongs to a family of distance functions known as Minkowski metrics (1864-1909 and who was highly esteemed by Hilbert; Hey & Walters, 1997)
- Any distance function which takes all the distance components and raises them to the same power, say r , ($r \geq 1$ and is a real number) and takes the r^{th} root of the sum is known as a Minkowski metric
- Psychologists who study attitudes have available two parameters that can be varied; the number of dimensions they think underlie the variable (note that they do not know, they are guessing, albeit an educated guess) and the value of the Minkowski constant, r
- These two values can be adjusted to fit the data at hand
- Due to the 'ease' with which data can be utilised within such a framework, psychologists do not hesitate in doing so and hence feel that abandoning this quantitative imperative would be a step backwards perhaps (and hence not in keeping with a more progressive stance; at least this may be the misperception)
- Psychologists are not, it seems, bothered by the number of questions that can be raised against the use of such multidimensional quantitative hypotheses and as such
- "will remain locked inside their own closed system of quantitative thought" (Michell, 2003, p.24)
- The place for a qualitative psychology is maintained and the overpowering need to remain transfixed within a quantitative system (which is clearly at odds with a number of research directions in psychology, among them, dynamic assessment) is bordering on delusion. Dynamic assessment, as posed throughout this study, should align with a more qualitative approach or if it is to continue on a path of quantification, should pursue measurement tools utilising conjoint measures. To employ "quantitative techniques" not befitting the research questions asked is counter-intuitive

However, the place for conjoint measurement is found in this predicament and interestingly enough, was already in existence as early as 1901 with Hillier's work. This predicament was overcome by the simultaneous derivation of a measure without the need for its intensive measure to be concatenated, a necessary condition for additivity within the physical sciences. Adams and Fagot (1959) (in Schönemann, 1994) first proposed a scaling method of measurement by the simultaneous scaling of two measures; hence co-joint or conjoint measurement. Additive conjoint measurement is the most popular of this type of scaling technique (Schönemann, 1994). This was deemed a suitable technique (albeit an indirect route for identifying additive structure; Michell, 2002) to replace the void left by the lack of concatenation constructs in psychology. What makes this a fundamental measure though is questionable as the logic behind this proceeds on the basis of two measures which are not themselves fundamental and with the joining of the two a fundamental construct arises. However, it is due to the joint distribution of item responses that latent trait models are testable in the first place (Borsboom, 2005). Latent trait modelling cannot be tested directly for any particular item because the underlying independent trait is latent and therefore not known (whereas if CTT posits known *a priori* ability distributions it is unable to tell whether endorsements are due to ability or item functioning), but it can be tested indirectly through the joint probability distribution for the items responses but the model is refuted by one single instance of axiom violation (if double cancellation does not hold for instance; see below) (Borsboom, 2005; Ellis, 1990). In other words, it specifies the conditions under which the structure of the correlation between factors provides information about the underlying attributes (Michell, 2002).

In order to obtain a measurable variable which can be concatenated, two or more joint factors apply to some event which is made manifest via these joint independent variables. For instance, the amount of progress made between pre and posttests covaries with concomitant drops in the rate of questions asked during mediation. Potential, as dependent variable is evident via (i) the amount of progress (answers successfully completed after training) and (ii) the drop in the rate of questions asked or aid sought. Two independent conjoint factors which make manifest this potential might include, as an example, the type of mediation utilised (i) long-term or short-term and (ii) intensive small-scale qualitative or cursory large-scale interventions. "Potential", like its "intelligence" counterpart is a construct which is not isomorphic to quantitative measurement which need not mean that they are not amenable to non-quantifiable measurement (recall Michell, 1999). Coombs, Dawes and Tversky (1970) state that

- by simultaneously measuring both the dependent and independent variables (Anastasi, 1988) and
- assuming that the empirical system is sufficiently rich in these types of measures (which dynamic assessment has yet to make obvious) and
- by axiomatising the conjoined ordinal measures into interval scale measures
- an additive construct is obtained for psychological measures (potential / intelligence) where there was none before. The notion of additive representation takes the place of interactions within analysis of variance
 - The difference between analysis of variance and the additive model is that the former seeks to determine whether the cell means are descriptive of additive combination of their column and row components whereas the latter model seeks to monotonically transform scale values in such a way that the requirement of additivity is adhered to via the transformed cell values. Such transformation will exist dependent on three axioms being satisfied, namely the double cancellation and solvability as well as the

Archimedean conditions⁸⁸ (Kline, 1998), of which only the first will be briefly sketched (note that conditions 1-10 above in 4.4.1.2 and 4.4.1.3 will need to be taken into consideration when interpreting the following):

Double Cancellation

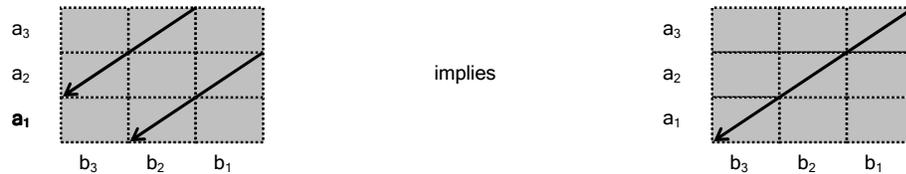
Double cancellation is a consequence of additivity (Borsboom, 2005). As is now known, additivity necessitates that objects (a , b) can be represented as $f(a) + g(b)$. Assume the following matrix consisting of Factors A and B.

		Factor A		
		1	2	3
Factor B	1	(a_1, b_1)	(a_2, b_1)	(a_3, b_1)
	2	(a_1, b_2)	(a_2, b_2)	(a_3, b_2)
	3	(a_1, b_3)	(a_2, b_3)	(a_3, b_3)

If certain pairs of values of A are ordered by \geq then other particular pairs of values will also be ordered. As with transitivity, \geq is a simple order (Michell, 1990). Within conjoint measurement, the transitivity of \geq on A is a special case of cancellation (Borsboom, 2005):

- (a_2, b_1) has to be \geq for instance (a_1, b_2) IFF
- $f(a_2) + g(b_1) \geq f(a_1) + g(b_2)$
- If this is so, then (a_3, b_2) \geq (a_2, b_3) and $f(a_3) + g(b_2) \geq f(a_2) + g(b_3)$
- Thus, $f(a_2) + g(b_1) + f(a_3) + g(b_2) \geq f(a_1) + g(b_2) + f(a_2) + g(b_3)$
- which is simplified to $f(a_3) + g(b_1) \geq f(a_1) + g(b_3)$.

Additivity implies that (a_2, b_1) \geq (a_1, b_2) and (a_3, b_2) \geq (a_2, b_3) then (a_3, b_1) \geq (a_1, b_3) which can be proven and is known as the condition of double cancellation. This is represented in the following two matrices. Double cancellation (the second conjoint matrix) is a consequence of the first conjoint matrix. Double cancellation thus provides indirect evidence for quantitative structure and conjoint measurement allows for the determination of attributes' additive structure as opposed to their merely being ordinal (Michell, 2003).



- The Rasch model is a particular example of such conjoint measurement with an underlying stochastic structure (Brogden, 1977; Embretson, 1996; Embretson & McCollam, 2004; Michell, 2002, 2003; Perline, Wright, 1999; Wright, & Wainer, 1979; Rasch, 1980). In other words it is a probabilistic model which models the probability of a response and does not model the actual response (Brogden, 1977) (see section 4.4.1.3 below for IRT model information) or in Rasch's (1980) words "a means of describing a series of events which cannot be predicted to occur as definite points of time, but to which probabilities of occurrence may be ascribed" (p.36). Part of converting qualitative ordinal level data into interval data would be to apply such a stochastic measurement model (Wright, 1997a). Rasch utilised Poisson's distribution of exponential additivity as it enabled the equation of two tests (rather than items) to be independent of a distribution (Jansen, 1994; Rasch, 1980; Wright, 1999). The importance of probability as sub-discipline within mathematics, statistics and measurement can be clearly traced from the early writings of Bernoulli, Poisson and Bayes.⁸⁹ They have played forth in behavioural statistics as well as measurement and IRT as can be seen in the Rasch model and even though Rasch started from a probability angle, the resultant IRT curve was a logistic model (Baker, 2001). "Rasch models construct conjoint additivity by applying inverse probability to empirical data and then testing these data for their goodness of fit to this construction" (Wright, 1999, p.80).
- This discussion on conjoint measurement as illustrated by the Rasch model is pre-empting the brief introduction to IRT below but it is necessary at this juncture to qualify why the Rasch model is one of conjoint measurement. The following is taken exclusively from Borsboom (2005). Additive versions of latent

⁸⁸ See Borsboom (2005, p.98); Coombs, Dawes and Tversky (1970, pp.26-29); Michell (1990, p.72) who present 36 substitution instances of double cancellations; Luce and Tukey (1964, p.3) and Perline, Wright and Wainer (1979) for more extensive explanations.

⁸⁹ It is perhaps necessary at this juncture to point out some of the advantages as well as disadvantages of utilising Bayesian statistics within a psychological assessment situation. For instance (Kingsbury & Houser, 1999) utilised Bayesian priors for a student scoring procedure which proved useful in an IRT adaptive test set-up. Two students evidencing the same true score and same test performance will receive different Bayesian level estimates if their priors are similarly different. Although true scores are unobservable, test scores are not and if they evidence equal results it will be hard to understand why different scores are eventually allocated to these two students. Cognisance must be taken of such situations.

trait models, such as the Rasch model hypothesises expected item responses to be logistic functions of the latent variables; this function, which encapsulates the subject i 's response to an item, j , is as follows:

$$P(U_{ij}) = \frac{e^{\theta_i + \beta_j}}{1 + e^{\theta_i + \beta_j}}$$

- $P(U_{ij})$ is the probability of a correct response; β_j is the location of item j on the θ scale where the probability of an endorsement would be 0.5
- Item response probabilities are then monotonically transformed and evidence simple additive representation and the above model is then rewritten as follows:

$$\ln \left[\frac{P(U_{ij})}{1 - P(U_{ij})} \right] = \theta_i + \beta_j$$

- \ln is the natural logarithm.
- The axioms of conjoint measurement are applicable to the model in its stochastic form if both the following hold
 - the probability, $P(U_{ij})$ is transitive;
 - i.e. if $P(U_{ij}) \geq P(U_{kl})$, and
 - $P(U_{kl}) \geq P(U_{mn})$, then
 - $P(U_{ij}) \geq P(U_{mn})$
 - and if it is connected
 - Either $P(U_{ij}) \geq P(U_{kl})$ or
 - $P(U_{kl}) \geq P(U_{ij})$ or both
 - because probabilities are numerical and numbers are ordered which are a result of Kolmogorov's probability axioms (see Kolmogorov's role earlier in this chapter)
 - independence is still upheld as both item difficulty and person ability are independent variables
 - the Rasch model will uphold double cancellation (assuming it to be true) because as was shown above,
 - If $\theta_2 + \beta_1 \geq \theta_1 + \beta_2$ and
 - $\theta_3 + \beta_2 \geq \theta_2 + \beta_3$ then
 - $\theta_2 + \beta_1 + \theta_3 + \beta_2 \geq \theta_1 + \beta_2 + \theta_2 + \beta_3$ resulting in
 - $\theta_3 + \beta_1 \geq \theta_1 + \beta_3$ which upholds the double cancellation

The solvability condition implies that either the values of a and b are equidistant or that they are rational numbers (Krantz, 1964); i.e. personal ability and item difficulty are continuous (Borsboom, 2005). The Archimedean condition limits the degree to which differences can be infinitely larger than any other difference within the conjoint matrix and this is independent of the column one wishes to inspect (Michell, 1990; Narens & Luce, 1986) resulting in person ability and item difficulty being unbounded (Borsboom, 2005). In effect, conjoint measurement takes its lead from analysis of variance models in which dependent variables vary alongside the joint effect of at least two variables. A crossed factorial design tests for the manner in which the dependent variable can be represented, either as a sum of the rows or columns and thus illustrates its similarities with this design (Perline, Wright & Wainer, 1979). Since the Rasch model is additive it is considered a form of fundamental measurement procedure utilised within psychometrics but as to whether it is truly scientific is questionable, according to Borsboom (2005) and Kline (1998).⁹⁰ Item response theory will be discussed below but it was necessary to introduce certain concepts in the conjoint discussion in order to elucidate the notion via such as model. Also, it was necessary to highlight the role played by conjoint measurement as fundamental measurement within such a model. This is because the Rasch model is used in newer models within change assessment research designs, such as Embretson's multidimensional Rasch model for learning and change. In order to maintain consistency in argument it needs to be illustrated that the Rasch model, as one instance of IRT, test theory is used in newer models and in particular can be utilised within a dynamic assessment framework thus upholding a scientific approach to the subject matter yet edging closer to fundamental measurement within a qualitative domain.

⁹⁰ It is interesting that Borsboom (2005) states that Kline (1998) has adopted the view that psychology can only be scientific if it is measurable. Yet Kline (1998) specifically states that what there is to fundamental measurement as viewed through the Rasch model (as substantiating fundamental measurement) is questionable in terms of its being scientific!

4.4.2 Test theory

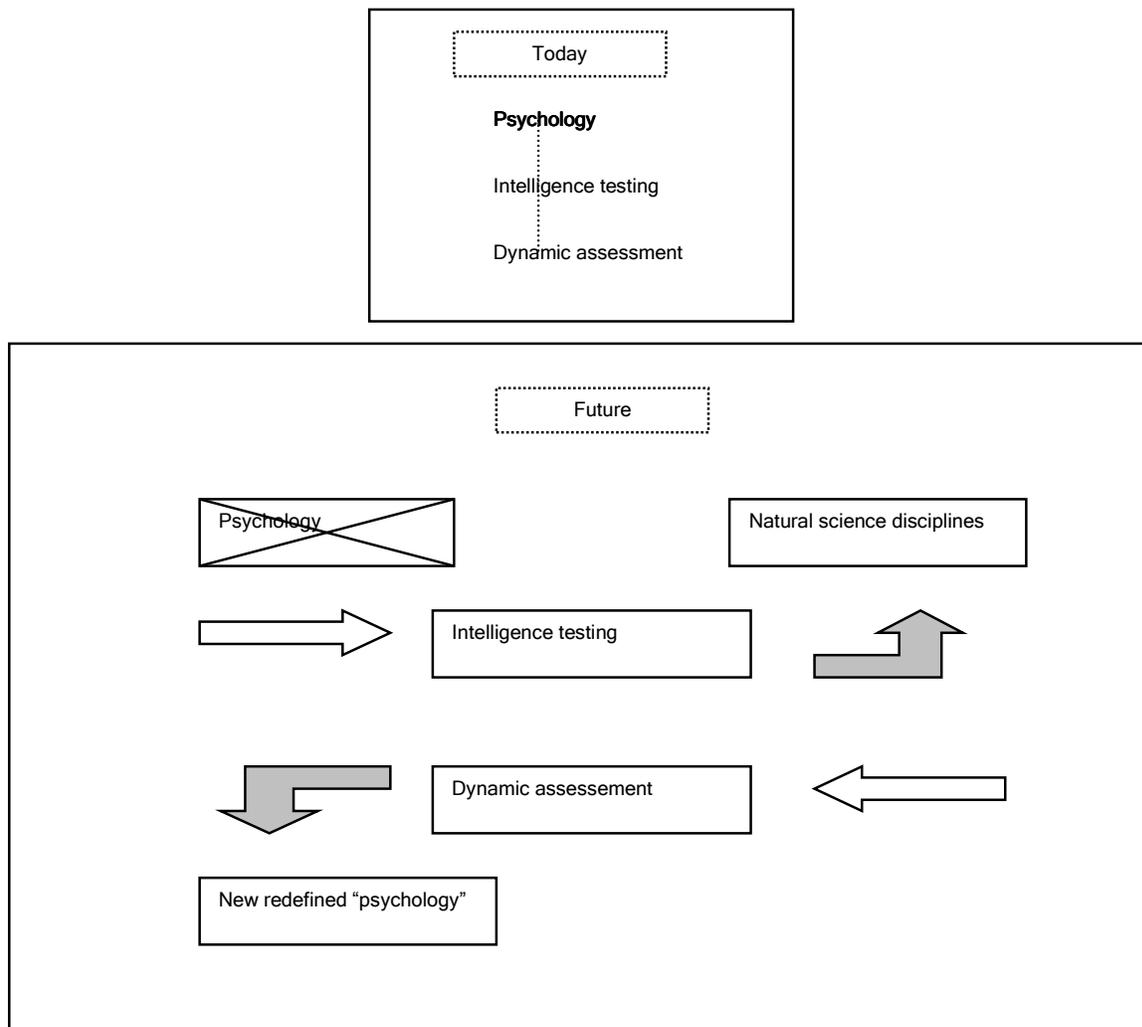
Blinkhorn (1997) humbly yet purposefully states that test theory is an undefined area of study which does not define what in fact takes place during test-taking but is a framework from which to view “categories of methods, models and techniques associated with the construction and evaluation of tests it is about attempting to fit tractable statistical models to test score data” (p.176). Test theory, he adds, is a collection of computation techniques with statistical manipulations attempting to give the enterprise credence. Reflecting concerns that Borsboom and Mellenbergh (2002) raise about the lack of concern regarding construct identification and the lack of clarity surrounding what test theory propounds as viable constructs, Blinkhorn (1997) admits of psychologists’ unwavering regard for the supposed accuracy and meaningfulness of what test theory has to offer by way of tests. McDonald (1999) maintains that test theory is nothing more than theory underlying psychological tests or measurement which is in turn derived from psychometric theory. But most importantly is his regard for the mathematical and statistical undergirding of psychometric theory “it is necessary to recognise from the beginning that test theory is essentially applied mathematics, overlapping with statistics” (p.3) this underscoring the need to re-look mathematics and statistics within the social sciences. Embretson (1987) states that psychometric test theory is being increasingly considered as a science, hence the need to assess what exactly is meant by social science (see chapter 3). Does trait underlie behaviour? Should behaviour (including intelligence and potential) even be considered as traits or should such constructs be considered as states as Feuerstein advocates? Lord and Novick contend that it does (1968). Psychological processes, whatever they happen to be, physiological, behavioural, bioecological and so forth is not consistent with a theory of underlying traits and this is simply not reflected in test theory (Blinkhorn, 1997). State and trait models cohere with test theory models but process and mechanism are left unaccounted for. Trait theories conveniently dismiss situation (which dynamic assessment as manner of assessment does not). The trade-off between relevance in a modern scientific age is to allow for as parsimonious a model as possible and such models are often found in test theory (Meier, 1994).

Test theory’s roots are located in individual difference research but has long since become reminiscent of statistical sophistication which has everything to do with large samples and population parameters - there is an irony in this state of affairs. The argument is such that no amount of statistical analysis and reconceptualisation can change what is necessitated, which is a measure of a psychological behaviour. Chapter 3 looked at the realm of psychological behaviour within science. No amount of rescue attempts at salvaging what is purported to exist within an ethereal realm can be consistent with a scientific approach and statistical test theory can perhaps be regarded as just one such attempt. If psychological theory cannot satisfy for itself a coherent definition of what it seeks to understand it seems pointless to bring in the services of other sciences in an attempt to rectify the situation. Psychology needs a theory consistent unto itself and cannot rely on methods and techniques of which it largely does not consist. Once again, psychological theory and statistical theory are investigating two very different areas of concern and through the interface of test theory it is assumed that contact is being made. Mainstream assessment purports to locate differences between individuals on test scores which is obviously what one is going to find in the manner in which it is secured. Psychometric theory’s historical context is couched in intelligence assessment hence the emphasis on intellectual factors in the growth and development of various test theories, although self-report questionnaires were being drafted in the 1920’s (McDonald, 1999). Dynamic assessment is predicated on another philosophy altogether and possibly should not (and in some instances does not) equate itself with individual difference research. Descriptive research is very powerful and often underestimated and one cannot deny glaring differences between various groups on various tests of intelligence for instance. The tests are so designed.

Dynamic assessment is not so designed and hence does not align itself with mainstream assessment. If mainstream assessment is so bent on statistically attuning itself towards finer discrimination between and within tests resulting in less rather than further progress, dynamic assessment should steer clear of yet even greater reform in the manner of assessment and test theory. Its place was never there to begin with and should subsequently not be there now. There exist alternative beginnings, different philosophies disparate tools and different methodology. Why the need to emulate mainstream trends which have in any event not proven terribly helpful or progressive? It is necessary to once again punctuate the story with yet another reiteration of the same sentiment which is being echoed throughout this thesis: psychology as a discipline needs to realign itself within a sphere which it can call its own and remain progressive. Currently, it is spread into domains with which it is clearly at odds and is struggling to maintain its existence. Dynamic assessment should be placed within an entirely different realm in comparison to current mainstream testing which itself should be placed within physiological and natural science methodology. There is nothing right or wrong about techniques utilised within such separate realms but the various movements within the discipline cannot co-exist in their current form, as the field will continue to be riddled with misgivings such as with which this thesis has been concerned. Figure 63 illustrates the reshuffling that needs to take place if dynamic assessment is to thrive in a progressive environment where it is unencumbered by the reigning tradition of psychology. Psychology should seek to dismantle its supposed unity as it is currently defined. The discipline already crosses so many boundaries that its existence is defined by nebulosity and uncontrolled spread. It should redefine itself according to its own progressive dictates and subsume dynamic assessment as method of assessment. Current intelligence assessment should shift over to natural science methodology where it can proceed unhindered by ill-defined behavioural constructs which are not and never have been defined in psychology.

Nevertheless, test theory's main preoccupation throughout its nascent years was the concern with reliability; reliability of the test as ascertained from parallel forms of the same test purporting to measure the same construct, or the same test delivered at different intervals over time (Spearman) which lead to work within classical test theory. Guttman's⁹¹ work situated reliability in terms of items and further elaboration of this work led to the development of generalizability theory as an extension of classical test theory. The Guttman structure can be seen within IRT models which were to come later (McDonald, 1999) and was able to deal more effectively with binary response options with which classical test theory could not. However, Guttman's scales were only ordinal and unidimensional (akin to the common factor in factor analytical approaches; Bond, 2000) (Kline, 1998). Modern test theory then, is really an amalgamation of continuous development within psychometric test theory and each theory exists because of the need to further refine the theory. There is thus a common thread running through these, at times, parallel and sequential developments. Due to time and space limitations within this thesis the discussion detailing classical test theory and item response theory will assume a background knowledge pertaining to these theories. A brief introduction to CTT and IRT as it pertains to dynamic assessment is given in Murphy (2002) as well as Murphy and Maree (2006). This discussion will be detailed from the outset.

Figure 63 The discipline today and in the future: a possible scenario for maintaining both mainstream and dynamic assessment methodologies



⁹¹ A sociologist. When one thinks about who contributed what to the discipline of psychology and why it was they did so, a clearer picture starts to emerge as to why they conceived of the methods they did.

4.4.2.1 Representational measurement: the legacy of Stevens' arbitrary powerful scales of measurement

Representational measurement is an apt description of what it entails; a representation of reality to numericism. Stevens (1946) eschewed the notion of additivity by stating simply that an attribute, represented numerically on a scale, would suffice as measurement (Kline, 1998).⁹² Borsboom (2005) maintains that the representationalist view on measurement is in fact constructivist and not operationalist even though Stevens' scales are considered as operational. Operationalist and representationalist views are often considered synonymous (Poortinga, 1991). It is constructivist due to its scales being constructed representations of data within a highly restricted representational framework of what constitutes a measurement scale. By assigning numbers to observables via rules, he avoided the necessary aspect of concatenation. Four conditions need to be upheld if such representation can conceivably take place and is taken largely from the originators of representational measurement foundations, namely Scott and Suppes (1958) as well as Suppes and Zinnes (1963) (in Narens & Luce, 1986). Measurement obtains when the following is made manifest:

- An ordered relational and operational structure provides the bedrock of empirical reality such that
 - $\chi = \langle X, \geq S_1, \dots, S_n \rangle$ where $\geq S_1, \dots, S_n$ are primitives of the structure. These primitives are the empirical relations on X .
- Axioms restrict the structure which reflects this truth about the empirical reality (recall axioms' functions as possibly undervivable within its own system; that is, they are a given which needs to be accepted for theory or structure in this case, to be at all meaningful). Hence, Narens and Luce (1986) state that these axioms are putative empirical laws. This of course is the crux of the issue in which it has been stated over and over that the degree to which psychological attributes can be axiomatised within statistical and mathematical as well as measurement theorems remains questionable
- A numerically based relational structure is comparable to the above mentioned empirical structure. Namely,
 - $\mathfrak{R} = \langle R, \geq R_1, \dots, R_2 \rangle$. R is the subset of real numbers and R_i represents the relations and operations comparable to the empirical relations and operations given above
- Lastly, proof of mapping is needed which illustrates that the ordering between empirical and relational has been preserved from χ to \mathfrak{R} . Structural preservation from one system to another is conducted homomorphically in such a manner that the structure is now related as a scale; enter Stevens (1946)

Writing in 1946, Stevens nowhere mentions in his trend-setting article, the controversy surrounding the assignment of numerals to characteristics of behaviour that are not necessarily amenable to such treatment; in fact it is almost taken as a given that psychological constructs are amenable to such tactics. Taking his lead from N.R. Campbell's classical treatise on measurement (Kyburg, 1992; Ross, 1964) he states that measurement is largely a semantic exercise. This is easily surmounted when devising suitable scales for such measurement and his conception of measurement is essentially a representationalist one (Michell, 2002) which is really just a way of stating a rule for the assignment of numbers to concepts (Ellis, 1966; Ryan, Lopez & Sumerall, 2001). Stevens was either unaware of the potency of the counter argument or had chosen to sweep it under the rug of uncomfortable questions. Stevens' operational scales of measurement have been with psychometrists ever since (Michell, 1997, 1999; Ross, 1964) upholding a firm positivist assumption of operationalisation equal to that of observed relations. This in no way diminishes his contribution to the clear understanding of the roles played by each scale but eschewing the issue of paramount importance as to whether one can truly measure something psychological in the first place is not good science. Having been influenced by Fechner's psychophysics and Spearman's quantitative science, the scene was thus clearly set for Stevens' scales of measurement to make its presence felt within a positivist mode and framework (Michell, 2002). He synthesised various emphasised aspects within works by Russell (1903), Johnson (1936) and Birkhoff on whom he leaned and utilised the former's numerical representation of order (Narens & Luce, 1986); Johnson's dual consideration of classification and ordering and the latter's theory of numerical transformations (Michell, 2002).

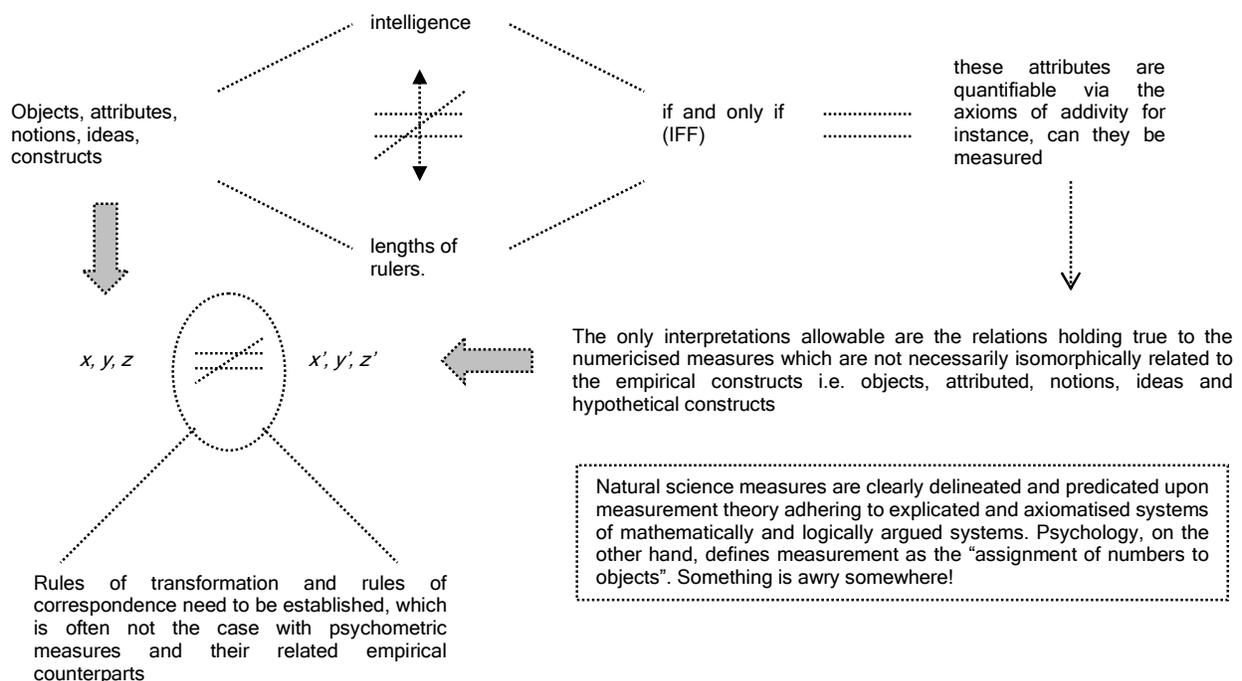
Appearing the 1940 Ferguson Committee of the British Association for the Advancement of Science which for ten years deliberated the matter of measurement and setting about ascertaining the veracity of quantitative sensory measurement, Stevens managed to provide for his scales of measurement an operational veneer. He succeeded in circumventing the issue of quantifiable constructs in psychological measurement by assigning numerals to these constructs (Stevens, 1946) and described statistical procedures for the scales for which they were "permissible" (Velleman & Wilkinson, 1994). The committee's results were evenly split between those advocating that measurement in psychology was impossible versus those in favour of it

⁹² This is reminiscent of the underlying rationale before Stevens; "if relations of qualitative increase or decrease are understood by analogy with quantitative change, and the mathematics of quantity is viewed as a mere formalism as suitable for applications in all disciplines, then psychological attributes that seem to be ordinal can be hypothesised to be quantitative (Michell, 2003, p.10).

(Borsboom, 2005). Steven’s notions of measurement were, however, further refined by future generations of researchers who stipulated that it was the properties of psychological constructs rather than the constructs themselves that were being measured (Crocker & Algina, 1986). The committee did not really come to a decisive conclusion (Luce, 2000). The assignment of numerals or numbers to empirical events, was in Stevens’ mind an isomorphic occurrence which is similar to the depictions illustrated in figures 34 and 54 as stated above (Ross, 1964). The scale transformations between ordinal and ratio for instance is an arbitrary one, as Ross (1964) highlights the very important fact that such transformations are applicable to the scale properties themselves but have no bearing on what is being scaled in the first place, a view that Michell (2002) shares. Hence, there was no explicitly defined philosophy behind the purported isomorphism between the formal aspects of the scale and the empirical domain to which these formalities applied. This lack of acknowledgment is a two-tiered fault (Michell, 1999, 2000) and is tantamount to the burying of heads in sand. The first error was in presuming answers to questions not empirically investigated (are psychological properties measurable in the sense of assigning numbers to them?) and the second higher-order error emanating from this error was the rank acceptance of a faulty definition by either ignoring it or glossing over its illogicalities (Michell, 1999, 2000). Figure 64 depicts the typical mode of operation of measurement. Note, however, that this illustration rules out any necessary equality between supposed quantified psychometric structures. The error continuously practised by psychometrists is in taking from a physicalist notion of measurement ideas which are absorbed in that system and transplanting these notions in another system altogether. As Ross (1964, pp.96-97) states:

the kind of psychological measurement theory expounded here appears to differ from the corresponding physical measurement theory as the descriptive semantical rules of this kind of physical measurement theory all referred to empirical terms. This difference stems from the fact that our exposition of the physical measurement theories is based on the assumptions of *absolute reproducibility* of physical experiments, an assumption which is not fulfilled in practical work (own emphasis).

Figure 64 Rationale behind measurement: ill-conceived notions of quantifiable constructs



4.4.2.2 Classical test theory – a goal for inference⁹³

The development of true score theory in classical test theory may be tentatively traced to the nineteenth century theory of errors (Stigler, 1999) during which time early statistical methods of treating astronomical data in the latter half of the eighteenth century were to meet later with the mathematical theory of probability early in the nineteenth century. The pivotal issues surrounding accurate data in terms of the relativism of observations in astronomy drove scientists to speculate on the relativism of 'correct' observations. Thus emerged the idea of focal points of observations being equal to 'truth' in addition to error (random and/or

⁹³ Regarding the early role of statistics in both natural and social sciences, Stigler states "what Newton had given astronomers and experimental design [had] given the psychologists: a goal for inference" (1999, p199).

systematic; Crocker and Algina, 1986) (Stigler, 1999, p.190). This is akin to the notions of later psychometric concerns with observed scores equalling true scores in addition to error: X (observed score) = T (true score) + E (error score) (Borsboom & Mellenbergh, 2004; Crocker and Algina, 1986; Daniel, 1999; Embretson, 1976, 1999; Kline, 2005; Marcoulides, 1999; Meier, 1994; Rust & Golombok, 1992; Scheuneman, 1991; Smit, 1996; Suen, 1990). The reliability (or 'attenuation' as Spearman first made use of this concept; Du Bois, 1970) resulted in a mathematical model originally developed by him in 1904 for the application in the area of intelligence measurement within the social sciences (Crocker & Algina, 1986; Du Bois, 1970; Murphy & Davidshofer, 1998).⁹⁴ "Despite the unity of statistics [the role played by statistics in the natural sciences vs. the role played in the social sciences], there are fundamental differences, and these have played a role in the historical development of all these fields" (Stigler, 1999, p.199). Classical test theory (CTT) can claim to be the earliest theory of measurement and is synonymous with classical reliability theory, true score theory, true score model and random sampling theory as its main aim is to estimate the strength of the relationship between observed and true scores (Suen, 1990). Common linkages between CTT and psychometrics are often made along with the assumption that the former is in fact the latter in totality (Gipps, 1994) and forms of assessment such as dynamic assessment have been lodged under the rubric of educational measurement. Due to the variability of the error score, possible changes within test scores can be ascribed to change or modifiability but because it is random it is without diagnostic value (Wiedl, 2002). The true score model, according to Borsboom (2005), is operationalist because the true score is defined in terms of the test score; in other words, its operationalises the notion of true score. In fact CTT and generalizability theory are considered as two approaches falling under the rubric of random sampling theory (Marcoulides, 1999). The underlying philosophy assumes statistical models which cater for an infinite number of testings thus allowing for a truer picture of the ability being assessed, or as Kerlinger (1981) puts it only an omniscient being would really know the true score. The mean score in CTT is the mean score of an infinitely long test (McDonald, 1999) (CTT's test-dependence; McCollam, 1998). This mean varies depending on the properties of the population being sampled thus according prime status to population parameters (CTT's group-dependence; Kline, 1998; McCollam, 1998), an aspect abolished in probabilistic item response models (Kline, 1998; Rasch, 1980). The deployment of various statistical techniques to achieve just this is the main rationale behind CTT.

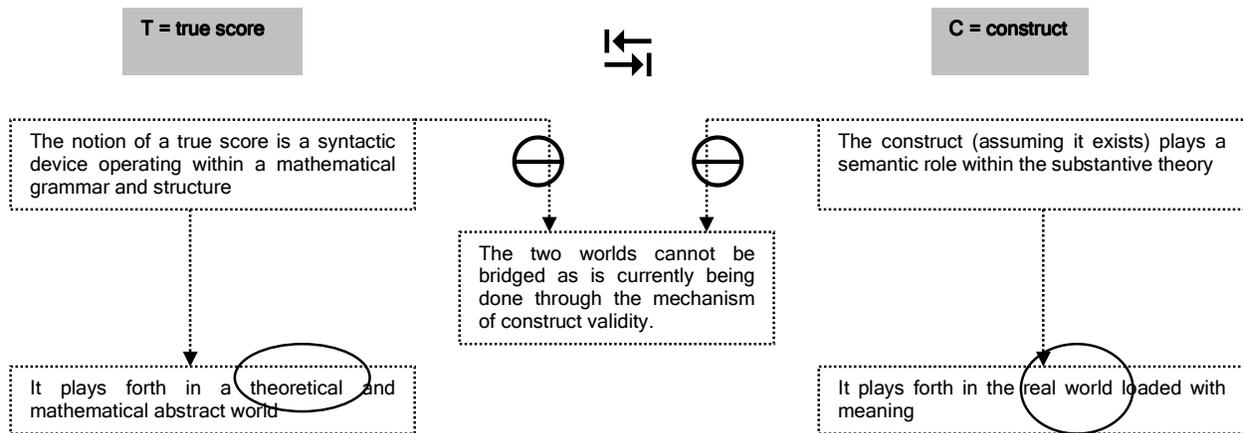
Reliability estimation, error variance and standard error of measurement lead to the estimation of a true score even though the only manifest score is the observed score. CTT only allows for ordinal scaling interpretation for the raw scores whereas item response models, via logistic functions, allow for interval level scaling (ability is estimated as a probability with comparable ability estimates for any calibrated item) (Embretson, 1983; McCollam, 1998). The total test is important to CTT whereas modern test theory emphasises the item, although there are latent trait models which are developed at the level of the test (Maxwell & Delaney, 1985). CTT principles are in fact derivable from special cases of IRT but the reverse does not hold (Embretson & Hershberger, 1999). Although it may seem that the death knell has struck for CTT, the future advancement and development of generalizability theory which encompasses CTT principles as well as IRT principles indicate that doing away with CTT would be rather drastic, notwithstanding the considerable number of test batteries still in vogue which are built on CTT principles. The situation of CTT and IRT is reminiscent of the parallel situation of dynamic assessment and static assessment; both models should be retained for what they can advance as they differ in their information yielding functions. Perhaps the most important unobtainable aspect within CTT is the theory's inability to consistently work out a reliable gain score from a pretest to posttest scenario. Utilising algebraic conversions, Pearson's correlations, variance of true scores and adhering to the central tenet of observed score equalling true score and error score, the reliability of the difference score is high only if the correlation between the two scores is low (Suen, 1990). Perhaps the most controversial aspect of CTT is the nature of the as yet improvable or at least the indefinable relationship between the true score and how it supposedly represents a construct. Once again, dynamic assessment needs not confine itself to measurement theories which have unresolved issues of their own. It does not necessarily have to align itself with any current theory of measurement in fact. CTT is underscored by the notion of a true score being the result of observations dependent on error distributions. Change can yet be accommodated within CTT as Wiedl (2002, 2003) illustrates in figure 65 below but there is a need to determine and identify the correct model within different dynamic assessment set-ups; change within IRT will be assessed below. Agreement of true scores with constructs is referred to as the fallacy of Platonic true scores which, if recalled, made its first appearance in the discussion on the Platonic realm of mathematical entities. The Platonic true score is not consistent with CTT (Borsboom & Mellenbergh, 2002). This line of argument takes us back to the whole notion of the representationalist view of measurement. CTT as incorrectly assumed by some psychometrists, is thought to equate the realm of true scores (the syntactic) with the construct (the semantic). Figure 66 illustrates this erroneous equating of true score to construct.

⁹⁴ What sent the social sciences down a path discernibly tangential to the natural sciences in terms of statistical manipulations of data was the idea (predating even Fischer's work on agricultural experimentation) of randomised experiments (Stigler, 1999); as exemplified in the work of Charles S. Peirce who made use of a version of an experiment already developed by Fechner (an experiment on sensation and stimulation; Sahakian, 1981) and in so doing provided an even more robust result due to his use of a blind randomised experiment (Stigler, 1999).

Figure 65 Change of performance in terms of CTT (Wiedl, 2002, 2003)

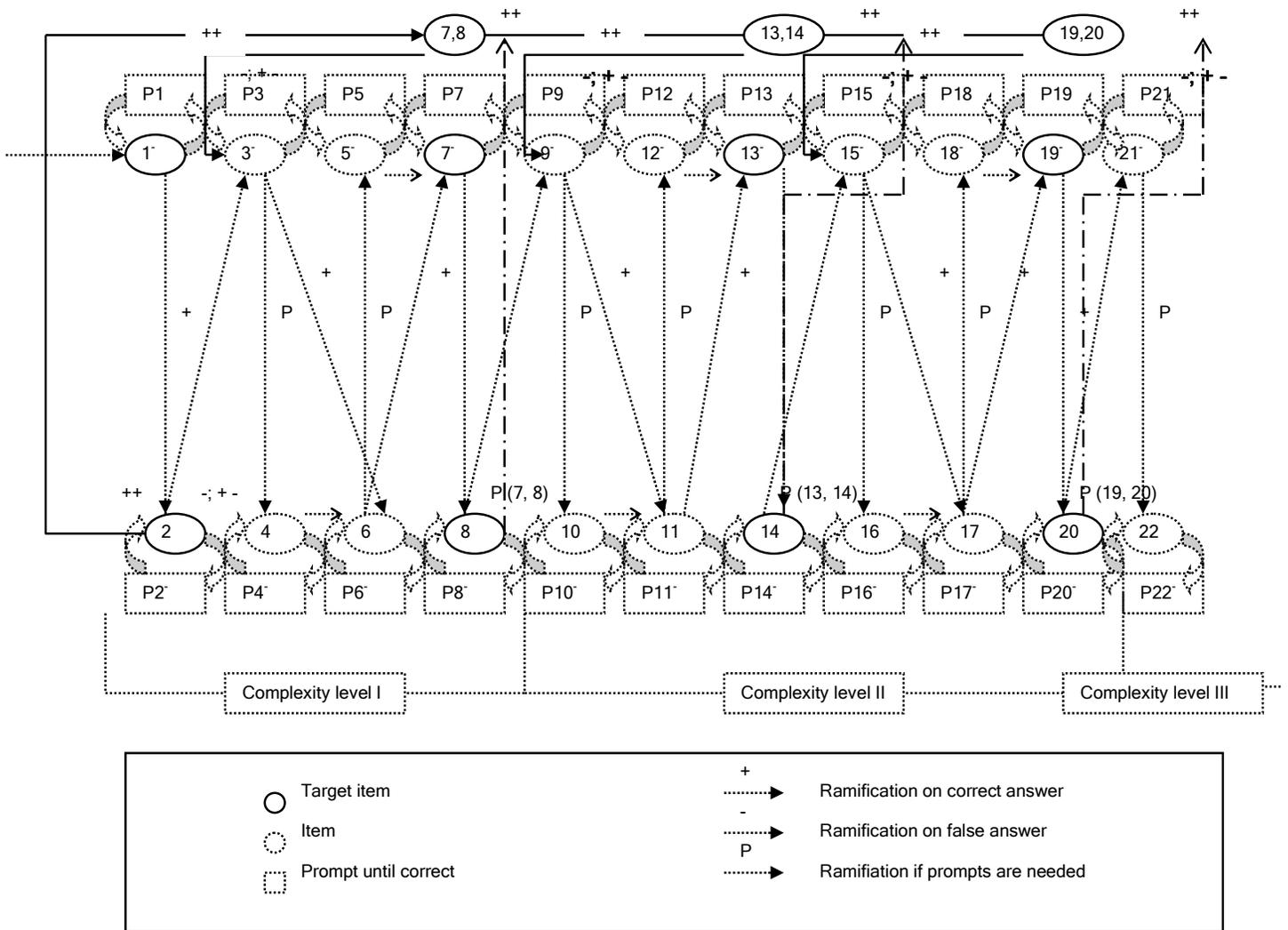
<ol style="list-style-type: none"> 1. $X = T + E$ 2. $X = T_S + T_V + E$ 3. $X = T_S + E_{RAN} + E_{SYST}$ 4. $X = T_S + T_V + E_{RAN} + E_{SYST}$ <p>X: performance on test T: true value E: error term s: stable v: variable ran: random syst: systematic</p>	}	<p>Possible ways of how change can be accommodated while still remaining within CTT tenets. The problem is to determine which of the models is the most accurate given the circumstances. Wiedl (2003) is very cautious however when he likewise cautions the reader to the as yet ill-defined construct being assessed; “no assumptions are made with regard to what is measured and the implications of this. The definitions thus cover different theoretical concepts” and refer, in turn to, the “zone of the proximal development’, ‘learning ability’, ‘cognitive modifiability’, ‘baseline reserve capacity’, ‘developmental reserve capacity’, ‘responsivity to intervention’ and ‘rehabilitation potential’” (p.95).</p>
---	---	--

Figure 66 Erroneous equating of true score and construct



Change can be captured within CTT and depending on what the goal of the assessment/intervention is, success need not be the outcome of IRT change-based models (or process models; Guthke, 1982) alone. The context, as with most things psychological, is prevalent and largely determines the what/when and how of assessments, so if the nature of the situation calls for CTT-based models then there are at least such dynamic assessment models which attune themselves to this framework. Wiedl (2003) mentions two such dynamic assessment paradigms located within CTT; the test-training-test paradigm and the paradigm of continuous integration of interventions into the test, the latter will now be briefly discussed at it pertains to dealing with the construct of change within CTT (compare to IRT). A partial glimpse of Guthke’s Diagnostic Program for the Assessment of Reasoning abilities (ADAFI) is illustrated below in figure 67. The construct of change is discussed further in section 4.4.2.2.1 but is brought in here due to its applicability within CTT.

Figure 67 The integration of dynamic assessment into the testing process: Guthke's ADAFI (Wiedl, 2003, p.96)



The ADAFI is still housed under CTT and thus conforms to its leading tenets so espoused and endorsed by the popular rendering of what it means for psychological measurement to be objective, valid, verifiable and in general robust. Change is accounted for via a process of correct and incorrect answers to items and is mediated during promptings. This model is cognisant of the task characteristics, the testee's ability level and how these parameters vary throughout the testing process (Wiedl, 2003). However, change as understood within change-based IRT models as another construct is not the same here - "the focus is not on change, but on the level of performance that can be achieved and the amount and quality of hints or time that the testee needs" (Wiedl, p.96). Nevertheless this manner of assessment is still dynamic. The point now would be to compare the IRT change-based models for dynamic assessment to ones such as this, offered by Guthke and seek the more attractive alternative which is dependent on the critical notion of context.

follows on next page ...

Continued from previous page ...

The algorithm is easy to follow and follows a logic which encompasses both item and person information. For instance, in keeping with the above model the following algorithms can be specified for the following items or questions (Q):

- If Q1 is correct move to Q2; if Q2 is incorrect prompt and try again; if Q2 is correct on first try move directly to Q7 and Q8; if Q7 and Q8 incorrect move back to Q3 and if correct move to Q6 (Q5 is presumably too easy at this stage for the person following this pattern); if Q7 and Q8 correct on first try move directly to Q13 and Q14 and likewise carry on in this fashion
- If Q2 is incorrect prompt and try again and move onto Q3; If Q3 is incorrect prompt and try again; if Q3 is still incorrect prompt before moving on to Q4; if Q3 is correct in first attempt move to Q6 (note that this person follows the pattern for the person above but only after having reached Q6)
- Q6 is the first location where an item is skipped presumably to determine the level of functioning; this is most likely done purposely. Respondents answering from the Q3 and Q4 are both directed to Q6 thus skipping Q5, however, should Q6 pose problems, prompts will direct respondents back to Q5 or else they proceed to Q7 directly
- The algorithm proceeds in this fashion for the rest of the items
- Item difficulty is known as well as person performance and in this manner it is very similar to the notions underpinning IRT change-based models discussed below

The construct of intelligence as presumably measured by various intelligence tests cannot be equated as they are expressed as functions of operations which differ between such tests. "Intelligence may be a psychological construct, but its explication requires that we don the hats of other kinds of investigators and try to integrate their varying perspectives" (Gardner, Kornhaber & Wake, 1996, p.134). Also "I do not believe that psychology can be reduced to the brain; but I do believe it can be put into correspondence with it and this should happen, because study of the brain can help to clarify the postulates that our psychological theories must have" (Pascual-Leone, 1997, p.80). A question which is implicitly implied but one which is not often stated directly, is the usefulness of intelligence, which, would seem at first glance to be quite obvious, intelligence aids our survival (in whichever milieu this survival happens to express itself), which is echoed in what Johnston, Partridge and Lopez state; "the value of intelligence lies in its ability to allow *rapid adaptation* to occur within the life-span of the individual organism" (own emphasis) (1985, p.487). Thus in extending this argument further, is 'rapid adaptation' nothing more than learning to adapt, thus learning to train oneself; learning to 'learn the situation' so to speak in as quick a time-span as possible? Is intelligence, by extrapolation then, nothing more than learning to change? If this is true, then assessing for change should be a main concern within intelligence testing.

This argument, of course, hinges on the above definition of what intelligence is. Intelligence assessment is almost entirely dependent on how one chooses to define it. If the true score can not be equated with the construct score as argued above, then it stands to reason that each construct score represents something else and does so via different operationalisations. In essence, an intelligence score on one test cannot be equated with an intelligence score on another test which leaves one with the uncomfortable question of what the tests are measuring. CTT expresses a construct only in so far as it pertains to specifics which is tantamount to saying that depending on which method you use to derive the gravitational constant, the result will not remain constant but will need to be re-examined and considered from a specific operationalist point of view. Such variation cannot be good for any discipline calling itself a science. "CTT-derived scores predict performance only at the point represented by the obtained test score. Rasch-based interpretation predicts performance on any task given its scaled distance from the individual's measured ability" (Woodcock, 1999, p.126). The need to trace an underlying variable which can be equated across tests is found in latent trait models. Parametric modern test theory models are instances of generalised linear item response theory of which the Rasch model is one such example (Borsboom & Mellenbergh, 2002). CTT true score resembles the IRT latent trait but whereas CTT is an unrestricted model IRT is not in terms of placing restrictions on the observed data. CTT posits hypotheses of true scores based on an infinite number of attempts to observe it, whereas IRT does not. However, Blinkhorn (1997) is not convinced of the veracity of the Rasch model in terms of what it can offer over and above CTT and he maintains IRT's stance as redolent of psychophysics. He refers to its supposed benefits as a mirage due to the model's inability to fit all types of data. There is thus disagreement within the literature over the various CTT and IRT models. This is precisely the point made in chapter 3 in which higher order problems are being felt in the chain in the lower areas within the discipline. The errors need to be rectified at the top. CTT and IRT models are robust in many instances, are mathematically sound (most of the time) and statistically amenable to manipulation but no amount of bolstering of techniques can do much to solve a largely philosophical dilemma emanating from the top most level of concern - the discipline itself.

4.4.2.2.1 Malleable stability - the gain score issue

“Th[e] gain score information has a controversial history in the psychometric world because of the unreliability of the scores and the dependence of the posttest score on the pretest” (Lidz, 2002, p.123). The paradox of the gain score can be summed up as follows: if a test-intervention-retest research design based on changeability via dynamic assessment theory yields low psychometric reliability then this is an indication of poor test construction. Score changes thus threaten psychometric properties of tests under the classical test mode (Hamers, Hessels & Pennings, 1996). However, pretest and posttest scores within a mediatory context are supposed to evidence poor reliability for the sole reason that the ability level undergoes changes due to inherent potential manifestation (via practitioner action; Elliott, 2003) which would otherwise never become evident (Embretson, 1992, 2000). The use of IQ tests within educational settings is its greatest liability (Ramey & MacPhee, 1981), which is, frankly, bordering on the absurd! Stability and modifiability models therefore preclude each other and in this respect are in stark opposition (change brought about by an intervention during a test is considered a threat to psychometric validity; Hamers & Sijtsma, 1995; Sternberg & Grigorenko, 2002) but can also be viewed as complementary depending on the nature of the measurement context (Jensen, 1992). Nevertheless it can be unequivocally stated that “A primary goal of LP (learning potential) assessments is to defeat the very predictions made by traditional IQ scores” (Glutting & McDermott, 1990, p.398). Recall that psychometrics developed specifically to test for a unitary concept of underlying intellectual functioning (Davis & Anderson, 1999). Dynamic assessment’s posttest scores are often enhanced predictors of future static and dynamic assessment intellectual achievements which results in asking the question of how valid the original static (pre-test) scores are in the first place (Tzurriel, 2000a). Traditional learning theory only accounts for changes in simplistic posttest - pretest score scenarios and it often happens that initial task scores are more variable than posttest score results which evidence less covariation over time and practice periods (Ackerman, 2005). CTT’s preoccupation with stable individual differences is particularly evident within psychometrics (Jensen, 1992). Wilder’s Law has been recognised for some time, which attests to the fact that a gain made over a testing situation is a function of his initial ability (Guthke, 1993b). Vernon (1952, 1954 in Lidz, 1987a) had conducted research into the gain score issue as early as the 1950’s detailing research dated to 1920’s. Mainstream psychometric theory by and large assumes a stable trait underlying ability (Irvine & Dann, 1991; Jensen, 1992) which should not change and classical test theory’s *g* undergirds this philosophy. Dynamic assessment’s predicates are however entirely different and the possibility of the construct changing exists in a framework where such change is hoped for (Sijtsma, 1993b). Recall, however, “the fallacy of inferring fixity in ability from constancy in status scores” (Lohman, 1997b) which argues for malleable ability even though scores may indicate otherwise. Modern test theory, through model development which encompasses changes underlying ability as latent trait, is able to cope with this very situation (Embretson, 1987) and is expanded on later. Three main issues within the gain score debate can be summarised as follows (Embretson, 2000):

- i. The paradox of reliability which is highlighted in the absurd situation of decreasing reliability of supposedly stable traits from pre to posttest situation; the so-called “unreliability-invalidity dilemma” (Bereiter, 1967). CTT requires stable reliability from one situation to the next. Dynamic assessment does not fit the CTT model well and as such evidences decreased reliability from pre to posttest situation. What is needed is a psychometric model which accounts for change as reliable. In this regard Bereiter (1962, p.7) sums up by stating that “the functions of prediction and of education are opposed rather than complementary” where dynamic assessment can be regarded as educational and mainstream psychometrics the predictive area of concern. CTT’s source of error is dynamic assessment’s bedrock of method! Bereiter wrote this in 1962, and IRT response models were being propounded in various forms as early as 1952 - 1960 (Embretson, 1997b) so it is a curious fact that dynamic assessment has had to wait so long for awareness into change properties to surface. Bereiter (1967) sites the situation as being the cause of deficient statistical methodology, which within CTT it is, but not within IRT as expanded on below
- ii. The change evidences a negative false correlation with initial status (Bereiter, 1967; Embretson, 1991b, 2000), the so-called “over-correction-under-correction dilemma” (Bereiter, 1967). McDonald (1999) states that if the underlying trait between a pretest and posttest score is similar the error variances are additive but it is contested whether the same trait is being measured
- iii. Scale units do not have a consistent meaning; there is no common metric (Cronbach & Furby, 1970). Large changes from an initial low ability as opposed to small changes from an initial high ability are not comparable which makes meaningful conclusions irrelevant. There is a need to ensure that a change of “one unit” at a low initial ability is equivalent to a change of “one unit” at a high initial ability level. Bereiter (1967) refers to this as the “physicalism-subjectivism dilemma”. In other words should the scale units remain the same or should they be arbitrarily changed to accommodate changes (physicalism) or should scale units change along with changes in the construct (subjectivism)? It is noteworthy at this juncture to point out Bereiter’s early 1967 claim that psychologists have tended to avoid working in areas such as this due mainly to the perceived insurmountable problems evidenced from the lack of available statistical technique, which was already being developed at that time in any event. Nevertheless, as Michell (1999, 2000) was to echo three decades later in his trenchant critique of psychologists simply ignoring these issues, it seems that such ignorance is rife through time. Differences in error measurement exists between initial and posttestings which nullifies any attempt to equate the two scores (Embretson, 1987)

Recall that most scales utilised within psychological research designs only ever reach ordinal scaled measurement as designated by Stevens (1946). Interval level scales are necessitated to make manifest and meaningful the differences in change scores. Lastly, measurement error as calculated within CTT framework posits a population parameter error source from variance ratios and hence standard error of measurement applies to all scores whereas in IRT the standard error of measurement, although generalising across populations, differs across response patterns (Embretson, 1997b). Most of the South African research in dynamic assessment surveyed in 2002 treated gain scores in the manner befitting CTT and not IRT (Murphy, 2002; Murphy & Maree, 2006) and at times treated gain scores in the fallacious manner of simply subtracting the pretest score from the posttest score and measuring change as simple gain (Cronbach & Furby, 1970; Embretson, 1987) or realising “potential” as initial scores subtracted from gain scores (Swanson, 1995). One notable exception was the study by De Beer (2000) which employed IRT via computerised adaptive testing to construct and validate a dynamic assessment instrument. The three issues mentioned above are framed within CTT framework which is precisely where the problem lies. Lidz (2002) states that one way in which the gain score issue can be approached is to ascertain statistical information regarding both the pre- and posttest errors of measurement. “If the obtained score outpaces either of these, there can be more confidence in the significance of the gain, and the issue of the practice effects can be ruled out” (p.123). However, Lidz (2003) emphasises the time and mathematical expertise that will be needed to implement various models which can accommodate change within their models; most dynamic assessors are, after all, practitioners working in the field. Change scores which reflect intra-individual changes within a test situation were critically reflected on as early as the 1930’s in Europe (Wiedl, Guthke, Wingenfeld, 1995). Early works in Europe that were veering in the more process oriented direction in terms of assessment were pressing for conceptually different techniques of assessment, although these thoughts had, as yet, to manifest in actual test application (Wiedl, Guthke & Wingenfeld, 1995). Early on Cronbach and Furby (1970) championed the residualised gain score recommended by DuBois (1957 in Cronbach & Furby, 1970). Here, the posttest score is expressed as a deviation from the posttest-on-pretest regression and the remainder or residual information of the posttest score that is not linearly predictable from pretest scores is taken as the change score. This was done in an attempt to appease the then current testing establishment in terms of extracting true gain from pre to posttest measures. Residuals remained free from the linear predictors made on the basis of the pretest scores. Change-based measurement, should by its very nature, estrange itself from stable trait measurement. Yet another reason for dynamic assessment to disassociate itself from mainstream measurement (Meier, 1994).

4.4.2.2.2 Generalizability theory

A gradual development of theory resulted in CTT’s vague conceptualisation of random error into a more contextualised account of error specific to the context at hand and although contributed to by many psychometrists, generalizability theory’s origins is cited as culminating in the work of Cronbach, Gleser, Nanda and Rajaratnam (1972) (in Marcoulides, 1999; McDonald, 1999; Murphy & Davidshofer, 1998; Suen, 1990). The theory allows for a greater distinction between multiple simultaneous sources of error by resolving error into various components and determining the degree to which each component contributes error and yields more information about the measurement procedure as opposed to CTT (Cronbach, 1990; Marcoulides, 1999) hence its reference as a theory of multifaceted errors of behavioural measurement (Marcoulides & Drezner, 1997). Clearly, the influence of analysis of variance and the role that this method of data analysis played in the development of generalizability theory is obvious in its emphasis on levels of analyses (or facets within generalizability theory). In essence it is a statistical theory which is concerned with the dependability of measurement (Marcoulides, 1998) and has developed new ways of thinking of reliability (Crocker & Algina, 1986). It seeks to know how generalizable such measurement is (Lunz & Lincare, 1998) and is an extension of CTT (Embretson, 1999) but is more liberalised in its approach (Marcoulides, 1999). Apart from generalizability theory, the other most common model for analysis involving judgements is multifaceted Rasch measurement (Lunz & Lincare, 1998; Marcoulides & Drezner, 1997) which accounts for variations in such a way that bias within various contexts is removed from score interpretation. Reliability is considered within the test situation and error is considered the product of the test scenario.

Generalizability theory makes a less strong assumption about parallel tests and refers to this assumption as the domain-sampling approach. As a theory of reliability, CTT assumes that tests are random samples of items from a domain (Nunnally, 1978). Underlying this assumption is the notion that tests can be said to be randomly parallel if the items are random samples emanating from the same set of possible items (Suen, 1990). The ability to generalize from one set of results to another is the main focus of this theory and its initial aim sought to answer the question of the degree to which generalizations could be made from one set of results to another in various conditions (Murphy & Davidshofer, 1998). Analysis of variance is employed in generalizability theory in an attempt to account for multiple sources of error as opposed to the more conservative CTT approach (Crocker & Algina, 1986; Marcoulides & Drezner, 1997; Kerlinger, 1981). Other methods estimating variance components (a major aspect within this theory due to the information provided on error source variance) exist other than analysis of variance, such as Bayesian methods and restricted maximum likelihood methods, but analysis of variance has proved to be the easiest to handle (Marcoulides, 1999). Generalizability theory can account for various reliability queries whereas CTT can account for only one reliability measure, namely, Pearson’s correlation. Variance plays a major role in this theory due to the manner in which it accounts for accuracy (reliability) of measurement across scores; in other words how error is treated (Murphy & Davidshofer, 1998). Within CTT the rationale behind measurement is to assess people whereas generalizability theory accounts for the assessment of multiple sources of error because it is able to handle more dimensions than CTT. If measures change or new

dimensions are added, generalizability theory is capable of withstanding these extra sources of error, however the analysis will change depending on the context. By generalizing from one scenario to another, decisions are based on pre-empting measurement errors in new applications which is where the familiar G-study (from the generalizing of error) and D-study (the decisions subsequently carried out from g-studies; Crocker & Algina, 1986; McDonald, 1999) originates. Error source is extrapolated back from the use of the specific test, underlying the call to understand why and how the test is utilised rather than merely looking at scores (Murphy & Davishofer, 1998; Rust & Golombok, 1992; Suen, 1990). This is perhaps its biggest drawback however as few practitioners have the resource to continuously revalidate tests (Rust & Golombok, 1992; Kaufman, 2004). Moreover, it is considered a conceptually different approach to CTT as opposed to being statistically different (Murphy & Davidshofer, 1998). These added dimensions are accounted for in this theory. Variation across one such dimension can be construed as true variance and the rest are due to measurement error or facets of measurement. These facets are the equivalent to factors as utilised within analysis of variance. Facets are applicable within and across domains and hence the theory accommodates multiple levels (McDonald, 1999).

Depending on the number of dimensions, the number of facets of measurement will vary. Research designs employing multiple levels of assessment are amenable to such analysis especially where more than one rater is necessitated by a test across more than one level of assessment; an issue which has plagued dynamic assessment inter-judge reliability when attempting to identify cognitive deficits (Reschly, 1997). Although it is strongly advocated that IRT be looked at with greater regard by dynamic assessment researchers, generalizability theory can offer qualitative and more clinically aligned dynamic assessment practitioners greater scope in adjudicating reliability of scores across multiple levels of analysis depending on how the dynamic assessment intervention proceeds. X raters can assess y individuals on z instruments (including observations) across a number of occasions. Embretson (1999) points out that unlike most applications of IRT, generalizability theory considers the impact of various measurement conditions on the trait level and thus can analyse the differential effect of various measurement conditions on the psychometric properties of the test. Marcoulides (1999) highlights his work into models of generalizability theory which he views as an extension of the theory as well as a special instance of IRT especially where judgments play a role in scores. His extended generalizability theory is able to account for latent traits such as ability and item difficulty as well as rater severity. Wilson and Wang (1995) investigate via their model of IRT the issue of different modes of assessment and include rater estimates with some very interesting results as they compare multiple choice items, rater estimates and open ended questions and the resulting IRT functions. Embretson (1999) cites this new work in generalizability theory as not having been thoughtfully considered by the broader community and although there is literature on the matter, it is relatively recent. It is emphasised that attention be turned towards the newer models of generalizability theory as well as IRT models within dynamic assessment, specifically generalizability as it can accommodate qualitative ratings which is an area of concern to dynamic assessment in its more clinical forms.

4.4.2.3 Modern test theory

The ideas behind modern test theory as exemplified in item response theory are not new and extend back to Spearman's 1904 paper on factor analysis, who in addition also pioneered many procedures within CTT (Borsboom, 2005; Embretson & Reise, 2000). The notion of a latent trait underlying various observed behaviours has remained with test theory in a variety of guises (having been theoretically and practically constructed in more primitive form in the 1920's in Thurstone's work; Embretson, 2004). Its modern renditions still assume a latent trait to operate meaningfully within instruments assessing for various traits (Borsboom, Mellenbergh & Van Heerden, 2003) and in many instances this underlying trait is considered unidimensional (Hambleton & Rogers, 1991). However below will be shown the need for and development of multidimensional trait models specifically to estimate changes in ability, or potential. Interest in latent trait models dates to the late 1970's and early 1980's but due to its perceived technicalities only later filtered through slowly into the realm of intelligence research (Whitely, 1980). Kormann and Sporer (1983) for instance comment on the intractable problem of dynamic assessment test validity and the lack of a general theory of change measurement even though attempts within IRT had already started to address these issues; but in all fairness, the techniques were decidedly under-developed as far as technological implementation was concerned and the authors do note the then-available contributions to the debate.⁹⁵ The main difference between latent trait models and CTT is that the former attempts to account for the mechanisms which generate the data (Borsboom, 2005). In other words, it is a top-down model or process as latent trait models or theory illustrates how parameters relate the latent variables to the data according to the data generated under the particular model (model fitting) (Borsboom, 2005). This is of particular importance currently as psychometric theories of intelligence have mostly prescribed construct definition via statistical and measurement techniques rather than the construct defining the necessary psychometric structures (Hunt, 2005). However, in all its sophistication the broader field of IRT is nevertheless concerned with mathematical models of ability traits (Carroll, 1996). Moreover, scaling is

⁹⁵ Interestingly enough, one of the researchers cited as having contributed to the measurement of change debate in Kormann and Sporer's (1983) article is one R. Schwarzer whose meta-analytic software programme is utilised in Appendix 1. Also of interest is the fact that as early as 1945 analyses at the level of item responses was considered as warranting attention from intelligence researchers (Rapaport, Gill & Schafer, 1945 in Kamphaus, Petoskey & Morgan, 1997).

justifiably interval-level based, parameter estimates are invariant and a common scaling of items and individuals is made possible (Embretson, 1996).

Paralleling trends within the factor analytic tradition, IRT models developed in various ways resulting in models that could deal with dichotomous variables, polytomous variables as well as categorical variables (Borsboom, 2005). Conditionalising on the latent trait scores (assumed to underlie the observed variables) will result in the statistical independence of the observed variables and is referred to as the principle of local independence, specifically within a subpopulation located at a “single point on the latent trait scale” (Crocker & Algina, 1986, p.343). Local independence allows for information pertaining to the conditional individual responses to be sufficient in determining the conditional probability of the response patterns for a set of items (Glas, 1997; McCollam, 1998; Van Schuur, 1997). In other words, responses to items are independent of one another (Embretson, 1983; Pfanzagl, 1994). If the latent variable is held constant the probability of endorsing two or more items is the product of the probabilities of endorsing each item (Coombs, 1967). Factor analytic treatment, confirmatory factor analysis, the parallel development of latent variable analysis with dichotomous variables and the resultant growth of item response theory utilised for both continuous and categorical and polytomous observed variables can all be traced back to the fundamental notion of a latent construct (Borsboom et al., 2003; Heck, 1998). The notion of a latent construct is a burdensome assumption to make especially as so much theory growth and development has centred around it and Cronbach and Furby (1970) mentioned similar comments as to the unlikely event of ever truly defining constructs via psychometric models despite their sophistication and development. This was stated over thirty years ago and is still being heatedly discussed in the literature today. For instance how is the measurement of latent traits to be interpreted? Borsboom (2005) offers two interpretations, which he states, is largely semantic.

- i. Stochastic subject interpretation
 - a. The person is central to this interpretation. Item responses are regressed on the latent trait and the probability of a correct response will change as the latent trait changes
 - b. Similarly to CTT though, the score on item j for subject i , $\Sigma(U_{ij})$ is exactly the same as i 's scores on j IF the latent variable model is true
 - c. Characteristics of individuals are being modelled
- ii. Repeated sampling interpretation
 - a. Here the latent trait model is reconceptualised as a repeated measures interpretation where the repeated sampling is central
 - b. The focus is on populations as opposed to individual subjects as in the stochastic interpretation
 - c. This interpretation focuses on the repeated samplings of item responses from populations evidencing probability distributions which are conditional on the latent variable. These populations all have the same position on the latent variable and the parameters are thus related to the latent variables
 - d. Repeated sampling from the same population will thus result in sub-population means
 - e. The crux: it is not the individual subject's probability that is of concern; rather the focus in this model interpretation is the probability of drawing a person that endorses the item from a population with the latent trait level sampled
 - f. In other words; population means on θ is distributed f over item responses u_{ij} ; hence the expected item response is $\Sigma(u_{ij} | \theta_i)$
 - g. Person-level random variation is not accounted for
 - h. Sub-populations means are being modelled

Substantiating the theoretical entities which are assumed to underlie the observables (learning potential) is the task of psychological assessment but doing so through conventional means, such as IQ test concurrent and predictive validity (Embretson, 1992; Fernández-Ballesteros & Calero, 2000; Guthke, Beckmann & Stein, 1995; Hamers, Hessels & Tissingh, 1995; Hessels, 2000; Jensen, 2000; Resing, 2000) or correlating results from widely used intelligence batteries is dangerous in terms of ill-defined initial constructs present in these original studies or criteria (Guthke, 1982; Kline, 1998; Pennings & Verhelst, 1993; Wallis, 2004), as has been argued in this thesis thus far. This manner of validating the construct of intelligence is circular in argument (Heitz, Unsworth & Engle, 2005). Rather, performance change criteria should be utilised (Hamers, Hessels & Pennings, 1996). Guthke, Beckmann and Stein (1995) conclude that an external concurrent valid criterion is unlikely ever to be found for learning potential tests and as such construct validation should proceed along the lines detailed by Cronbach and Meehl (1955). This author maintains that such external learning potential criteria can be located or developed and will not necessarily have to resort to nomological networks of findings. As mentioned, researchers such as D. Borsboom concur. One need only look towards the reductionists within physiological psychology to find common ground or other process-based measures of intellectual development. The utilisation of change-based IRT models is also another avenue to pursue (more on this below). Criterion validity within dynamic assessment thus poses a difficulty within the mainstream set-up (Lidz & Thomas,

1987). Dynamic assessment would perhaps be better off in utilising ipsative⁹⁶ assessment, which is of course a move away from standardisation. If, as the Spearman effect attests to, tests loading higher on g tend to discriminate more regarding differences between people, such g -loaded test are already biased against the supposedly disadvantaged groupings dynamic assessment intends to assess as it has been reported in the literature over a span of many decades that certain groupings tend to perform poorly on intelligence tests in comparison to other groups (Loehlin, 2004). Why would we stack the odds against ourselves by utilising such g -loaded test in the first place? Of course there is no simple answer to this particular dilemma. Embretson (1983) refers to this concurrent construct validating procedure as the nomothetic span of the test which is indicative of the degree to which the test can differentiate between individuals. Intelligence tests usually have very good nomothetic span⁹⁷ but are not always able to specify the underlying construct being assessed. This is in addition to the literature replete with near zero correlations between tests assumed to measure the “same” construct (Guilford, 1982). If the underlying construct of initial ability, intelligence, is different to the underlying construct of learning potential then it is hardly surprising that the two do not correlate meaningfully in many instances as the constructs are supposedly measuring different traits (Guthke, Beckmann & Stein, 1995).

Does the substantive theory necessitate the model (Borsboom & Mellenbergh, 2004)? This is echoed in the same plea for substantiating the theoretical hypothesis within null hypothesis significance testing, which as we have seen above, is not a falsifiable manner of progress. IRT models are in fact falsifiable (one of the rare occasions within the psychological discipline where such as situation prevails) as the IRT model employed to fit observed data may not necessarily fit the data and thus not adequately predict or explain the data (Hambleton, Swaminathan & Rogers, 1991). Does the latent trait underlie the observable variable or do we construct the latent trait from the observable? Borsboom et al, (2003) discuss at length the philosophical issues surrounding just such a notion by ascribing to various interpretations (constructivist and instrumentalist) but agree that realist interpretations of this notion can be the only solution. The latent trait can never be assessed directly (otherwise there would be no use for tests!) but it can be assessed in terms of the joint probability of the results it implies via joint measurements made indirectly (Michell, 2000). We still have not adequately addressed the issue of which is paramount. The trait’s manifestation through observables or our constructing of the trait via the observables. Addressing the latent trait from a constructivist position entails a dependent trait and not an independent trait which is of course not sound science and secondly, addressing the trait from an instrumentalist position results in numerous operational definitions for the very same construct (as each operation will need to be defined for each test) (Borsboom, Van Heerden & Mellenbergh, 2003). Entity realism ascribes to a correspondence of truth view and as mentioned is the view taken by the authors. Purely mathematical and statistical analysis will relinquish the need to even describe the trait in any form other than one of mathematical concern. However, psychology is clearly not a sub-discipline of mathematics and it is very easy to see how the discipline often loses itself in a morass of statistical and mathematical trickery.⁹⁸

The Rasch model as exemplified within item response theory is, as mentioned above, a form of conjoint measurement as it ascribes to the necessary requirements for additivity over and above its ordinal status. The Rasch model is not a model constructed in isolation from its historical context. “Test-free linear measures were latent in Campbell’s 1920 concatenation, In Fisher’s 1920 sufficiency, in Levy’s 1937 and Kolmogorov’s 1950 divisibility; clarified by Guttman’s 1950 conjoint transitivity; and realized by Rasch’s 1953 additive Poisson model” (Wright, 1997b, p.43). Borsboom and Mellenbergh (2004) state that the Rasch model is not the only IRT model to avail of itself of conjoint measurement and that there are other probabilistic latent variable models which can be used. Ability and difficulty level are both continuous quantitative attributes in a monotone model of which the Rasch is one such example. If Jane’s probability of endorsing an item increases with the difference between her ability and the difficulty of the item, then so it must be with Peter also, dependent on the fact that the difference between Peter’s ability⁹⁹ and the difficulty item is not less than the difference for Jane. This is referred to as a monotone item response model (Cliff, 1992; Michell, 2003a). Algebraically this looks as follows:

$$P(x_{ij} = 1) \geq P(x_{hk} = 1) \equiv (a_i - d_j) \geq (a_h - d_k)$$

It can be seen how the Rasch model and its additive conjoint characteristics allow for two simultaneous attributes to be quantitatively concatenated, thus allowing for fundamental measurement. This notion of double cancellation is extremely important in this argument and is the reason why it was explained at length above. The following argument is Michell’s (2004) but is considered so imperative in this discussion on dynamic assessment’s status that it needs to be précised. Recall in the discussion on classical test theory above that error is unknowable and that true score derivation and interpretation is tentative at

⁹⁶ Ipsative is being referred to in its educational sense in which assessment is compared to the individual’s past performance and does not refer to self-ratings by the individual.

⁹⁷ From a g point of view this is hardly surprising. The retort to this would be that g is the common factor.

⁹⁸ Is there really still a “psycho” in psychometrics? Consistent with Blinkhorn’s (1997) concern about the tenuous link between test theory and psychological theory. Suen (1990) maintains that there is a consistent move towards cognitive theory-based psychometrics but that psychometric education programmes will need to consider more closely the issue of mathematical programming into subsequent courses.

⁹⁹ Ability is used throughout the discussion, but IRT is successfully utilised in personality assessment as well where “ability” is merely reflective of endorsing an item or not and in this context does not refer to ability as is understood within intelligence assessment (Hershberger, 1999).

best and egregiously wrong at worst. The one parameter Rasch item response model is a probabilistic model which is predicated on the probability of an individual endorsing an item dependent on the individual's purported ability level and item characteristics. The individual's ability as well as random error is taken into account. If error was not considered and the individual's response to an item was endorsed then it could be said that the ability equaled the difficulty parameter of the item. Thus one only knows an ordinal fact. However, item responses modelers assume quantitative information (and not merely ordinal - see the above discussion), from the distribution of the error component. If the model is true then the error distribution reflects the quantitative structure of the attribute but if the attribute is not quantitative (and so far, potential hasn't been proven as such) then the supposed distribution of error is completely vacuous. "Here, as elsewhere, psychometricians derive what they want most (i.e. measures) from what they know least (i.e. the shape of "error") by presuming to already know it" (Michell, 2004, p.126). Now, if random error is retained, but the shape is admittedly unknown then the best we can do is to assume ordinal structure of the attribute in question (potential) unless the double cancellations as discussed above manifest and hence, even utilising IRT models is suspect when it comes to assessing for psychological constructs. A partial solution to this has been put forward by Mokken (1971) (in Michell, 2004) in the utilization of a non-parametric item response model, which, according to Michell, has been largely ignored. The core differences inherent in IRT (new rules) and CTT (old rules) can be summarised as follows (Embretson, 1997b):

- i. Old rules - CTT
 - a. Rule 1 - standard error of measurement applies to all scores within a population
 - b. Rule 2 - if reliability needs improvement, the test is usually made longer
 - c. Rule 3 - in order to successfully compare tests utilising different forms, test parallelism and/or test equating is necessitated
 - d. Rule 4 - score distributions are necessary if particular scores are compared
 - e. Rule 5 - if the data yield normal raw score distributions then interval level interpretations can be made
- ii. New rules - IRT
 - a. Rule 1 - standard error of measurement generalises to the population but differs according to response patterns
 - b. Rule 2 - as IRT is item-based, making the test longer is ineffectual. In fact it can be shown that shorter tests are more reliable
 - c. Rule 3 - test equating is optimal for different ability groups
 - d. Rule 4 - item based scores are paramount. The distance from items are required and not an entire score distribution
 - e. Rule 5 - interval level scales are inherent in the model (conjoint measurement ensures a defensible and justifiable manner of quantifiable measurement)

Can dynamic assessment which is predicated on change-based measurement and the derivation of potential not avail of itself a manner of such measurement? In this way, at least adherence to robust modeling is accounted for, it remains within a scientific nomological network of scientific progress, pursues replicable research, and still maintains its qualitative nature of clinical intervention assessment mode. This is one such attempt to place dynamic assessment within mainstream intelligence assessment whilst still participating fully (and perhaps even more so) in a scientific sub-discipline. What is being addressed here is firstly the need to determine whether psychological constructs are measurable and secondly to determine a method suitable for allowing such quantitative structure to become manifest which is precisely the dictates set forth by Michell (2004). His argument centres around the need to substantiate the requirement for additivity and to thoughtfully ask if such requirements are being met. His answer to this is that for much mainstream psychometric practice this issue has been ignored. It is not so much that quantification of attributes cannot be made as much as the fact that these questions are never asked in the first place.

Dynamic assessment can embrace conceptions of the nature laid down by Michell (2003, 2004) and can utilise the recommendations set forth. Such an avenue is available and taking cognisance is what is now necessitated. However, one must not fall into the same trap and declare that dynamic assessment constructs are themselves amenable to quantification. After all, adhering to a model which requires two simultaneous attributes and in so doing making manifest a third measure does not mean that the two original attributes are quantifiable. *All that is being done is arguing for the quantifiability of the model and not the traits.* In other words, quantification can never be detected directly but only via a means of monotonic transformations of the observed variables which fit the model (Borsboom & Mellenbergh, 2004; Cliff, 1992). This is highlighted because it is the crux of the argument. Is potential, like intelligence, measurable? The monotone Rasch model may well avail of its structure to make allowances for the additive conjoint structure of the attributes but what of the attributes? "At present, psychometrics does not possess theories of sufficient richness" (Michell, 2004, p.124). Items are usually constructed using face validity (it is an art rather than a science; Chipman, Nichols & Brenna, 1995; Embretson, 1983), of what we think measures something, of what we assume to be the case and what experience has presumably taught us. When items are dropped from item banks (item attrition hovers between 30% - 50%; Embretson, 2005) rarely is it the case that the actual reason for its being dropped is known or even questioned (Meier, 1994). If this is the case for the dropping of items, what can be said of the retention of others? What makes these items suitable? Luck, chance or statistical spread? It is unnerving to think that the situation presents itself in this fashion.



The unresolved question of validity makes its appearance in this particular segment of the argument. Utilising criterion measures as correlations of predictive success is hardly much of an answer to the question of what it is we measure (Barrett, 2005). How do we know for sure what it is the construct is? To unequivocally state that test A predicts test B is saying absolutely nothing about what underlies test A or B at all. Our tools may be robust, developed and highly advanced, but the point of origin which is the fundamental question of quantification is not being addressed, at least not adequately. The threefold problem looks as follows (Barrett, 2005):

- i. We do not know the nature of the criterion against which we wish to test for predictive or construct validity
- ii. We assume it is additive and hence measurable
- iii. We causally link the new construct to an “established” criterion and do so within what is perceived to be a normative framework (there is no normative framework, recall our lack of progress in even defining what is meant by intelligence). A nomological network does not entail construct validity (Barrett, 2005; Borsboom, Mellenbergh & Van Heerden, 2004), because just by saying that we all agree $x = y$ does not make it so. Science does not work that way and should not work this way for psychology. One need only think about item bias and how items which are considered biased are removed from item banks for certain groups. Are the items ever looked at in terms of why they may have led to under/overprediction in the first place? This would entail going back to the nature of the construct. Most often these items are merely removed (Van de Vijver, 1991)

How do we adhere to a scientific discourse as discussed in chapter 3? Such deep philosophical issues are also found pertinent by Borsboom and Mellenbergh (2004). What of the discipline in this regard? Once again, the need is felt to state that assessment as sub-discipline should lodge itself firmly within one camp or the other, not both:

- a. it is either a scientific field of enquiry molding closely to the natural sciences and is able to prove or at least spring from an axiom of measurability (recall the discussion on improvable axioms in the section on mathematics)
- b. or it should retract its position within this realm and focus on its exclusive qualitative clinical interventionist strategy format to aid individuals in reaching what is now considered to be their potential but in a newer light.

Regarding axioms in (a) above, Borsboom and Mellenbergh (2004) say as much when they state that certain axioms will just have to be adhered to in order for the continued use of a probabilistic model such as certain IRT models. The accepted “fact” within the model is its assumption that what it measures is in fact quantitative. But is this good enough? After all this is precisely the concern that Michell (2003, 2004) has and it is not a concern to be shrugged at and swept under an assumption within a model. Does the model purport to *represent* (an isomorphic relation that is mapped from the attribute to the numerical?) or does it purport to *explain*? (Borsboom & Mellenbergh, 2004). This takes us back to chapter 3’s discussion on the veracity of psychology as a formal science - does it seek to state or explain? This depends on its nature: instrumental, realist, constructivist and so forth. One cannot isolate such a thesis by engrossing oneself in the level of explanation attained only in this chapter. The need to look at a broad spectrum is warranted which is precisely what has been done thus far. It is at this point, where the issues discussed in chapters 2, 3 and 4 cohere. Right here in the debate on the nature of quantification, the nature of the tools used and the nature of what it is we psychologists are trying to do within science. The question asked here cannot and should not be asked only of assessment or dynamic assessment in particular but of science in general (Borsboom & Mellenbergh, 2004).

4.4.2.3.1 Modern test theory change models – an answer for dynamic assessment’s change score problem

As prelude to this section on modern test theory’s attempts to address change within a test, Kormann and Sporer (1983) offer a description of the state of affairs regarding the measurement of change. Even though theirs is a dated one the historical record is nevertheless accurate and hence warranted as it depicts the gradual process of development of ideas towards the understanding of change assessment originating from classical ideas to modern test theory ideas of change. Figure 68 below illustrates this.

Figure 68 Kormann and Sporer's (1983) overview of change measurement as viewed both classically and alternatively¹⁰⁰



¹⁰⁰ Citations in this figure are not referenced in this thesis.

- *Embretson's multidimensional Rasch model for learning and change (MRMLC and MRMLC+)*

Reflecting on the gain score issue as discussed above in 4.4.2.2, the need for dynamic assessment to answer psychometric questions has been partially addressed since the late 1980's but has been known for many years prior (Guthke, 1982; Embretson & McCollam, 2004; Wilson, 1989) with modern test theory models allowing for changes in ability within a reliable model which attends to the issue of the change score paradox (Bereiter, 1962, 1967; Embretson, 1992; 2000; Embretson & Prenovost, 2000; Lidz, 1991). It is not a novel idea and perennially resurfaces when aided by the requisite technology but which will hopefully not fall in to the trap envisaged by Cole and Valsiner (2005) who summarily state that old and worn-out theoretical ideas resurface when in fact they should not. The most important aspect within these general models is their original concern with theory. Specific cognitive developmental theories (and more emphasis on cognitive processes assessment in general; Roberts, Markham, Matthews & Zeidner, 2005) posited by developmental psychologists serve as point of origin which is becoming increasingly focused on within modern psychometrics (Draney, Pirolli & Wilson, 1995; Embretson, 1994, 1997a) as opposed to the traditional (and still current) notion of measurement implying inference from observations to theory (Ippel, 1991). Theory is attended to via the tool and not the other way round (Wilson, 1989) as evidenced from early work on intelligence assessment which occurred largely in theoretical voids (Kaufman, 2004). Confirmatory IRT models assume that performance is determined by underlying abilities or traits which can be accounted for through the model (Kline, 1998), such as is the case with Embretson's multidimensional Rasch model for learning and change (MRMLC) (Embretson & McCollam, 2004) which is a special case of structured latent trait models (SLTM) (Embretson, 1997a, 2000). The development of confirmatory factor analysis has led IRT researchers to utilise confirmatory as opposed to exploratory factor analytic models. The former were typically the models utilised in early intelligence research (Keith, 1997) and is rightly stated to be an imperative tool towards the understanding of intelligence research in general (Schulze, 2005) precisely because of the link between theory and model and its respective philosophical underpinnings (Mulaik, 1988). Confirmatory models allow for hypothesis testing, whereas exploratory models' underlying nature and number of factors is unknown (Nunnally, 1978) relating back to sciences' origins as inductive and deductive knowledge-acquisition device.

Theory predominates over model and confirmatory models allow researchers to model the underlying theory (Heck, 1998). Due to the MRMLC'S constraining of item discrimination to unity, evidence of its familial tie to the Rasch model¹⁰¹ can be seen (Embretson, 1991a, 1991b) as the Rasch model is well suited to incremental development within individuals (Bond, 2000). The need to impose such linear constraints is due to the need to allow hypothetical factors (latent trait, ability, intelligence or potential) to influence the observed responses (Fischer & Tanzer, 1994). However, similar constraining of modifiability averages is not permitted as their increases or decreases (Sijtsma, 1993b) over time are reflective of the latent response of potential (Embretson, 1991a). It is pertinent to recall the vast literature on ability change over tasks where correlations decrease between ability and task performance as skills develop (Ackerman, 2005). Its multidimensional status is implicitly implied within dynamic assessment as there is more than one ability accounted for (Embretson, 1987). Dimensionality of a test refers to the number of traits necessary to achieve local independence, however, local independence and dimensionality are not the same concept. The one follows on from the other. Validity within psychological constructs is based on the assumption that there exist x amount of latent traits which account for items being locally independent (Crocker & Algina, 1986). It is worth reiterating that no matter what model, no matter how sophisticated the mathematical modelling, no matter how grand the statistical techniques; construct validity remains with the realm of theoretical speculation. Rasch himself emphasised this in his book on probabilistic models where it is stated that only an *estimate* is given of θ , "in fact the best estimate obtainable but θ itself can never be determined empirically" (1980, p.77). Note his confidence in that the facility available to estimate θ is the best one available (and at the time is was) but that this recognition in no way impinges on empirical construct validity. This loops back to the discussion on psychology's methodology within science. Empirical testing of constructs might well be possible and indeed valid within a realm concerned with the physiological aspects of physiological measurement (see chapter 2). However, the case for dynamic assessment and its agenda is far from such empirical methodology and should maintain focus on qualitative endeavours, yet remain scientific in its own right. Rasch, it seems, was saying just this. Models aid, simplify and ease a number of issues, but even extensions of generalizability theory and various IRT models cannot yet account for valid constructs. Dynamic assessment's "potential" constructs would be better founded and validated within a more qualitative set-up. In essence what is necessitated by such a model of change is what Fischer (1997)¹⁰² describes as an ability which is viewed "by a point in a multidimensional parameter space, and development being understood as a migration through that space" (p.189).

¹⁰¹ The traditional unidimensional Rasch model did not allow for item subset comparisons "which are required to operationalise cognitive processing variables"(Embretson, 1997a, p.225). Hence the improvement along with the need to measure multiple dimensions (learning potential or change in ability).

¹⁰² Fischer (1997, p.190) gives a compact outline of the various models he and his co-workers have devised. Dichotomous, polytomous and frequency data are modelled according to the number of occasions individuals are assessed as well as the nature of the items utilised (same or different items).

Unlike its predecessors such as Fischer's (1972) linear logistic latent trait model with relaxed assumptions (LLRA), Embretson's MRMLC is able to account for individual differences within training. The LLRA cannot measure individual differences and assumes all subjects learn at equal rates which is clearly not in keeping with dynamic assessment theory (in Embretson, 1991a). Likewise, Anderson's (1985) multidimensional Rasch model for repeated administration of the same items to the same individuals over a period of time was unable to account for change parameters for individuals although it was able to determine the impact of time or treatment on the theta distribution (in Embretson, 1991a). Initial ability along with modified abilities are assessed in which the initial ability and subsequent modified scores are related via the Wiener process¹⁰³ (the design structure for the MRMLC; Embretson, 2000) which increases over time and across conditions (variance will increase alongside decreases in correlations). The Wiener process allows for each ability level to be assessed seeing as it is being modified throughout (Embretson, 1996). As with the Rasch one parameter logistic, the MRMLC is defined by an initial ability level and the difficulty level of the item. However, in addition to the one parameter model, successive modified abilities are weighted between successive conditions and via the Wiener process ability levels are differentially involved. The generic structure Λ would like something as follows:

Table 18 Structure change dependent on modified ability through successive occasions

Λ	Occasion	Ability		
		θ_1	θ_2	θ_3
	1	1	0	0
	2	1	1	0
	3	1	1	1

If a typical dynamic assessment intervention worked perfectly, the structure in Table 18 would manifest, reminiscent of the perfect Guttman scalogram (Coombs, 1967; Kerlinger, 1981). Due to the model's similarity with the Rasch-family model, measurement is justifiably interval-level applicable. The initial ability θ_1 is utilised in all three occasions; θ_2 is the second ability which is θ_1 in addition to modification and θ_3 is θ_2 in addition to modification. θ_1 remains involved in each subsequent condition as does θ_2 and θ_3 in their subsequent inclusions (Embretson, 2000). The initial ability level (pretest score) is thus utilised throughout the successive modifications. The general structure for Embretson's (1991, 1992, 2000; Embretson & McCollam, 2004) MRMLC is as follows:

$$P(X_{i(k)} = 1 \mid \theta_j, \beta_i, \lambda_{i(k)m}) = \frac{e^{(\sum_m \lambda_{i(k)m} \theta_{jm} - \beta_i)}}{1 + e^{(\sum_m \lambda_{i(k)m} \theta_{jm} - \beta_i)}} \quad 104$$

θ_1 is the initial ability level; $\theta_2, \dots, \theta_{jm}$ represent the modifiable abilities across different conditions; β_i is the difficulty for item i which is considered as constant across the various conditions; $\lambda_{i(k)m}$ is the weight of the latent ability m in item i within condition k which is specified as either 0 or 1. The structure Λ represents "matrix weights for the items within the k conditions in estimating the m abilities" (Embretson, 2000, p.513) as seen above in table 18. Dynamic assessment theory pivots the central notion of change of ability and here the assumption is being made that potential can be thought of as an ability. This is an issue some mainstream psychometrists contest (Embretson, 1987, 1995), hence the term multidimensional in the model, it is assessing more than one dimension; namely potential (Embretson, 1995). Originally, unidimensional models were considered robust enough to manage multidimensional traits especially when traits were highly correlated¹⁰⁵ but this is not always the case especially in change-models (Adams, Wilson & Wang, 1997). Ability and potential represents two variables, pretest scores are unidimensional and posttest scores are multidimensional (due to the introduction of modifiability) (Smith, 1997). The most common reason for the use of change scores is to operationalise the concept of such an ability (Cronbach & Furby, 1970; Pennings & Verhelst, 1993). It is essential to do so as dynamic assessment construes learning potential as an ability (Hamers & Sijtsma, 1995). The question now raised would be: if intelligence is not definable as a construct, is learning potential? Moreover, if learning potential is linked to intelligence in some form or another, the questions asked of intelligence research can be directed at dynamic assessment. If

¹⁰³ Embretson (2002) notes that although the Wiener structure is used in the MRMLC it is not the only system that can be used. Others include the Helmert structure which may possess better measurement properties because of the greater number of items in the estimation of ability and modifiability but these modifiabilities are more global as opposed to condition-specific. "The Wiener simplex model is particularly appropriate for structuring parameters in a Rasch model since the properties of equivalency of measurement scales between tests and additivity of the effects are compatible between the models" (Embretson, 1991, p.499). Another design embeds an orthogonal polynomial design where abilities are represented as global, linear, cubic or quadratic change. Data fit for each system is not a concern but the choice of system is determined by measurement error which varies across the systems and secondly, theory should dictate which system should be used.

¹⁰⁴ For the sake of consistency, notation representing ability (θ) and item difficulty (β) will be kept consistent throughout even though authors may choose to utilise other notational forms.

¹⁰⁵ For intelligence assessment is this yet again reminiscent of g ?

dynamic assessment ceases to conform to mainstream assessment it might not have to even answer the question at all. Embretson's (1992) mathematical modelling of these manifest changes from pre to posttest has evidenced the change in the construct representation of the ability which results in changes of test validity as well. Embretson's (1991a, 1992, 1995, 2000) model:

- Predicts pre-posttest reliability by factoring in modifiability within the multidimensional scale through additional dimensions in the posttest results and does so simultaneously using maximum likelihood¹⁰⁶ for all item responses over time. Via maximum likelihood estimation, error variance for abilities is known. As a repeated measures design, means and standard deviations increase across occasions
- Accommodates varying meanings attributable to changes in raw scores at different initial ability levels. In other words ability estimates are invariant across the sets of items and these ability estimates are not biased due to the difficulty level as they would prove to be in CTT and because the scale is no longer an ordinal one, interval-level changes are meaningful. Irrespective of initial ability levels, modification has the same impact on log odds on any of the items which is a direct consequence of utilising the Rasch model. Initial ability and modified abilities are now additive (see section 4.4.1.2 above on the importance of additivity for quantification of constructs)
- As per dynamic assessment theory and because of its IRT model status, there is an expectation of different raw scores for persons who manifest equal levels of potential or modifiability when their initial ability levels differ. In addition to this, different abilities can be estimated from the same items because it is the response pattern that is assessed (Wright & Stone, 1979). This allows for change to be viewed as a separate ability
- Removes scaling artefacts that contribute negative spurious elements within the correlation change scores and are thus partialled out. Problems associated with measuring change are essentially due to scaling and regression (Klauer, 1993; Schöttke, Bartram & Wiedl, 1993)
- Performs the functions inherent within unidimensional IRT models as performance is ascertained from response patterns and is not based on linearly derived total scores. Due to the original Rasch unidimensional model's lack of flexibility (Wang, Wilson & Adams, 1997) the MRMLC provides more robust initial ability estimates as well as introducing modifiability estimates which are corrected for differences in initial ability
- Allows for confidence intervals to be estimated for each person's ability and modifiability estimations and these confidence intervals can be compared across design structures (recall the plea in section 4.3.1.2 where confidence intervals were increasingly called upon to perform the function of power in NHST)
- Yields different results from CTT when viewing change scores, "statistical results depend on how change is measured" (Embretson, 2000, p.518)
- Does not repeat items. It does so to avoid the classic problems associated with repeated measures which include practice effects, retention of items and response consistency effects. Local independence, one of the main characteristics of IRT is thus upheld

The model would:

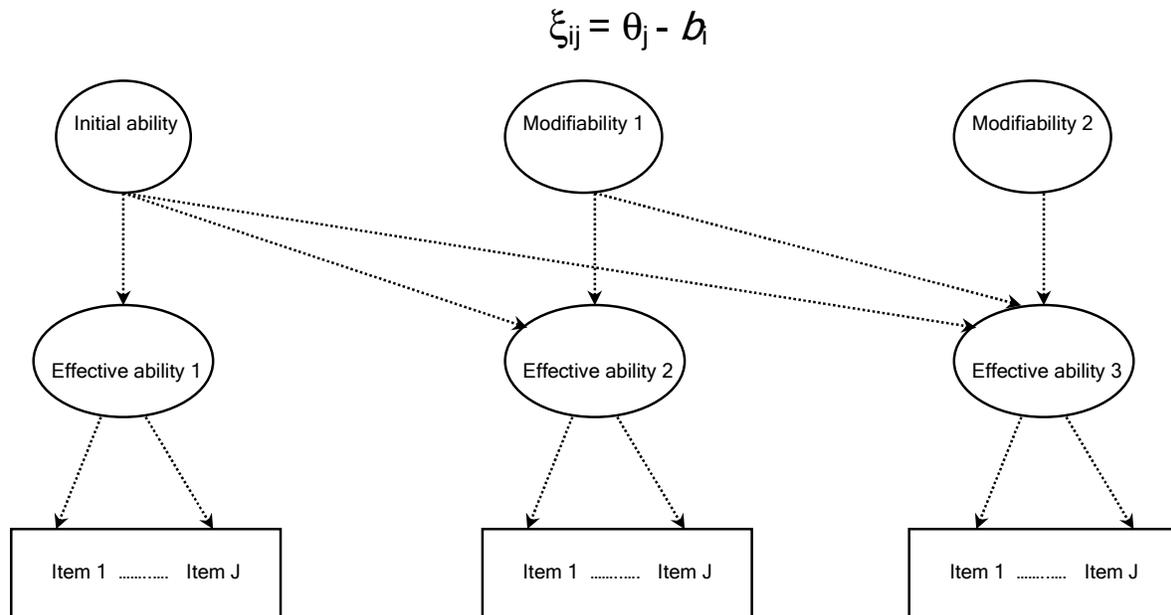
- be better utilised within situations where modifiability was the main target of change measurement
- not necessarily yield greater predictive validity with fixed content tests, as pretest and posttest scores would suffice
- need multiple groups so as to observe each item within each condition
- need to condition on the ability at each occasion within each group in order to estimate item parameters
- be limited in its efficacy to link ability changes or learning to requisite changes in the cognitive processes underlying such changes. In other words, the theory-substantive fit could be better defined

Embretson's (1995) MRMLC+ extends functionality of the MRMLC discussed above by resolving issues surrounding individual change and allows for adaptive testing so as to estimate learning which it links to the substantive changes in processing and knowledge. In other words it links the difficulty of the task to features that influence cognitive processing (Embretson, 1996). The improved version of MRMLC includes structural models of item difficulty and includes the facility to link changes in the model to changes within the construct (Embretson, 1995). MRMLC originally estimated the difficulty parameters for each item whereas MRMLC+ contains models for both person and item. Difficulty item parameters are replaced with new model predictions. Figure 69 illustrates the relation between initial ability level, modified abilities and the subsequent effective ability which results from the accommodation of initial and modified abilities in the model (Embretson, 1991b). Cliff (1991a) refers to this process as a path

¹⁰⁶ Recall Bayes' theorem. The maximum likelihood approach is a special case of Bayes and is used under certain circumstances (Suen, 1990). Three methods of maximum likelihood are utilised for Rasch models: joint maximum likelihood (point estimates for persons and items), conditional maximum likelihood (item estimates only) and marginal maximum likelihood (item parameter estimates and sample distribution parameters) (Roberts & Adams, 1997). Bayesian parameter estimation is also another model which uses a uniform prior function but it is recommended that the distributions for θ and b be normally distributed (Suen, 1990). Once again it is evident how important Bayes theorem is to modern day test theory; a theorem, which in Bayes' own time, was all but forgotten. Maximum likelihood is not the only mechanism through which estimations of parameters can be made. Heuristic or approximate procedures can also be utilised (Crocker & Algina, 1986).

model due to each stage in the process being reliant on the values of the previous stage. The current score reflects the previous score in addition to an increase that presumably arises from mediation (or some treatment or in the case of lifespan development, maturation) that has occurred in the interval since the previous measurement. It is worth asking the question at this stage: what happens if mediatory intervention makes no difference or results in a decrement? This echoes Collins' (1991) question as to whether development is reversible and how longitudinal change-based models are able to effectively handle such data.

Figure 69 Effective ability as consequence of initial and modified ability within the Rasch model (Embretson, 1991b, p.189)



The South African study by De Beer (2000) which, as mentioned above, utilised IRT as means of constructing and validating a dynamic assessment test recognised the importance of Embretson's model but differed in its emphasis on learning potential as exclusively defined by the difference score. De Beer's (2000) study utilised the pretest and difference score in defining learning potential and not just the change score. De Beer's (2000) study utilised computer adaptive testing (CAT) as platform for IRT response modelling for a dynamic assessment and as Embretson (2005) states, the future of CAT has already proven expedient (Drasgow & Olson-Buchanan, 1999; Kingsbury & Houser, 1999). The next phases in this development are underway, namely, adaptive item generation which generates new items which avail of the most information pertaining to the individual. Not only does the computational system, which is premised on artificial intelligence, choose the next appropriate item but it does so by merging psychometric theory with substantive cognitive theory. It is paramount that the stimulus value of the item in terms of its cognitive consequences on item functioning be understood. Embretson's (2005) item generation system is based on "the cognitive design system approach" published in 1998 (in Embretson, 2005). Cognitive design emphasises the impact of the underlying processes involved in solving items, the impact that this process has on performance as well as considering the impact of the stimulus features on the process (Embretson, 2005). This approach takes information from both worlds so to speak:

- ⇒ the psychometric world gives to the design information on item difficulty and item discrimination
- ⇒ the cognitive world gives to the design, information processing difficulty which is derived from an information-processing approach to the study of cognitive. This stresses problem solving, use of abstract reasoning and working memory capacity for instance
- ⇒ items which embody the same sources and levels of cognitive complexity emanate from the same structure type but these structures can differ qualitatively from item type to item type. Research is ongoing in an effort to bridge the development of these structures to variables within the cognitive model

Of note to this study in particular, is that this approach has been extensively applied to advanced matrix tests including the Raven Advanced Progressive Matrix Test, a test often utilised within dynamic assessment studies and also considered a good indication of intelligence (of both g and Gf ; Wilhelm, 2005) within mainstream intelligence assessment. The elements thus far:

- ⇒ development and deployment of substantive cognitive theory (see chapter 3 and its discussion on substantive theory as well as chapter 4 which argues for substantive theory development alongside its



statistical entity status within statistics). Embretson (2005) supports Cronbach and Meehl's (1955) treatise on construct validity and how it is defined within a nomological framework. This is, however, contrary to the views espoused by Borsboom, Mellenbergh and Van Heerden (2004) for instance. The latter authors prefer to plumb epistemological and ontological depths when it comes to this issue, whereas Embretson shows a preference for immediate solutions in a practical world. Both sides have much meaningful insight

- ⇒ development and deployment of robust psychometric theory (further developments in measurement foundations)
- ⇒ the utilisation of both CTT and IRT indicators of validity and reliability (in so far as this can possibly be attained)¹⁰⁷
- ⇒ further development and deployment of multidimensional IRT models which are able to account for changes in underlying ability (the perennial change-based approaches which are increasingly becoming psychometrically justifiable)
- ⇒ the advancement in CAT which utilises information and practice from various disciplines such as artificial intelligence, programming languages and neurology. This echoes the stance taken in chapter 2 in which reductionism as approach was applauded in its efforts to understand various sub-systems to a greater extent. Once each individual system is more fully understood, the closer the goal of unification appears in terms of bringing much of this knowledge together in a framework which will serve to elucidate the underlying processes within intelligence assessment. The discipline of dynamic assessment especially should take heed of these novel approaches

- *Wilson's SALTUS model*

Wilson (1989) developed his developmental Saltus model (from the Latin "leap" and by implication "leaping" from one stage to the next) which is also an extension of the Rasch model (Embretson & McCollam, 2004). Utilising various tasks at each developmental level, state changes are measurable but Wilson (1989) assumes that development proceeds in a stage-like manner and that it is not continuous (Wilson & Draney, 1997). This is contrary to available evidence which suggests cognitive development is continuous or at the very least undecided with much research involving mathematical theories of complexity (complex systems theories), chaos and catastrophe¹⁰⁸ as well as sophisticated computational models (Brainerd, 1993; Feldman & Fowler, 1997; Fischer & Rose, 1998; Gelman, 2000; Halford & Wilson, 1980; Johnson & Gilmore, 1996; Mareschal & Shultz, 1997; McCollum, 1997; Molenaar & Raijmakers, 2000; Preece & Read, 1995; Schulze, 2005; Shayer, 2000; Suizzo, 2000; Van der Maas & Molenaar, 1992). Modern theory and practice have called into question some of Piaget's results¹⁰⁹ (Gamlin, Elmpak, Franke-Wagar, Lotfabadi, Nourani & Wang, 1996; Nicolopoulou, 1993). That critical periods of development exist is not in question (Carlson, 2002). Non-linear models of development including catastrophe theory model the development process as one of continuous, stable yet abrupt change and is characterised by attractors (stable aspects of development) which can change radically and unpredictably (Sternberg & Grigorenko, 2002). It is vital to understand the cognitive theory behind the use of various models as they clearly effect interpretation and understanding of the underlying rationale. Was the model built to mould the theory or has theory been tweaked to mould to the model? The assumption here is that if researchers or practitioners do not agree with the stage-like development as advocated by Piaget for instance, then perhaps this model is not the one to pursue within a dynamic assessment set-up positing change. How exactly does the change occur? This is one example of a question that needs to be addressed before models are used. These changes are defined on two types of discontinuities; first and second-order changes. First-order changes are sudden changes which occur within a single ability and second-order changes are those manifesting across at least two domains. The Rasch one parameter model is utilised for the first-order changes and Saltus model is utilised for second-order changes. These first and second-order changes are the result of Fischer, Pipp and Bullock's (1984) (in Wilson, 1989) work on discontinuities within cognitive development; an instance, as stated above, where model was derived from theory. Wilson's (1989) Saltus model is a refinement of Guttman's scalogram model (as can be seen in Table 69 above) in which one item per level exists (Embretson & McCollam, 2004).

¹⁰⁷ Recall the thesis running throughout: no amount of technical application, development and sophistication can alter the one remaining variable. That variable is the origin of test items which are devised by people for people by hand on an intuitive basis and there is nothing scientific about this in the strictest sense of the practice of science. As argued throughout this thesis though, a science can be practiced within psychology without necessarily having to adhere to natural science tenets.

¹⁰⁸ Which does tend to make one sceptical of over simplistic modelling of change!

¹⁰⁹ It is at times unfair to be overly judgmental towards researchers in years gone by especially when their original contributions were so ground-breaking at the time. Piaget was situated with a context of his own and should be judged accordingly and perhaps not so ruthlessly discredited (Campbell, 2000; Kuhn, 2000; Perre-Clermont, 2000). Piaget's constructivist account of development is relevant in approach even today (Karmiloff-Smith, 1994) although much of his theory has been criticised (Anderson, 1992; Bates & Elman, 1994). His framework can serve as point of departure for comparative behavioural analysis regarding the development and trajectory of primate cognitive development (Doré, 1991). This highlights the similar evolutionary background shared among primates but this is a discussion for another time and place. But as an aside, it is remarkable to see the differences in animals who have led lives enclosed in poor environments and who have received subsequent exposure to better environs; is this not a manner of allowing the animal to negotiate in a zone of yet-to-be development?

Due to problems within Guttman scaling, Wilson (1989) utilised Rasch's probabilistic approach thus circumventing the adherence problem which manifests when persons not adhering to the scalogram ordinal model are discarded. In addition to the usual parameters (person ability and item difficulty) Saltus includes parameters influenced by cognitive theory which account for stage influences on item types. Wilson (1989) investigated data sets according to three cognitive developmental models in an attempt to apply his model and determine the fit with each set according to each theory. Wilson's (1989) reiterative concern echoes Guttman's similar concern; scale analysis does not define content, only substantive theory can! Recall this study's preoccupation with the need to re-look philosophical issues within psychological concerns. Refining and tuning tools of measurement will never solve a substantive problem, no matter how sophisticated the instrumentation is. Utilising modern IRT multidimensional change models can only aid in better fit to data¹¹⁰ and theory provided the theory is, via quantification, amenable to such measurement in the first place. Wilson's (1989) structure for Saltus is governed by a logistic function as in the Rasch one parameter model but the ability / item parameter ($\theta_i - \beta_j$) consists of additive elements for person, ability, item difficulty and level and is expressed as such:

$$\theta_i - \beta_j + \sum_h \phi_{ih} \tau_{hk},$$

$$i = 1, \dots, N, j = 1, \dots, L.$$

k = item type is a known function of j with at most K types; τ_{hk} is the level parameter for the subject group h and item type k . ϕ_{ih} is the selection factor which indicates group membership with $\phi_{ih} = 1$ if person i is in group h and 0 if not. Depending on the number of person groups, h , $\phi_i = (\phi_{i1}, \dots, \phi_{ih})$ must be estimated from the data. The person group $h(i) = h$ and item type $k(j) = k$, the probability of response y_{ij} is as follows:

where $\Psi(\sigma_{ij})$ is the logistic function

$$p(y_{ij} | \theta_i, \beta_j, \tau_{hk}, \phi_{ih}) = [\Psi(\sigma_{ij})]^{y_{ij}} [1 - \Psi(\sigma_{ij})]^{1 - y_{ij}}$$

$$\begin{aligned} \Psi(\sigma_{ij}) &= \frac{E(\sigma_{ij})}{1 + E(\sigma_{ij})} \\ &= \frac{E(\theta_i + \beta_j + \sum_h \phi_{ih} \tau_{hk})}{1 + E(\theta_i + \beta_j + \sum_h \phi_{ih} \tau_{hk})} \end{aligned}$$

Each person from each group, it is assumed, will apply strategies pertinent to that group across all items and under the assumption of local independence the conditional probability of a response patten $y_i = (y_{i1}, \dots, y_{iL})$ is as follows: (see section on Bayes conditional probability above) Wilson (1989) emphasises again the importance of substantive theory because the function $h(j)$ is known *a priori*. Group membership is determined from the data based on observable responses but the ability that led to that observed function was not known):

$$p(y_i | \theta_i, \beta, \tau, \phi_i) = \prod_j [\Psi(\sigma_{ij})]^{y_{ij}} [1 - \Psi(\sigma_{ij})]^{1 - y_{ij}}$$

- *Other models*

- i. Multidimensional random coefficients multinomial logit model

Other IRT models which encompass change in their structure have been developed, and include among others, the multidimensional random coefficients multinomial logit model which contains Embretson's MRMLC structure in addition to many other models developed since Rasch's 1960 model (Adams, Wilson & Wang, 1997). This model is an extension of the original unidimensional model developed by Adams and Wilson (1996) (in Adams et al., 1997). The unidimensional random coefficients multinomial logit model (RCMLM) as precursor to the multidimensional random coefficients multinomial logit model (MRCMLM), allows for flexible custom-designed testing situations and extends other generalised Rasch models (Wilson & Wang, 1995) which impose linear structures on item parameters, similar to Embretson's MRMLC's item constraint. It utilises marginal

¹¹⁰ The author is aware of the reigning controversy surrounding the debate of fitting the data to the model (Rasch) and fitting the model to the data (one, two and three parameter IRT) (Andrich, 2004).

maximum likelihood estimates for parameter estimation (Wilson & Wang, 1995) which aids in dealing with inconsistent estimates provided that the population distribution is specified. If not, conditional maximum likelihood estimation is utilised (Adams, Wilson & Wang, 1997). Joint maximum likelihood estimates become inconsistent as the sample sizes increase thus the utilisation of marginal likelihood estimates (Wilson & Wang, 1995). It allows for different numbers of categories in different items, differences in rater estimates and avails of multilevel structures (Wilson & Wang, 1995). Given the unique nature of dynamic assessment and the underlying ability changes, it is hypothesised that the Adams et al., 1997 model can perhaps be utilised for certain dynamic assessment situations. Specifically, the immediate attraction here is that the RCMLM incorporates a scoring function which enables a flexible relationship between qualitative aspects of the performance on an item and the quantitative level of performance the response reflects; i.e. scores above 0 and 1 can be allocated. The following exposition is taken from Adams, Wilson and Wang (1997). The RCMLM model can be written as follows:

$$P(X_{jk} = 1; A, b, \xi | \theta) = \frac{e^{(b_{jk}\theta + a'_{jk}\xi)}}{\sum_{k=1}^{k_j} e^{(b_{jk}\theta + a'_{jk}\xi)}}$$

And the response vector model as follows:

$$P(X = x | \theta) = \Psi(\theta, \xi) e^{[x'(b\theta + A\xi)]}$$

With

$$\Psi(\theta, \xi) = \left\{ \sum_{z \in \Omega} e^{[z'(b\theta + A\xi)]} \right\}^{-1}$$

Each item, i , consists of k_i+1 response alternatives ($k=0, 1, \dots, k_j$). X_j is the vector valued random variable where X_{jk} is 1 if the response to item i is category k and if not, then 0. X' is the response pattern realised via the lower case x and x' . ξ are item responses of p parameters which as mentioned above are linearly constrained. The linear combinations result in design vectors a_{jk} ($i = 1, \dots, n$ and $k = 1, \dots, k_j$) each with length p . These are collected into a design matrix A . The scoring function mentioned above, is represented by a response score b_{jk} and is collected as a vector b' . In the simple logistic model 0 and 1 are utilised as labels to indicate both the response category and the scores; which also includes figures above 1 in developed models. θ is of course the latent variable.

The extension of this RCMLM model is the MRCMLM¹¹¹ which takes into consideration a number of traits, D , as underlying performance, hence its multidimensional status. The D -dimensional latent space is once again represented by a vector $\theta = (\theta_1, \theta_2, \dots, \theta_D)$ (Wang & Wilson, 2005) and represents a random sample from a population. The scoring function is now a vector and not a scalar as in the RCMLM because the response category, k , in item i , corresponds to a $D \times 1$ column. A response in category k , on dimension d ($d=1, \dots, D$) of item i is now represented as b_{jk_d} . These scores, in turn, are collected into a vector b_{jk} and are collected into an item scoring submatrix, B_i and lastly collected into a scoring matrix B . ξ and A are defined as they were for the RCMLM model where they define the RCMLM for the item (Wilson & Wang, 1995). The probability of a response in category k of item i is as follows for the MRCMLM:

$$P(X_{jk} = 1; A, b, \xi | \theta) = \frac{e^{(b_{jk}\theta + a'_{jk}\xi)}}{\sum_{k=1}^{k_j} e^{(b_{jk}\theta + a'_{jk}\xi)}}$$

And the response vector model as follows

$$f(x; \xi | \theta) = \Psi(\theta, \xi) e^{[x'(B\theta + A\xi)]}$$

With

$$\Psi(\theta, \xi) = \left\{ \sum_{z \in \Omega} e^{[z'(B\theta + A\xi)]} \right\}^{-1}$$

The difference between the RCMLM model above and the one to the left (the MRCMLM), is that θ is a scalar in the former but a $D \times 1$ column vector in the latter. Also, in the former model, the response k to item i is a scalar b_{jk} and in the latter model this is a vector b_{jk} . $D = 1 + \exp(\theta_1 + \xi_1) + \exp(\theta_1 + \theta_2 + \xi_1 + \xi_2) + (\theta_1 + \theta_2 + \theta_3 + \xi_1 + \xi_2 + \xi_3)$.

¹¹¹ Wang and Wilson (2005) state that MRCMLM runs in ConQuest and is implemented with the marginal maximum likelihood estimation. The SAS NLMIXED procedure can also be used for fitting linear and non-linear models of the MRCMLM (Wang & Wilson, 2005). In addition it also utilises an empirical Bayes method (i.e. multivariate normal distribution is assumed and the vector-covariance matrix is empirically estimated (see the discussion on Bayesian statistics as under-utilised statistic within the social sciences above).

ii. Unified model for assessment

The unified model for assessment, although not developed with the change construct in mind is, according to Embretson (2000), relevant to the identification of the underlying attributes which have been successfully mastered (DiBello, Stout & Roussos, 1995). The unified model captures substantive cognitive theory as well as psychometric item response theory in one model. By utilising substantive theory the model is able to relate information pertinent to practical contexts in which specific information pertaining to cognitive areas is needed. Incorporation of educational assessment via cognitive theories into a model which is able to amalgamate both psychological research and robust defensible psychometrics is starting to come of age in models such as proposed by DeBello et al., (1995). Regarding conventional latent trait models:

- ⇒ Very few continuous latent traits are evidenced as underlying responses
- ⇒ A model is said to be multidimensional only when item responses can be shown to be conditionally independent when equal ability individuals are sampled
- ⇒ This multidimensional trait then accounts for the covariation between items
- ⇒ Such an approach, although successful with broad-based ability traits, are not able to render anything useful in terms of cognitive specificities which is needed in practice during remediation and so forth

Newer models:

- ⇒ Accommodate discrete cognitive skills leading to latent class analysis
- ⇒ When the probability distribution of classes (initially not known) is known as well as conditional item response probabilities the latent structure will be known
- ⇒ The research in this area is still being conducted and as a result, if too many classes are utilised, the model becomes unfeasible. Likewise, too few classes results in the resemblance of this model to the multidimensional models
- ⇒ The model can account for where the individual is placed within a cognitive space and can determine the accuracy of the assessment - once again it can be seen how substantive theory and psychometric modelling can be melded
- ⇒ A fine balancing act is necessary if researchers are to extract the most pertinent cognitive information from test responses in such a manner as to allow for a greater number of parameters, without the size increasing to such an extent that the model becomes unwieldy. In order to aid in making the model finer grained, certain computational mechanisms are needed, such as neural networks and Bayesian inference networks
- ⇒ The trade-off between fewer parameters versus less accuracy is the model's proficiency at detecting small changes in responses
- ⇒ The model incorporates various sources of stochastic response variation (as yet unaccounted for in uni- and multidimensional trait models, factor analytical models) such as:
 - Differential strategy usage - i.e. the model may predict certain strategies and not others, but if the item is endorsed utilising unaccounted for strategies this needs to be factored in
 - Completeness - which reflects the as yet to be completed list of skills, procedures or other attributes which are used by individuals but are not listed in the model
 - Positivity - where for some reason the correct strategy is not used even though the ability is present, or the requisite ability is not present but the correct strategy was used
 - Slips - random errors made by individuals for any number of reasons
- ⇒ The model is discussed in detail in DiBello et al., (1995) from which the above has been summarised. The research presented by these authors, although now over ten years old, is indicative of where future latent trait models are heading and is currently cited by eminent researchers in the field (Embretson, 2005). At the time of the 1995 publication, though, there were still various issues which needed attention in the model such as model calibration and the need to incorporate more cognitive operation necessary for diagnostic assessment. The research was still at simulation phase

iii. Linear Partial Credit Model

Ponocny and Ponocny-Seliger (1997) have developed a programme which manages various models within the area of change; L_{PC}M (Linear Partial Credit Model). It fits models for two or more time points and can do so for unidimensional, multidimensional and mixed dimensional items sets with both dichotomous and polytomous data in an attempt to measure change. The programme is written in C and initially ran in DOS, but was, during 1997, being developed for a Windows based environment.

Dénouement

The culmination of mathematics, statistics and measurement and what these seemingly disparate areas of concern mean for test theory cannot be more clearly seen than in Wright's (1999) table which he has entitled "An anatomy of inference" which

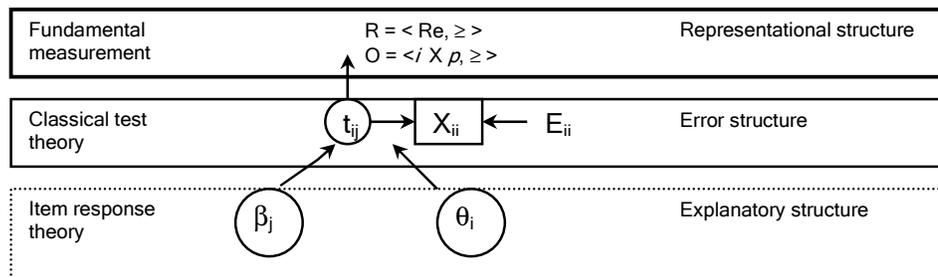


follows below in table 19. Test theory is an evolving discipline with a large literature attesting to various models' strengths. It is worth citing directly from Borsboom (2005) regarding the three main theories concerned with the measurement of the mental: "classical test theory is basically about the test scores themselves, representationalism is about the conditions that should hold among test and person characterises in order to admit a representation in the number system, and latent variable theory is about the question [as to] where the test scores come from" (p.121). This statement can be visually reflected in the following figure, 70.

Table 19 An anatomy of inference (from Wright, 1997, 1999)

An anatomy of inference		
Obstacles of <i>raw data</i> inference	Solutions to <i>inference</i> from raw data	Inventors
<i>Uncertainty (do not know)</i>	<i>Probability (know at least something)</i>	Bernoulli, 1713 (section 4.2)
have => want	binomial odds	Bayes, 1764 (section 4.3.1.1)
now => later	regular irregularity	Laplace, 1774 (section 4.2)
statistic => parameter	misfit detection	Poisson, 1837 (section 4.3.1)
<i>Distortion (from observation to conceptualisation)</i>	<i>Additivity (making visual sense of observation data by making it linear in our two and three dimensional space)</i>	Fechner, 1860 (section 4.4.1.1)
nonlinearity	linearity	Helmholtz, 1887 (section 4.4.1.1)
unequal intervals	concatenation	Campbell, 1920 (section 4.4.1.1)
incommensurability	conjoint additivity	Luce / Tukey, 1964 (section 4.4.1.1)
<i>Confusion (the need to separate and study one dimension at a time)</i>	<i>Separability (causes as separate parameters make for our quantification; i.e. item or person parameter)</i>	Rasch, 1958 (section 4.4.1.3 in which Rasch (1980) is used)
interdependence	sufficiency	Fisher, 1920 (section 4.3)
interaction	invariance	Thurstone, 1926 (sections 2.7.1.1.; 4.4.1.3 and 2.8.2)
confounding	conjoint order	Guttman, 1944 (section 4.4.2)
<i>Ambiguity (based on our still arbitrary decisions as to what underlies behaviour)</i>	<i>Divisibility (there are known solutions to dealing with such ambiguity)</i>	Levy, 1937 (section 4.4.1.3)
of entity, interval	independence	Kolmogorov, 1950 (section 4.2)
and aggregation	stability	Bookstein, 1992

Figure 70 Relations between fundamental measurement, CTT and IRT (Borsboom, 2005, p. 129)



Fundamental (representational) measurement:

- ⇒ the true score can be instrumentally compared \geq to the product of person ability and item difficulty ($i X p$)
- ⇒ the true score is numerical and can so be ordered
- ⇒ the effects of latent ability and item difficulty are independent, items and subjects can be ordered instrumentally
- ⇒ this results in $O = \langle i X p, \geq \rangle$
- ⇒ if the model is true and because ability and item difficulty are independent, true scores can be transformed in an additive manner befitting Rasch model representation and
- ⇒ this results in the numerical relational system $R = \langle Re, \geq \rangle$ but “the representationalist theory becomes a liability, for it prisms apart the numerical concepts used in science from the empirical reality studied” (Michell, 1990, p.49)
- ⇒ together, $O = \langle i X p, \geq \rangle$ and $R = \langle Re, \geq \rangle$ form an additive conjoint measurement structure

Classical test theory:

- ⇒ is a theory of error structure
- ⇒ the true score is defined as a propensity distribution of expected values for individuals, i , and items j

Item response theory:

- ⇒ hypothesises the data generating process
- ⇒ variation on the latent trait produces variation on the true score
- ⇒ item difficulty produces such variation as well
- ⇒ it thus claims more than the fundamental representation, which itself cannot add anything other than a representation of the relationship
- ⇒ as mentioned, the IRT model is falsifiable on account of its not fitting the data. Fundamental measurement cannot avail itself of this even as it only functions as a construction of a representation and nothing more

4.4.3 Summary

“To measure” - this short clause is packed with underlying assumptions which are sometimes made explicit but are most often implicitly implied. This tacit agreement among psychologists as to what constitutes measurement, borders at times, on ignorance of these said assumptions. Measurement precludes quantifiable structures which, as a concept, necessitates an additive structure, units of concatenation, provable representation, fundamental qualities or extensive units or in the case of less identifiable areas of concern, conjoint measurement based on two or more intensive measures via the process of double cancellation as well as via solvability and Archimedean axioms. Do psychologists routinely consider the above? This question is often asked within texts detailing the shortcomings of a quantifiable psychology and social science. Measurement within psychology is intimately bound to its mathematical foundation, for it is upon a system of axioms which has allowed for its maturation into the discipline it is today. Also intimately bound to this measurement system are the statistical techniques employed to render inferential information to users. It is, after all, the users who are supposedly gaining from such a system of measurement. This thesis though is highly questionable as it cannot at this stage even be proven that the constructs being measured exist as defined according to the techniques used to measure them. Unfortunately it seems that constructs are defined by the measures used which is tautologous, unfair and is simply an egregious error in logical thinking.

Psychological measurement within the intelligence arena is often predicated upon a general construct seemingly prevalent in most sub-tests assessing for some sort of intelligence. There seem to be as many definitions of intelligence as there are tests yet the one pervasive finding is the statistical manifestation of a general factor. This is hardly surprising if most test batteries



assess on very similar items. Spearman's general factor and his development of a theory of mental testing attests to his insight and work but one must be mindful that there are possibly innumerable ways of assessing for constructs which do not necessarily fit the picture of a general factor underlying what we think is intelligence. Much work in this regard has been undertaken and has been gaining momentum since the early years of the twentieth century often concurrently to classical notions of test theories. The main argument within dynamic assessment, as this thesis contends is that

- Intelligence is difficult if not impossible to define
- The construct is assumed to exist in some fashion or other
- It is likewise assumed to be stable

Dynamic assessment not only has to contend with the above-mentioned three aspects but has the following added to its problems

- learning potential is difficult to define especially if one is to continuously define it according to mainstream ways of identifying it
- learning potential will need to extricate itself from the mainstream terminology otherwise it will never define for itself its own position - one can only proceed to step two once step one is complete; but step one is not even there
- learning potential assess for change which is entirely at odds with the notion of stable development and progress

The solutions to this dilemma are numerous but the following can perhaps be put forward as tentative attempts

- dynamic assessment should leave the realm of intelligence assessment as currently understood in mainstream testing
- intelligence assessment as understood in the mainstream realm is perfectly suited to further advancements on the physiological front as well as on the psychometric front
- the research efforts as evidenced by numerous erudite intelligence researchers cannot be ignored and in fact should be encouraged, however,
- dynamic assessment has its own unique philosophy and thus should contain itself within a similarly different philosophical and practical areas and hopefully in this way
- it will be able to more accurately define for itself what is meant by learning potential. The definition may well undergo radical transformation if defined within a completely different context

Measurement utilises elements as its population of manipulation. Such elements need to undergo transformations if they are to be accurately represented within a system of measurement. In order to effect this, correspondence rules need to be addressed so as to preserve the original structure of the element. Such transformations are governed by various techniques applicable to certain types of elements which is why due consideration for psychological measures is so imperative. Psychological measurement was thought about a great deal during the first four decades of the twentieth century and it is perhaps a pity that similar thought had not occurred on such a scale since for it has become routine to accept what is thought to be an accurate understanding of what it means to measure. Stevens, who is often viewed as the person starting a measurement revolution for psychology in the 1940's gave to psychological measurement, scales, which could handle various types of data. Unfortunately Stevens' ideas were not the only ones promoted but they were marketed well and were possibly understood the best. This situation is easy to understand once thought is turned towards the prevailing scientific spirit of the times, which has been discussed at length in chapter 3. Unfortunately, assigning numerals to objects is completely ill-befitting to the serious enterprise of psychology and its construct assessment. There is more to measurement than scales. Far more.

Primitive or extensive measures are quite easily come-by in the natural sciences where concatenation, representation and subsequent measurement is possible, at least more so than in the social sciences. Results in the natural sciences look robust and generally appealing to scientists who are able to measure even objects which are not observable, albeit in an indirect manner however psychology cannot even say it can measure observables let alone unobservables. Thus, in an early attempt to appease discordant opinion, measures were assumed to be possible and subsequent measures on many fronts were undertaken. Most notable are those measurements stemming from psychophysics which even in its name is redolent of a natural science endeavour. By utilising carefully worked out statistical science along with measurement theories, a robust veneer was granted to an essentially soft discipline. The unnerving aspect about this is that a perfectly scientifically valid qualitative science can evolve within psychology without its having to subscribe to natural science dictates - but it seems that the research world and the public at large still do not recognise this fact. The consequences play forth in areas such as dynamic assessment which seek to encourage and assess change in a manner quite foreign to mainstream assessment.



Physics is not psychology and psychology is not physics, so why then do we insist on utilising physics methods in psychology? It can be assured that physics does not utilise psychological methods in its studies.¹¹² Nevertheless in psychology's continual pursuit as robust science it devised for itself a basis of verifiable and justifiable measurement in the sense understood within the natural sciences. It accomplished this by representing constructs in a manner which allowed properties to become jointly assessed and hence amenable to measurement in the extensive sense of measurement. Intensive measures were now cased in extensive frameworks. Conjoint measurement, although still questioned as to its facility to really assign objects to measurement, has become an accepted part of modern test theories. Axioms of measurement are carefully considered within their mathematical systems and have proofs derived for them in order to validate their postulates. Much recent work in this field has been conducted, among recognised others, by Joel Michell, a researcher in Australia and Paul Barrett in New Zealand. Criticisms and thoughtful concerns regarding classical as well as modern theories of measurement placing special emphasis on construct validation has been conducted by, among recognised others, Denni Borsboom in the Netherlands. Their geographic locations may not, at first glance, seem important but when one considers the origins of dynamic assessment as emanating principally from Russia (Soviet Union), Israel, the Netherlands and Germany interesting questions start to be asked. Mainstream testing has had a bountiful history within Britain and America (although predicated on different philosophies throughout). These facts only serve to raise tentative questions at this stage. Why have alternative mechanisms of assessment principally arisen in countries outside the areas of mainstream assessment such as the United States and Britain for instance? South Africa, at present, is pulled both ways but seems to have sided more with the mainstream manner of assessment and basic underlying philosophy. Likewise notable criticism of theories of mental testing has emanated from countries outside the United States and Britain. Is this due to local minority and/or immigrant populations in countries outside the United States and Britain? These two countries have had to contend with an influx of immigrants especially since the Second World War whereas other countries have consistently existed with local minority groups. This brings to the fore the issue of cultural relevance and the use of tests to keep people out as opposed to letting them in. A contentious issue at the very least but a plausible reason as to why criticisms are lodged at mainstream assessment which emanate from countries outside mainstream concerns.

Newer developments within mental test theory are evidenced in the works of modern item response theory, a method whose time has now truly arrived. Due to the lack of computational power in the 1940-1970's many techniques were simply too complicated and power-hungry in order for them to be utilised, especially by mathematically unskilled psychologists.¹¹³ Classical test theory is a powerful theory of reliability and should not be neglected in the future but rather used concurrently with other newer techniques and test theories. Generalizability theory which is considered as a branch of true score theory allows for more flexibility in terms of relaxing certain assumptions within its predecessor. Modern test theory offers solutions to various measurement issues that have confronted psychology but one troubling aspect still remains, which is unlikely to be solved by any measurement technique and that is the burning issue of construct validation. For this, substantive theory is needed which in psychology's present-day status, needs to be revisited in terms of its methodology and philosophy. Modern test theory has partially aided dynamic assessment's quest to define for itself a measure of changeability, which in the true score model is almost impossible. These change score models were discussed in some detail so as to illustrate their encompassment of change within the measurement model which represent the new rules of measurement. This was done for a reason. By seeing which representations are being utilised for change assessment, the model can be questioned directly. There are numerous Rasch related IRT models available which avail of themselves mechanisms to assess change and only five were discussed. Clearly, if dynamic assessment is to stay within mainstream assessment, its best bet at this stage would be to follow up on and make use of such change-based IRT models whilst utilising useful information and insights from classical test theories and older models of reliability. If dynamic assessment is to forge ahead on its own unique trajectory then it should avail of quantifiable non-numeric measurements. Perhaps hybrid dynamic assessment models can coexist in both realms. The debate, however, has only just begun.

4.5 Integrating Madsen - attenuating the framework

In attempting to attenuate Madsen's quantifiable systematology it cannot be ignored that criticisms have been levelled at his technique and approach towards documenting a method of theory judgement (Brandt, 1984; Ettin, 1984; Schaeffer, 1984). What is central to this study is Madsen's technique and not necessarily his actual findings in his own area of interest and research. In other words problematic issues within Madsen's work such as reliability including selection, inclusion and rendition criteria (Ettin, 1984) are simply not agreed upon by critics and are not explicated to any great extent by Madsen himself. Hence the need to attenuate and support the need for utilising his approach; not to mention the rather obsolete usage of S-R terminology¹¹⁴ which is quite dated as the stimulus-response model of human behaviour is not only crude but incorrect due mainly to its simplistic

¹¹² The author is aware that there is a continuous flow between the disciplines in terms of importing ideas but there is not much crossover in terms of methodology; physicists do not infer unknowable causes from effects even in areas such as quantum physics!

¹¹³ Perhaps this is not always a fair argument. After all, psychologists are not mathematicians nor are they statisticians. However, this does not mean that utilisation of techniques without concern for how they were derived should proceed without some measure of understanding of the underlying mechanisms. To state that one is not informed about certain mathematical formulations is to take the easy way out.

¹¹⁴ The notation will be used throughout chapter 5 in keeping with Madsen's convention of scoring the HQ.

stance on human functioning. The author is of the opinion that Madsen offered his systematology of theories of motivation as one researcher's contribution to the field and did not, in contrast to what Ettin (1984) implies, expect that such an analysis be taken as a given in future readings of the text. This is open to speculation, yet there is sense in the criticism lodged at the lack of reliability and validity scaling for his technique and that it in fact does not, in its current form, warrant the status of empiricism (Ettin, 1984). Madsen's conception of a meta-theory has been classed as more of an heuristic aid than empirical certainty (Brandy, 1984; Ettin, 1984) and due to his approach being described as descriptively sound, cogent and parsimonious yet in need of revision in some quarters (his HQ system) the study utilises his approach.

In attenuating and building a framework for dynamic assessment and intelligence, all that has been discussed thus far in chapters 1-4 become enmeshed into the framework in such a way, as will be seen, to allow for greater flexibility within Madsen's original scheme in order to accommodate this study. Madsen, for instance, did not take cognisance of the prime considerations within psychology, namely the mathematical, statistical and measurement foundations permeating assessment; thus this will be worked into the framework alongside his original elements. The first consideration will be the culture and society in and from which various models and theories have developed. As already hinted at, different viewpoints have emanated from different countries possibly due to influences not always directly perceptible. The scientific community from which the individual researchers practise will also be highlighted. Within any model or theory, the nature of the science practised will be looked at and will include the nature of the empirical research, the theoretical thinking behind it as well as the underlying philosophy. Meta-levels of concern also focus on these three levels but extend their range somewhat to include philosophies of science, psychology and their respective histories. Recall that theories become data within the meta-theoretical framework and are treated at the empirical level. Theories are a culmination of deep-seated philosophies and hypotheses, conjectures as well as data whereas models function as heuristics which in turn are either theoretical, empirical or a mix of both. Theoretical heuristics serve to constrain whilst empirical serve to detail and locate. Such models are aligned according to their degree of abstraction: material (descriptive), graphic (explanatory), simulation (ontological) or mathematical (symbolic). These dimensions will serve to support the model framework. The data stratum is the level at which measurement philosophy enters Madsen's framework. As this is a vital component within the thesis in attempting to direct dynamic assessment's measurement schemes, it will need to be brought in at this stage.

Prime considerations as discussed in chapter four will thus form part of the data level analysis (even though it can be easily accommodated in the epistemological realm) and will concentrate on the statistical and measurement issues more so than on the mathematical ones. Hence the framework will proceed as follows: consideration of meta-level issues such as ontology, philosophy; hypothetical-level issues such as hypothetical terms, scientific hypotheses and the hypotheses system and data-level issues such as abstract data and concrete data issues. An attempt will be made to utilise Madsen's hypothesis quotient (HQ) and depending on the ease of application, the results should illuminate, at least at a very basic level, the testability of various models and theories. Visually the framework as illustrated in figure 71 will utilise the notation so as to ensure clarity. Thus when the models/theories are assessed according to various dimensions, notation such as "A(ii)" and "B(i)" and so on will be used. The hypothesis quotient will be worked out according to the model/theory's amenability to such quantified rendering. One very positive aspect or feature about this process, is that the same framework is utilised throughout thus yielding a standardised rendering of the models. Pitched at the same level, so to speak, there is at least some minimal basis according to which comparative assessment and judgments can be made. Thus, although not everyone may agree to the synopsis provided by working through the framework, one can at least use a common metric to assess these different models and theories and conclusions can be reached.

Figure 71 The attenuated Madsenian systematology framework

A: Meta-stratum (chapter 2 as necessary prelude)
A (i) Ontology:

- How does the theory/model comment on or is influenced by the conception of the human being, the mind/brain thesis and human freedom?

A (ii) Philosophy:

- How does the theory/model account for its epistemological modes of conception in terms of cognition and its relation to reality?
- How does the theory/model comment on or is influenced by meta-theoretical concerns such as nomothetic, hermeneutic and idiographic ideals?
- How does the choice of the theory/model's methodology impact on the outcome; what test methods and data language is utilised?

B: Hypothetical-stratum (predominant attenuation occurs within this level)
B (i) Hypothetical terms:

- How does the theory/model account for its ontological reference, existential form and function of hypothetical terms? In each of these classifications are concerns which will be noted, however should they prove unnecessary or superfluous within this thesis' context such concerns will not be dealt with, hence an attenuation of Madsen's framework.

B (ii) Scientific hypotheses:

- How does the theory/model account for its ontological classification of scientific (testable) hypotheses as well as its meta-theoretical classification of scientific hypotheses? In each of these classifications are concerns which will be noted, however should they prove unnecessary or superfluous within this thesis' context such concerns will not be dealt with, hence an attenuation of Madsen's framework.

B (iii) Hypothesis system:

- How does the theory/model account for its deductive explanatory system as well as its model explanations and what is the degree of abstraction involved? In each of these classifications are concerns which will be noted, however should they prove unnecessary or superfluous within this thesis' context such concerns will not be dealt with, hence an attenuation of Madsen's framework.

C: Data-stratum (inclusion of chapter 4)
C (i) Abstract data:

- How does the theory/model account for its functional relations between variables?

C (ii) Concrete data:

- How does the theory/model account for various forms of concrete data such as evidenced from behaviour?

C (iii) Prime considerations (chapter 4):

- How does the theory/model account for and defend its use of statistical techniques?
- How does the theory/model account for and defend its measurement models (mental test theories)?

4.6 Conclusion

Psychological assessment literature more often than not concerns itself with measurement issues as currently understood via mainstream assessment issues. Undergraduate and postgraduate assessment training very often comprises issues pertaining to classical test theory and the newer item response theories of measurement. Very rarely does the text delve into the mathematical foundations upon which these theories are predicated. Here reference is being made to philosophical mathematical issues and not only issues pertaining to test theory models per se. When psychologists purport to test, assess and in some or other manner measure constructs, they are leaning on a long history of mathematical dispute and disquiet of which many are blissfully unaware. The fact that mathematical tensions exist between what is knowable and what is not knowable hardly ever receives a mention in psychological assessment texts. Number crunching, statistical manipulations and the requisite details underlying measurement theory are often left to researchers within these respective fields. A continuous thread spans endeavours in mathematics, statistics and measurement where their culmination is manifested within assessment. Although the philosophical and historical foundations of mathematics may seem to be far removed from assessment upon first glance, it is vital that these issues are at least understood and looked at before tests are thoughtlessly applied in practice. Perhaps the most important fact for psychologists is that so-called firmly entrenched inviolable proofs are in fact predicated upon axioms which are themselves improvable. What does this mean for the practising psychologist? A great deal of angst and concern if one is to fully understand that cherished ideas of mathematical superiority and certainty are in fact not so superior nor so certain. "Axioms" as a term should perhaps be revised; perhaps we should consider them as temporary axiomatised systems in waiting. Probability as a concept has played a monumental role within social science statistics and measurement and has rightly pervaded (or invaded) various sub-disciplines to a large extent. Probability is part of mathematical thought and thus needs to be addressed when one seeks to understand what its function is within psychology. The probability of an event occurring, or that an experimental technique will work hinges on its mathematical predicates and history. The story of mathematics, statistics and measurement is a long one spanning millennia, so it can hardly be ignored. Having said this, the question is asked once again: why the need to consider mathematical foundations? It is hoped that this question has partially been answered.

The statistical foundations of the social sciences really is the work of giants on whose shoulders we gladly stand; however stagnation at such lofty heights may well prove to be our downfall. Pioneering statisticians devised schemes in manners that would leave most current statisticians and historians in disbelief. Computational power was not even a consideration during the heady days of Galton, Fisher, Neyman and Pearson (Karl and Egon) and constant reminders of the possible inadequacies of their insights has to be greatly tempered with their astounding achievements given their only source of computational power - their intellects. Hard work, perseverance, determination and considerable skill allowed such individuals, among others, to propel the nascent field of statistics in a direction which was to serve psychology for the next century. But all is not well. Consistent misrepresentation and misunderstandings of the rationale behind the use of many statistical techniques has partly resulted in an extended era of number crunching at the expense of substantive theorising. As stated in chapter 3, psychology (as with other disciplines) moves at the behest of policy, law-makers and society's needs and wants. Who are we to argue about grants which come our way if we can prove our techniques are psychometrically defensible and robust despite their logic being flawed? This decades-long stalemate will eventually lean in favour of one or the other direction. Pundits have lauded the entrance of thoughtful statistics and decried the continued use of outmoded and dated techniques within psychology specifically. Papers are cited as being peppered with p values, asterisked to such an extent that it becomes an almost gross violation of what it means to be significant. The crud factor has entered into a realm which can happily accommodate yet more data and more values less than 0.0001! How low can you go?¹¹⁵

The subsequent confounding of statistical significance with that of substantive significance as well as arbitrarily chosen labels at which to accept or reject findings within experiments which are themselves biased towards findings in one direction has made for experimentation within psychology an enterprise of anticipated number crunching. Where is the substantive psychology in all this? Dynamic assessment's process-based qualitative interest in the renewal of cognitive functioning, it seems, is ill-fitting to such a scheme. Can the complexities of cognitive maturation and skill development as observed within intervention processes really be contained within a framework and structure allowing for only certain levels of significance to be obtained within certain statistical tests? Brain development is not equivalent to the processes involved in determining the level at which certain metals undergo molecular decay and hence disposed of in a production line (where significance comes into its own and is especially applicable and robust; saving time and money). A re-look at the use of statistical practice among psychologists prompted numerous debates concerning the over-use of null hypothesis significance testing and under use of other manners of treating data (such as confidence intervals and use of Bayesian statistics for instance). However, this particular concern with statistics is perennial and has a history extending as far back as the 1960's. Dynamic assessment was cloaked in an already established framework of statistical assessment of data and as such, has hardly had much of a choice of techniques at its disposal

¹¹⁵ Given a large enough database, shoe size really can be shown to be significantly correlated to weather patterns.



especially if it was to become more generally accepted into mainstream assessment practise. Psychology and statistics within the social sciences developed almost in tandem and both fields informed and gained information from the other. There was no fully developed statistical framework according to which data could easily be analysed and custom treated depending on the context. Psychology had to develop in its own manner and borrowed techniques from statistical sources but also stimulated the development of novel and widely used techniques, some of which have spread to other disciplines involving the more natural science orientated.

Although useful, NHST testing necessitates a concerned revamp and a renewed look at Bayesian statistics is warranted, especially within the field of dynamic assessment which resolves to understand prior and post mediation performances. Instead of basing results on tests accounting for only information written in the test, a Bayesian approach can accommodate prior information on the individual which does not necessarily have to emanate from a test per se. This is fairer and more inclusive and in keeping with numerous aspects propounded by dynamic assessment's basic philosophy and serves as a reminder that there is more to inference than frequentist approaches. Results, when they are necessary, should attest to a flexible framework in which individual case studies are assessed on their merits and which do not yield to the strict and short-sighted "accept-reject" mode of practice within NHST. Studies should also be amenable to a falsificationist approach as opposed to a verificationist one. This larger reality, however, encompasses variables which cannot be ignored and in finding for itself another way of dealing with the statistical within dynamic assessment research, the broader scientific community needs to re-arm itself with a variety of tools which can be deployed. One such tool is the choice to allow non-significant studies into publications, as it is replication upon which a science should be predicated. Replication, it might be said, is taken into account within statistical modelling as in the case of the power of a test, which is cost and time saving; but when conducting substantive research or when research is at the level requiring more substantive work, then this spirit should not prevail. Dynamic assessment is at just such a level of substantive research.

Simply put, measurement entails numericising objects. Objects can be physical entities which can be directly observed or indirectly ascertained via means of skill inference from what can be observed or what is known. Indirect inference can proceed from objects known to exist but known only through mathematical abstraction. Many objects of the latter type have been predicted and later discovered yet remain unseen in the conventional sense. Psychological variables, however, cannot be said to have attained the status of existing object whether seen or not. Representation of objects does not proceed along a manner akin to merely assigning a number to such an object but includes rules and axioms of representation. Numericising such objects has posed, what some may consider, to be an insurmountable problem. Means exist through which such unobserved entities can manifest via numerical relations; this enterprise within psychology is called test theory and encompass older and newer forms of mental test theories, each with merits and disadvantages. Measurement is, however loosely, predicated upon a mathematical set of axioms, some of which have been delineated and proved. Decades of debate concerning the nature of mental measurement has periodically resulted in committees being constituted in an attempt to further understand the nature of measurement and its requisite tools and how best to go about the business of testing and measuring. The most salient question hovering above all this has been whether psychological entities, such as they exist, can at all be quantified.

Assuming a quantified structure, classical test theory progressed primarily from a theory supporting the notion of a general factor of intelligence and was preoccupied with the reliability of scores. True scores, in this model, are never directly knowable, rather, they are computed from error in addition to the observed score. Curbing or at least accounting for error as robustly as possible by factoring random and chance factors it is assumed that the true score is attained to some degree. Item response theory jointly computes individual ability along with the difficulty of item in a probabilistic model thus allowing for direct measurement of intensive measures, thus overcoming, at least partially, the deficits of numericising intensive measures that classical test theory could not adequately manage. In addition to this, change within testing is more readily accounted for in IRT which is now able to model change utilising various models of change-based multidimensional models, many of which are based on the original Rasch models. Such early models were being considered theoretically in the 1950's but due to lack of computational power were not amenable to practice testing. Luckily, there is evidence of a trend towards increasing use of IRT models, although some researchers still question the degree to which even these techniques (although superior in many ways) are able to account for a truly representational manner of measurement. The core question remains as to whether there exist any entities which can avail of quantification in the first place. This would of course necessitate a re-look at substantive psychology encompassing renewed interest in theoretical psychology, an issue discussed in chapter 3. This chapter concludes with the attenuated Madsenian framework which will be deployed in chapter 5.

A point has been reached in this thesis in which the developed framework can be utilised for purposes of exploring a meta-theoretical framework for dynamic assessment and intelligence. Chapter 1 briefly introduced the need for such a framework within the area of dynamic assessment research. Chapter 2, although in a measure testament to the author's particular view points, was a necessary prelude to chapter 3 which dealt with some concepts that were evidenced in Madsen's framework and in order to more fully appreciate some of Madsen's views certain dimensions received attention in chapter 2. Chapter 3, in keeping with the framework, discussed the importance of identifying psychology's formal existence as having sprung from a natural science orientation and that due to historical circumstances was framed and dictated to, to a certain degree by the



prevailing climate of the times. In addition to discussing the role of natural science explanatory mechanisms as well as those of social sciences, attention was turned towards psychological explanatory mechanisms. Chapter 3 continued with the detailing of a meta-theoretical framework as developed by the Danish psychologist, K.B Madsen. In pursuance of a goal of unifying framework, the development of a framework was not yet complete. The framework necessitated additions in terms of prime concerns running throughout psychological assessment. As mentioned Madsen's framework was originally developed with theories of motivation in mind and not psychological assessment per se. The need to include epistemological and ontological measurement foundations in the area of assessment is paramount when the future of dynamic assessment within intelligence is concerned. Thus chapter 4 needed to be incorporated into the existing framework. The emphasis within the framework is placed more so on the statistical and measurement foundations as it pertains to dynamic assessment and intelligence, however the mathematical foundation discussion was necessary in order to highlight various aspects within the statistical and measurement foundations. Chapter 5 applies the framework detailed thus far and attention is now turned toward this culminating chapter.

Chapter 5 Dynamic assessment and intelligence: exploring a meta-theoretical framework

5.1 Introduction

This introduction will take the form of a question-answer format. Adequately addressing the status of the field of dynamic assessment necessitates a return to origins; a re-look at the beginnings which spawned the development, growth and future trajectory of the field within the broader discipline of intelligence research within psychology, a social science often veering towards its natural science foundation (at least in the West). The author might well have considered the following as militating against the pursuance of this study and will embark on a brief diversion and play the role of devil's advocate. A few aspects will be analysed for the veracity of their claims:

1. Exploring dynamic assessment in its current form will suffice as there is simply no reason to look toward the past nor is a meta-theoretical framework warranted as this field is not yet developed to the point where such action need be taken. Moreover, why choose someone's work which is as dated as Madsen?
2. Questioning of psychology's historical background is without merit as one simply cannot rework history with the express purposes of suiting this study
3. Investigating domains further a field such as mathematics, measurement and statistics has no possible bearing on the past or future direction of a field as vast as dynamic assessment
4. Retracing obscure Soviet influences might be fulfilling as a past-time but can hardly offer much in the way of the future viability of dynamic assessment
5. Embarking on philosophical forays is all very well but cannot be said to always have a direct impact on practical matters anyway. Theory and practice do not always converse well
6. How can one possibly address the vast field of intelligence in such limited confines and then seek to somehow amalgamate dynamic assessment within it? Is this not arrogance?
7. Is there really a "problem" or "issue" at all to begin with? After all, is dynamic assessment (like any other area of concern) really in dire straits? Is one not perhaps making a mountain out of molehill?

To which cogent answers to a few of these aspects take the following form:

1. Unfortunately this is not the case. Re-evaluating past lines of progression within a field from as objective a stance as possible highlights what could possibly have been forgotten or simply not seen as important at the time. Hindsight is particularly acute and observations of how the field has progressed (or stagnated) could provide valuable clues to its future existence. Moreover, the author herself, although advocating the principles behind dynamic assessment cannot be said to be partial towards this manner of assessment in terms of many aspects underlying its rationale. Knowing the historical progression will indeed aid in the prevention of mistakes and hopefully we will not be doomed to repeat such errors. Regarding the second part of the claim; the necessity of a meta-theoretical framework, the field is at once old and young depending on who is read, what is stated and how it is argued and the point of view taken from the outset of such deliberations. It is particularly old in essence (principles) but is new in terms of modern legitimacy. A framework is sorely needed to bring together such varying strands. The framework chosen for this study's purposes not only provides some measure of containment but also posits guidelines in the analysis of models and theories within such a framework. Madsen came closest to what the author required in terms of just such a framework. Moreover, his framework was attenuated and added to in order for it to be even more tenable as constitutive guideline.
2. Often, the finer details pervading the historical record are either ignored or misinterpreted by future generations who simply acquire information from second-hand sources who themselves may be open to misinterpretation. This is not always the case of course, but nothing can replace the need to study primary texts. It is vastly illuminating to read original sources and to interpret anew and re-visit "common" sentiments expressed which were never "common" in the first place! New renderings of texts can bolster certain arguments for or against certain sentiments which may or may not be pervasive in a field of study. Perhaps the statement made above by the devil's advocate is too strong and it can be argued that not many people would in any event agree with the notion. A tamer re-working can be considered.
3. If there is one statement made by the devil's advocate that can be vehemently argued against, it would be this one. It is crucial if not imperative that the reasoning behind the required need to implement mathematical, statistical and measurement theories to the psychological and wider social domain be reconsidered. Merely continuing with a method of study or analysis simply because "this is the way it has always been done" does not make any scientific sense at all. The ongoing controversy surrounding NHST is one case in point. The author has not come across very much in the psychological literature by way of mathematical foundations being explored with explicit regard to psychological dynamic assessment issues. This is perhaps a rich area into which can be delved deeper. The role of measurement theory and statistical renderings of meaning are crucial areas to explore and new techniques can be

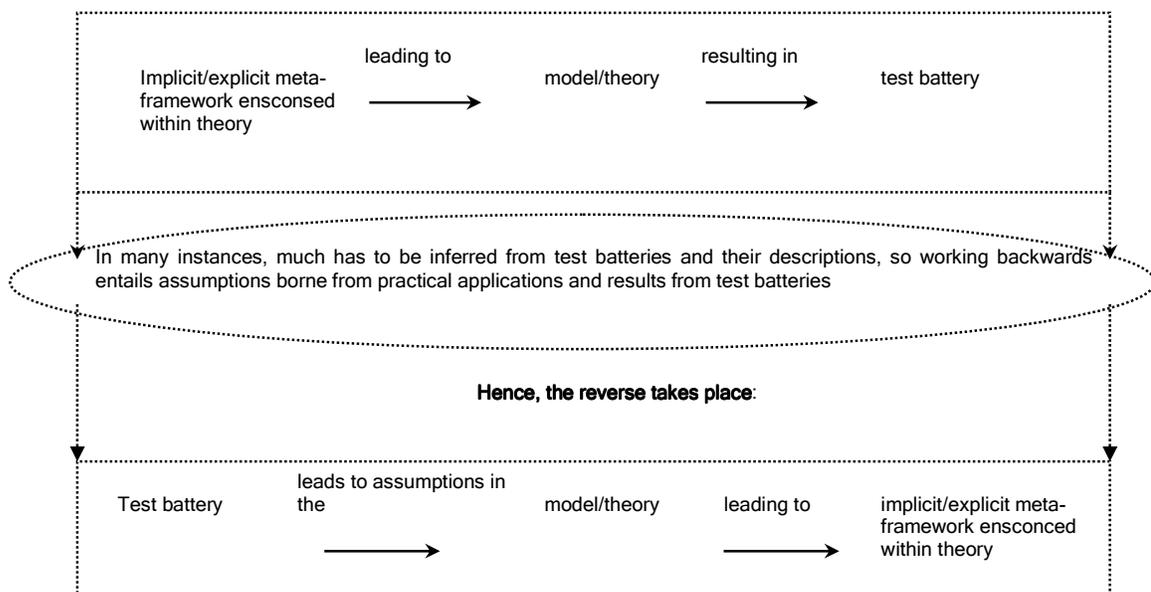
devised which can partial out certain negative features currently resident within some techniques attempting to explore and “prove” that theoretical constructs are synonymous with empirical constructs.

4. Vygotsky is perhaps the most often cited psychologist within Western educational and psychological theory and perhaps comes a close second to Pavlov.¹ The study focused more on Vygotsky’s prevailing context than on Vygotsky the man himself and there is a vast amount of information on that topic. However, retracing the steps to his formulation of what he considered to be potential is illuminating for one reason: that context influences the advancement of science. Simplistic as this insight may be, it is nevertheless an insight which attests to dynamic assessment’s growth in the psychological testing arena. Pushed from one side to the next, Soviet administration and philosophy overpowered many of Vygotsky’s contemporaries leading, in some cases, to their ostracism and even death in interment camps. That such circumstances propelled a man to change his model and thinking in to areas which is so popular today gives rise to great thought. These are of course only partial reasons given for the development of his thinking regarding dynamic assessment. His thoughts on the matter were not developed clearly however.

Having argued for and against what can be levelled at the author as poorly considered questions and also having thus served as introduction to this chapter the exploration will now commence.

5.1.1 Exploring the framework for dynamic assessment and intelligence: the eight considerations

The meta-theoretical framework can be deployed via the eight considerations which have been argued for in support of this framework. Madsen’s original framework has been attenuated and added to throughout the workings of chapters 1-4. Three main strata encapsulate the eight strands and concern ontology, philosophy as part of the meta-stratum; hypothetical terms, scientific hypotheses and hypothesis systems as part of the hypothetical stratum and lastly, abstract data, concrete data and prime considerations (mathematics, measurement and statistics) as part of the data-stratum. Dynamic assessment models will be discussed in their entirety and from as holistic a view point as possible within the confines and limits of this study (space and time). The intention of deploying the framework is not to scrutinise the theories/models on a statistical basis but to provide insight into their placements within their meta-framework. Often the case may be that hypothetical terms, scientific hypotheses and the data stratum concerns will need to be inferred from what is being promulgated by the model as typically expressed through the test battery which is very much the reason why dynamic assessment needs to be explored in terms of its fit in the broader scheme of intelligence assessment, because inference from test batteries to philosophical origins is a leap in terms of the assumptions made. Hence, the process looks as follows in some instances:



¹ If he has not yet numerically taken over citations and references to Pavlov. It is very possible that citations and references to Vygotsky may well eclipse those of Pavlov in the near future, although the author is merely speculating at this point and has no firm evidence to support this.

5.2 Dynamic assessment and learning potential models conceptualised within the framework

Various theories and models will now be investigated as to their placement within the meta-theoretical framework and in many instances such deliberations are resultant on test batteries' empirical results in practical situations. In keeping with the tenets of scientific discourse, no method, approach, theory or model is given precedence over any other and each will be discussed in turn. The order of the discussion follows arbitrarily in this regard. Choices of model/battery/theory inclusion are based on two criteria:

1. That the details of such models/batteries/theories be explicated in monographs due to the advantage of space; i.e. allowing the authors greater luxury in detailing their ideas. Article formats are not deemed adequate in terms of being able to explore more fully the material that is necessitated by such an analysis. This does not mean, however, that such studies cannot be included at a future date. Chapters, books and edited volume editions simply make more sense as chosen format when extended detours are made regarding rationale underlying choice of theory and model
2. Relevance of material. Of the monographs and edited volumes surveyed there is a certain degree of overlap in content such that one model is discussed in two or more separate publications. Hence the number of models included hereunder does not numerically reflect the number listed in such monographs and edited volumes. Also, a number of works and chapters within volumes do not discuss only models per se but also endeavour to discuss the dynamic assessment methodology in theoretical form, hence not detailing theory as such but general issues of concern to the field. Although forming a minority among those discussed in the following, dynamic assessment as approach is also treated and will be highlighted as such when such a case arises. It is expected though, that Madsen's HQ determination might not necessarily be of use in such cases as it is rendered void of relevance

Each model's discussion highlights a separate issue and this issue will be discussed in terms of the broader picture against which all other models are discussed (see Table 20 below). The choice of highlighting specific themes is not arbitrary but reflects a reasoned rationale. There are, no doubt, other aspects to highlight within any one model or approach but the one chosen in each instance warrants further discussion even though other models might share the same concern. A list of the models to be discussed is presented below followed by a brief listing of these above-mentioned concerns.

The following models are discussed:

- Section 5.2.1 - The Leipzig learning test (LLT) and the adaptive computer assisted intelligence learning test battery (ACIL): diagnostic programmes from Guthke and Beckmann (2000b)
- Section 5.2.2 - Assessing the learning potential for inductive reasoning (LIR) in young children from Resing (2000)
- Section 5.2.3 - Dynamic assessment for students with learning disabilities (Dynamath) from Gerber (2000)
- Section 5.2.4 - A model for dynamically assessing intelligence: differences in feedback from Carlson and Wiedl (1978, 1992, 2000) as well as Wiedl (2002)
- Section 5.2.5 - Dynamic assessment of the level of internalisation of elementary school children's problem-solving activity from Karpov and Gindis (2000)
- Section 5.2.6 - The learning potential test for ethnic minorities (LEM) from Hessels (2000)
- Section 5.2.7 - Swanson-cognitive processing test (S-CPT) from Swanson (2000)
- Section 5.2.8 - Feuerstein's Learning propensity assessment device (LPAD) from Feuerstein, Feuerstein, Falik and Rand (2002)
- Section 5.2.9 - The assessment of learning potential: the EPA instrument from Fernandez-Ballesteros and Calero (2000)
- Section 5.2.10 - Application of cognitive functions scale (ACFS) from Lidz (2000b)
- Section 5.2.11 - Analogical reasoning learning test (ARLT) from Schlatter and Büchel (2000)
- Section 5.2.12 - The Mindladder model: using dynamic assessment to help students learn to assemble and use knowledge from Jensen (2000)
- Section 5.2.13 - The cognitive modifiability battery (CMB) from Tzuriel (2000c, 2001)

Table 20 Highlighted issues within the selected models

<i>Section</i>	<i>Highlighted concern</i>	<i>Section</i>	<i>Highlighted concern</i>
Section 5.2.1	Nature of construct	Section 5.2.8	Founding tenets and controversy surrounding clinical and statistical decision-making
Section 5.2.2	Issues surrounding the modelling of “fuzzy” constructs brought about by mediatory attempts to alter it. Change as non-linear and geometric in nature. The manner in which statistical co-variation can truly be said to account for change	Section 5.2.9	Near and far transfer issues and how this is influenced by the hypothetical-substantive construct debacle
Section 5.2.3	Supported rationale for the reification of hypothetical constructs within a narrowly-specified and delimited construct	Section 5.2.10	Curriculum-based assessment. As well as the argument that cognitive assessment cannot be equated with ‘intelligence’
Section 5.2.4	Differences in feedback; assessment and testing	Section 5.2.11	Contention of dynamic assessment as <i>bona fide</i> model exhibiting its own constructs and eschewing traditional notions of dynamic assessment as either model or approach
Section 5.2.5	Differentiation between dynamic assessment as construct innovator vs. construct poacher	Section 5.2.12	Construction and nature of knowledge and how this aspect is altered. This construct can be said to underpin potential in its broadest sense. Impossibility of measuring stability and modifiability concurrently
Section 5.2.6	Rate of change as indicator of potential vs. the usual manner of rendering change as static pre and posttest score	Section 5.2.13	<i>A priori</i> assignment of rotation of factor loading and issues surrounding the nature of transfer. Also issues surrounding static and dynamic philosophical conceptualisations of assessment
Section 5.2.7	The veracity of dynamically altering a static conceptualisation traditionally within the purview of static intelligence assessment		

5.2.1 The Leipzig learning test (LLT) and the adaptive computer assisted intelligence learning test battery (ACIL): diagnostic programmes

Guthke and Beckmann's (2000b) Leipzig learning test (LLT) views the child as a developing organism and seeks to promote process of functioning whilst framing such development within a framework of language difficulties. Differentiating language performance among low performers but offering diagnostic value enhances this manner of test from pure classificatory grounds. The dynamic component arises with the nature of assisted feedback which is markedly stronger than its more simplistic renderings within static-based tests. The test's development has spanned almost two decades and places emphasis on concept formation and analogue with supportive results from a variety of sources attesting to this manner of assessment as valid discriminator of intelligence functioning within low performing individuals utilising developmental psychology theory during developmental progress. Evaluation takes the form of quantification of the amount of assistance required by the individual, the time consumed during testing as well as verbalisation of performance. The child is also assessed regarding their verbal feedback on their understandings of the concepts involved in correctly identifying the processes involved within the assessment. Statistical data indicate that learning potential is evidenced from this test which is not evidenced in normal school performance as initial impulsive behaviours are equalised throughout the assessment and children are levelled out in terms of initial understandings of the requirements necessitated from the assessment. This levelling process is brought about by the nature of the intervention-assessment where the assessment process allows for simulation of learning as would take place in the classroom where errors are corrected and moreover allows for the diagnostic feature to enter the assessment process. Probabilistic models are utilised, among other techniques, to empirically construct and score the test (via the sum of prompts required) and the test is normed thus aligning itself to standardised requirements. Points to ponder include the moderate correlation with the Wechsler, the fact that the LLT which is devoid of language correlates with the Wechsler verbal IQ and lastly, that initial LLT items correlate more with the Wechsler than the ability-testing items in the latter half of the test. Lurian and Vygotskian notions are heavily leaned upon giving to the assessment a stance other than that illustrated in more static-based intelligence assessments. Qualitative evaluation also forms part of the total assessment process where the nature of the errors made are tabulated and as they are strategically placed throughout the tests as distracters, the choice of a particular incorrect



distracter provides clues as to the manner of incorrect reasoning. The assessment is pro-training biased with less weight being attributed to control scores and the end classification of the child is based upon the nature of the learning process engaged in by the child. In sum the assessment highlights are considered as follows:

- Process-based assessment leaning on notions grounded by Vygotsky and Luria
- Allegiance to the dominating force of static-based criteria of normed standards and the perceived necessity of correlating scores and process with established static-based intelligence tests
- Utilisation of statistical techniques although with considerable variation in technique, including among other techniques probabilistic models
- Both quantitative and qualitative analyses carried out serving a two-fold purpose
 - Greater variety of diagnostic information available
 - The mimicking of school-style interaction where errors are not only highlighted but corrected
- Language-free tool for those in the lower performance area
- Greater weight granted to training as opposed to controlled testing, with both contributing to the final “score” or profile generated by the mixture of quantitative and qualitative analyses

These authors have also devised the adaptive computer assisted intelligence learning test battery (ACIL) which is a computer-based diagnostic programme for older students. One of the ACIL’s tests has already been viewed in chapter 4 (section 4.4.2.2) in terms of its contribution to the notion of the construct of change as described within both CTT and IRT based assessments. The ACIL’s predominantly diagnostic complement of test batteries rests upon the assessment of reasoning ability as presented via three modalities; namely figural-graphic (adaptive sequential learning test or ADAFI), numerical (the adaptive number sequence learning test or ADAFI) and verbal (the adaptive analogy learning test or ADANA). Citing face validity as credible marker of intelligence construct, relying on the factor analytic literature evidencing these three areas of functioning as indicative of reasoning ability and revolving around key aspects from cognitive-psychology, the computer-based assessment facilitates the ease with which these constructs can be assessed (Guthke & Beckman, 2000b). Vygotskian proximal development plays a major role in these tests alongside the diagnostic value they offer. Furthermore the assessment is targeted at individuals from less-than adequate environments, those adversely affected by psychological factors, poor performing scholars as well as serving as means of differentiating between these aforementioned aspects and poor performance brought about by brain trauma. Response times as well as feedback regarding specific errors ensure its diagnostic utility. The insightful use of cognitive theory as undergirding item choice reflects the thought that has gone into devising the item pool which was not statistically driven but theory driven which makes this collective battery ever more pertinent to this field (considering the arguments that have permeated this study). Care is taken to validate the construct being assessed for via the empirical items which seek to elicit manifest or empirical equivalents of the hypothetical construct. The deployment of the framework below encapsulates the spirit of both the LLT and ACIL.

5.2.1.1 A(i) Ontology

The model has been specifically devised with human development as central concern and has set out to accommodate such change, even to expect it; hence *a priori* assumptions about change are inherent in the philosophy behind the test. Not much is revealed as to the model’s concern with the mind-brain thesis but does emphasise the notion of freeing the developing child from constraints previously instituted by language as means of assessment. Being void of language and yet being able to reach classificatory status as to the child’s functioning and performance, the nature of what can be known about the potential functioning is affirmed.

5.2.1.2 A(ii) Philosophy

Having addressed what *can* be known (above) the task surrounding *how* this can be know is tackled from the view of cognitive developmental psychology citing Vygotskian notions of proximal development as well as Lurian neuropsychological cognitive processes. Diagnoses via the learning and training phase are ploughed back into reality so the link between test and practice is obviously manifest. The tool’s main purpose is to function as diagnostic aid which is yet able to facilitate prediction to a point. Nomothetic concerns filter into the rationale behind the development and deployment of the test batteries but are in essence idiographically situated where concern lies with the individual.

The whole system is one of individual concern and assessment but is done so through procedures both qualitative and quantitative suggesting a concern with normative veracity and local (individually tailored) applicability. The outcome is finely detailed yet measures constructs typically operationalised as learning potential. The transition from hypothetical construct to one of substantive construct is processed via measurement but more specifically is the manner in which the measures are obtained. The data language employed, as mentioned in chapter 3, tends to mimic the research design which is applicable to this approach. As hypothetical constructs need to be made manifest within an acceptable framework in terms of current envisioning of valid and reliable assessment, the couching of numbers within process is very evident in this model. Numbers are essential in order to convey what can be seen to be progress and development but is done so via detailed analyses of error detection and



the types of errors that are made elucidate the process of development. In a way, this represents a tangible balance to the whole procedure and gives one the idea that the assessment process could go further qualitatively but does not do so for reasons obvious to its marketability within the current assessment context. This prevailing scientific climate (if one can consider it as such in the first place) is at once a constraining feature and liberating one, depending on one's view point; hence the concern with placement within a meta-theoretical framework. The data language is thus typically framed in the common numericised tradition which for so long has dominated psychological research.

5.2.1.3 B(i) Hypothetical terms

The most obvious candidate for hypothetical term is potential which will be similarly highlighted in other dynamic assessments. Potential as an hypothesised construct represents an instance of a mentalistic hypothetical term. Other organismic hypothetical terms make their presence during the qualitative mode of enquiry into error type and description of manner of answering questions after mediatory efforts such as the increased length of time taken to answer questions in the posttest revealing a more sedate approach to answering than the prior impulsive approaches. The move away from ontological reference of hypothetical terms towards the realm of existential hypothetical variables takes the form of mediation which in this instance is achieved through the ramification algorithm clearly seen in the ADAFI. This is an instance of a process variable where change is effected via the process of inducing clearer thought as mediatory tool. Of the three classificatory hypothetical structured variables, the ones most measurable are directive variables and are the most commonly found variables whereas dynamic variables are more difficult to pin down due to their qualitative nature; which the authors have endeavoured to accommodate in their model via error detection analysis (where and how the error came about).

5.2.1.4 B(ii) Scientific hypotheses

The hall-mark of the scientific enterprise as should be clear at this point is the testability of its scientific hypotheses. Regarding this model, this is perhaps its greatest strength. So detailed and comprehensive is its logic and stratagem that replication is easily achieved. Of course one area of gain is another area's loss (pure subjective qualitative analysis). The hypothetical notion of learning potential is made manifest through designed algorithms (descriptive indicators tied back to levels of achievement) as well as more qualitative indicators of achievement such as the number of tasks attempted, the latency time taken, number of steps and number of prompts required. These terms allow for some means of anchoring the notions behind what may entail potential or at least what might be the result of potential (as it may not necessarily be directly measurable). The obvious contender for scientific hypothesis is learning potential as it can be assessed for and measured. That it exists is implicitly implied. The authors suggest a channel through which the assumed construct becomes manifest via the test situation. Concerning the ACIL, it is stated that "reasoning ability" as construct derived from content-related theory is measurable from items (evidencing situations) such as figure sequence, number sequence and verbal analogies and surfaces when an answer is provided. This is delineated as follows: construct - situations - behaviour. This scheme which is made clear by the authors ties in with Madsen's conceptualisation of existential and functional hypotheses. The former being translated into those presumably measurable aspects tested throughout such as reasoning ability, learning potential and inductive abstract reasoning. Changes in scores, pretest-posttest evaluations and the nature of errors made become functional in their capacity to partially yield what may be considered "potential".

Madsen's hypothesis quotient quantifies the relation between theoretical (hypothetical) and empirical (substantive) hypotheses thereby yielding some measure of testability and explanatory power of the theory. Theoretical hypotheses evidence functional relations between two hypothetical variables; H-H, whereas empirical hypotheses evidence functional relations between empirical and hypothetical variables; H-S or H-R. Relations evidencing data which are only empirical are also considered; S-R or R-R. The LLT is premised upon the measurement of the hypothetical constructs "abstract analytical reasoning" in particular inductive reasoning and "concept formation" and as such form functional relations with learning potential which is deduced from the reduced need to train on items. Relations between reasoning and potential as well as concept formation and potential can possibly be construed as two indicators of H-H relations. Empirical hypotheses are the scores resulting from the test battery's training and control phases which is a mixture of coded quantitative and qualitative assessment. The total complement of scores for the LLT is made up of six separate scores which place more emphasis on total training than on control scores. Madsen's treatment of H-H and H-R/H-S leads one to assume that somehow the nature of hypothetical and empirical indicators are of similar nature when in fact this is very difficult to support. It is unlikely that such a rigid utilisation of his HQ system could be beneficial to dynamic assessment models and theories especially if much of the background information is garnered from test battery information. The question that is asked then is how to go about rendering an HQ score without deviating too much from his original system.

It can already be seen that the LLT and ACIL models carry more weight in terms of H-S/R and R-R relations than H-H relations. Six LLT scores can be possibly² understood to represent H-R constructs as they are the results from explicit stimuli, namely, instructional and mediational strategies. Another question that could be asked of this system is how to utilise the scores; either six separate scores or their amalgamation into two final scores. Six scores will be chosen as the author is of the opinion that the battery leans heavily on empirical constructs being measurable and that the results are built up from these constructs thus allowing the test entrance into mainstream assessment. Utilising Madsen's formula and delineating the H-H and H-S/R constructs the following can be formulated for the LLT:

$$HQ = \frac{\Sigma(H-H)}{\Sigma(H-S) + \Sigma(H-R)}$$

Theoretical (hypothetical) scientific hypotheses

- H (intelligence) - H (abstract analytical reasoning)
- H (intelligence) - H (concept formation)

Empirical (substantive) scientific hypotheses

- H (intellectual improvement /learning potential) - S (training on tests)
- H (intellectual improvement /learning potential) - S (mediation via feedback - hints)
- H (trainability) -R (training score)
- H (trainability) - R (control score)

$$\Sigma (H-S) = 2; \Sigma (H-R) = 2; \Sigma (H-H) = 2.$$

$$\text{Hence HQ for the LLT and } = 2/(2+2) = 0.5$$

The higher the HQ the lower the testability making this instance of HQ = 0.5 quite testable indeed. Of course one needs to establish some sort of base measurement before anything definitive can be said. 0.5 may be construed as "high" or "low" depending on the range of scores across models/theories/test batteries within the domain of dynamic assessment. One can possibly keep in mind Madsen's own account of his global assessment of motivation theories and the HQ scores assigned to these theories. This is of course unrelated to the present study but when one considers that some of his models' scores were as high as 2 as in the case of Freud's theory as analysed in one work, a comparison immediately becomes apparent. Recall Skinner's theory score of 0 indicating extreme testability if one can phrase it as such. Once this range is explored a more rounded rendering can be given. In order to do so more such models/theories/batteries will need to be explored. Based on Madsen's own HQ studies within the realm of motivation studies however, it can be stated that this model is very testable. This makes sense given the heavy reliance upon measurement of hypothetical constructs, empirical results attesting to change and end scores derived from the number and quality of hints given throughout. Qualitative error detection is still subjective however, allowing for flexibility in the model via the system of hints which is nevertheless quantitatively coded. Utilising Madsen's formula and delineating the H-H And H-S/R constructs the following can be formulated for the ACIL (encompassing all three sub-tests):

Theoretical (hypothetical) scientific hypotheses

- H (learning potential/intelligence) - H (reasoning ability)

Empirical (substantive) scientific hypotheses

- H (learning potential) - S (systematic feedback/prompts/amount of assistance) - serves as a training function as well
- H (intelligence) - R (figural-graphic: various diagnostic levels assessed)
- H (intelligence) - R (numerical: various diagnostic levels assessed)
- H (intelligence) -R (verbal)
- H (latency as evidenced through learning potential) - R (latency time; time taken per item)
- R (number of tasks)

² This is a thesis, in other words a supported system (by way of logical argumentation and citation) of contentions. It is meant to invoke criticism in order to propel the system being developed and deployed and can in no way be said to represent *the* end product to such as exercise but *an* end product; one such possible end product. The whole point is to develop a workable tool which is currently in its nascent stage of development. Debate is not only warranted, but sought-after at this stage.



- R (number of prompts)
- R (number of steps)
- R (time required to complete the whole test)
- H (error-specific feedback) - R (coding of types of errors)

$$\Sigma (H-S) = 1; \Sigma (H-R) = 9; \Sigma (H-H) = 1$$

Hence HQ for the complete ACIL = $1/(1+9) = 0.1$ making the complement of the ACIL more testable than the LLT which stands to reason when one considers the scores made manifest from the various sub-tests in the ACIL.

5.2.1.5 B(iii) Hypothesis system

Mediating between the more grounded data-level and the meta-stratum level, the hypothesis system seeks to define via explanatory means the logic involved in the model. The deductive explanatory system is typified by the traditional premises followed by logically deduced conclusions, which in the case of the LLT would follow its facility in differentiating poor performing students from impoverished backgrounds from those hampered by language disorders via a process of mediatory compensation within a system of hints aimed at providing a micro-lesson within the learning potential paradigm (see section 5.2.1.4) below for a model exposition detailing a system of feedback). The process is emphasised in comparison to the end result or product being emphasised. Classification is considered a primary mover in performance within the LLT and as such is assessed in a myriad of ways along different criteria involving shape (form) and size moving through three stages of performance; assistance required, level of complexity and learning process across the full procedure. The explanatory system used here is more firmly aligned to the traditional nomological-deductive account of science as not much is left up to relativist accounts within the scoring procedure. An exemplary model is built for the ramification algorithm for the ADAFI sub-test of the ACIL which appears in figure 67 (chapter 4). This model traverses the model's landscape so to speak in a manner facilitating the understanding of what precisely is meant by change and how it is characterised. The model is rendered in two dimensions and graphically depicts the movement of how progress is made through the testing situation.

5.2.1.6 C(i) Abstract data

Here the territory becomes less firm as Madsen, as stated in chapter 3, relates the caution attached to the blurred boundaries between abstract data and hypothetical terminology. It is at this level that the science of everyday experiments and trials proceed and which can be evidenced in journal articles and book chapters and illustrates results from various test batteries deployed in practice. Functional relations and correlations form part of this level. The functional relation for both the LLT and various tests making up the ACIL take the form of assessment via hints and prompts, re-assessment and end results as diagnostic means and micro teaching environment. The only instance of a correlational measure to be noted within the LLT and ACIL is the correlation between the non-verbal LLT and the Wechsler verbal IQ scale which need not be causally related.

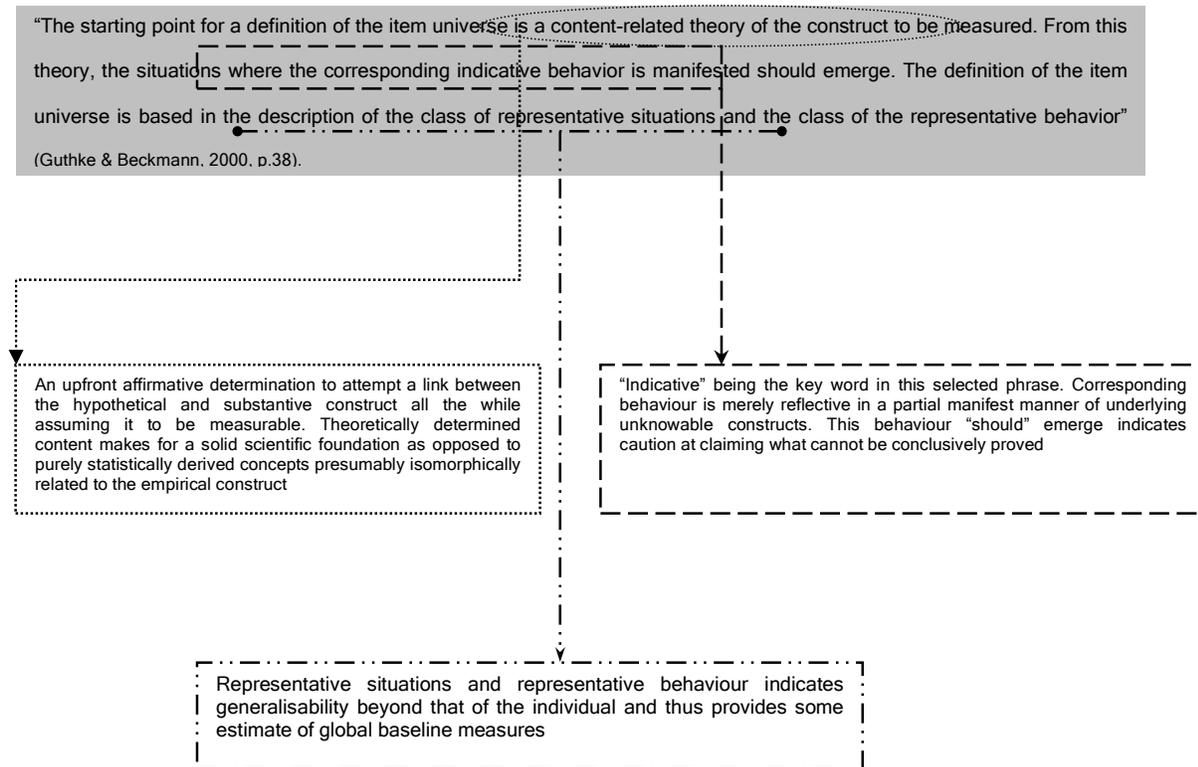
5.2.1.7 C(ii) Concrete data

This level or stratum is really the most concrete in terms of the entire framework and concerns itself with obviously manifest and observable behaviour either behaviorally or in terms of scores on tests items. For instance, working at a slower working pace illustrates less impulsive reactions to test items and greater attention to more difficult items. The mode of error detection is also qualitatively aligned and is scored on manifest behavioral outcomes. As with many other learning potential or dynamic assessment batteries, concrete data are prominent as they are often the only tangible aspects with which to work. The most obvious levels at which to work are the data-stratum criteria as they are most easily identifiable.

5.2.1.8 C(iii) Prime considerations

The authors are deliberate in their approach towards the assessment of potential and link firmly back to psychometric theory and the definitive role that traditional modes of assessment plays in adhering to valid and reliable norms as stipulated by governing bodies within psychology. The complexities surrounding the learning potential movement and its vast array of techniques are firmly bounded by the conventions of traditional methods of assessment. Lamentable as this may appear to some dynamic assessment purists (models of whom will appear later), the authors contend that in order to realistically align learning potential test batteries within the market place, such concerns need to be taken seriously and thus certain core psychometric concerns such as traditional reliability and validity are weighed heavily against the loss of such information within stricter dynamic assessment models. The authors achieve a versatile balance in maintaining classical test theory norms whilst introducing modern test theory concepts. The LLT and ACIL complement both ends of the continuum much the same way that many learning potential assessment tests do. Test items are individualized yet reliably scored in keeping with item response theory models of assessment and due to the comprehensive item pool, content validity is made more meaningful as items are based on cognitive and developmental theory and not sourced statistically. Hence, subjectively scored items (classical test theory) make way for objectively scored items (modern test theory). Moreover, adaptive testing fits the learning potential bill in terms of its

flexibility. The LLT procedure was developed utilising multidimensional scaling (MDS) or smallest space analysis, hierarchic cluster analysis (to determine learning types) and a probabilistic model. Norm referencing enables a baseline according to which some measure of generalisability can be enforced. MDS encompasses the usual relations found in correlation matrices in addition to offering more information as to their spatial distances thus allowing for more underlying dimensions to come to light than would be possible utilising only factor analysis. One must however acknowledge the role that subjective choice plays in determining which orientation of axis will yield the most plausible explanation of underlying related dimension which necessitates *a priori* decisions, which in a way, is similar to the concept behind face validity. Nevertheless, MDS allows for greater flexibility as it is not dependent on linear relationships and requires only that differences between variables are meaningful at the rank-ordered level; i.e. it is a non-metric approach. Immediately, the appeal of MDS can be seen especially when taken in tandem with the discussion in chapter 4 regarding the traditional scales of measurement and the necessary requirements that need to be adhered to in terms of allowable statistical analyses. The ACIL test complement is premised upon the face validity of reasoning ability which can be understood to be dangerous in terms of traditional nomological models of science as well as issues surrounding empirically measured hypothetical constructs. Translation from the hypothetical realm to one of numericised substantive realm is always a difficult area. Moreover, the authors state that the ACIL tests assess for not only actual competence but learning potential. It is worth citing the authors in this regard:



The ADAFI's item pool lends itself to modern test theory item analysis exceptionally well. Complexity levels between and within items make for fine discriminations per item and is tailored to each individual's manoeuvring through the item pool, thus yielding valid and reliable indices which are attuned to each individual. The number of alternative paths towards completing the test allows for its flexible utilisation across many “strains” of learning potential discriminations. The highlights in terms of “generalisable individualism” (if such a term can be accommodated for the moment) in all three sub-tests of the ACIL consist of:

- Multiple paths towards completion of the test
- Allowance of deviations which are specific in targeting the testee's path from no or little initial ability to successful completion of questions evidencing potential (not ability but potential, as it is assumed, however tacitly, that ability does not change!)
- Qualitative indices of performance from the reliance of error detection and the subsequent quantification of such coding
- Fine discriminators in terms of reasoning rules that need to be learned and applied in the test. Understanding colour but not form is indicative of errors made in terms of sequence progression
- The inbuilt teaching component of the test allowing for not only diagnostic inferences to be made but practically ploughing back into the testee some form of teaching lesson. This is done by illustrating the heuristics involved in learning the analogical reasoning tasks

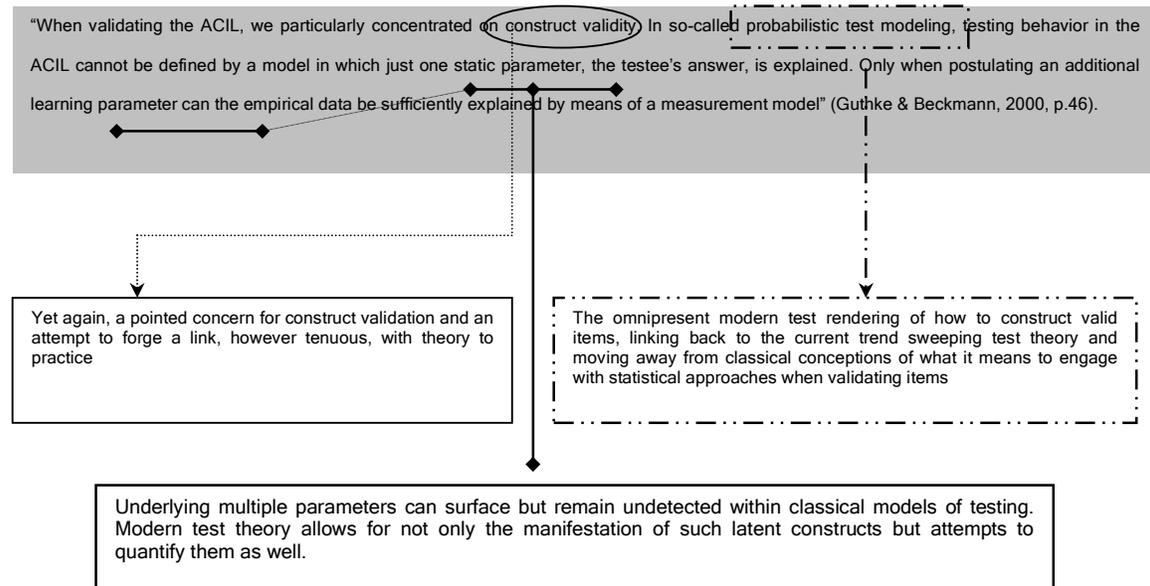


- Various dimensions which are assessed in terms of answering questions; time, latency, number of steps and so on

Traditional determination of reliability is ascertained by Cronbach's alpha which, as expected, turns out to be lower than the norms for traditional mainstream assessment. Should this even be of concern?

As potential is sourced (assuming there is potential to be sourced)³ and is premised on change (see chapter 4 for discussions on this topic) surely the reliability indices will be of no use especially if the items change in the nature of their construct. If they do not, then high Cronbach alphas are acceptable. It is assumed that the authors felt compelled to insert a paragraph attesting to some form of reliability. The author maintains that this should not be done for the sake of placating standardised assessment methodologists but should reckon with full force that their methods are applicable to their domain of potential assessment and argue for it as such. Paying allegiance in such scanty manner to both sides is part of dynamic assessment's muddle. Take a view and run with it. The case stands with the authors' attempts at proving validity of the ACIL complement yielding predictive results from criteria usually made use of, namely, school marks. Once again, we encounter a vicious circle. Why validate in such a manner at all? Unfortunately, as with Vygotsky's manoeuvrability within the Soviet system, so too are researchers forced to move around in a system necessitating the delivery of goods as the prevailing climate sees fit. That such a pragmatic reason should propel concerns within a discipline supposedly progressing in terms of scientific dictates (notwithstanding that such dictates are robustly nomologically centered) is all the more reason to ponder the strange circumstances. However, in their defence they are keenly aware of modern test theory's attempts at providing far superior indicators of valid measures as they pertain to hypothetical constructs and as with reliability, the following is cited as evidence in support of their use of modern test theory as it pertains to validity-bearing concerns:

³ Recall the author's opinions regarding the need to assess dynamically in the first place. Dynamic assessment is not or at least should not be a tool utilised for the sole purpose of "finding" potential when there may be none to be found. Dynamic assessment is not necessarily the disadvantaged person's leverage to reach into the world where such heights are not necessarily attainable. It should rather be a tool for finer discrimination at a level heretofore not recognized by mainstream assessment. The two types of assessment cannot be thus diametrically opposed, neither philosophically, nor scientifically. The author is at pains to emphasise this point.



Indices of learning ability within the AZAFO are calculated and derived from z scores which indicate not only increased difficulty levels of items but ease with which such items are completed. The lowering of z scores obtained by a testee as the difficulty increases is taken to reveal potential. The LLT and ACIL complement of tests is at once a marketable test given the current pervasive (but ill-informed) psychometric test view popularly understood to be “correct” and an endeavour into some uncharted waters (not all is unknown). Can the balance be maintained? The HQ allotted to these batteries is yet another indicator of their testability as models but has shown substantial gains in proving that latent constructs may well be quantifiable if this is the route that is to be taken. Two key words spring to mind when contemplating these two batteries which serve as conclusion: “testable” and “modern”.

5.2.2 Assessing the learning potential for inductive reasoning (LIR) in young children

Resing’s (2000) approach towards fairer assessment encompasses notions of Vygotskian ZPD and its cognitive development operationalisation as well as cognitive training research. Embedded in the method and nature of the approach is a hint structure as initially developed by Campione, Brown and colleagues. This research into the learning potential paradigm is noted for its reliance on structured hints. The reverse figure of the number of hints required exemplifies the width of zone of proximal development available. The fewer the hints the larger the width. Of importance in this model and approach towards assessment is the acknowledgement of the uncorrelated nature of process and product with four main classes of groupings coming to the fore:

- i. Those performing highly on conventional assessments but poorly on potential assessments (this can be explained in two ways: a ceiling level being reached; at least with older dynamic assessment tests and disruption of already-functioning cognitive systems. It can then also be argued that alternative approaches towards the solutions of problems exist which are not accounted for in learning potential approaches)
- ii. Those performing highly on conventional assessments and highly on potential assessments (expected)
- iii. Those performing poorly on conventional assessments and highly on potential assessments and (most common)
- iv. Those performing poorly on conventional assessments and poorly on potential assessments (can be accounted for)

(i) and (ii) can be said to manifest similar IQ scores yet need very different intervention strategies, if at all, whereas (i) and (iii) evidence different IQ’s, hence they may necessitate similar intervention strategies. There is not one singular method that will work across the board. The system or approach utilised within the LIR is predicated upon a now readily identifiable construct of inductive reasoning operationalised through analogy tasks; in this instance verbal and visual exclusion tasks. This reliance on traditional constructs as meaning-making (see section 5.2.11) is most often the route followed within dynamic assessment, hence its possible classification as construct-poacher through approach versus construct meaning-making anew. Learning potential is related to intelligence via its operationalisation of inductive reasoning. The link is thus forged and it is not only intelligence that can be defined by way of shaping the environment but learning potential as well. Although there is a move away from product to process within conventional intelligence circles, the emphasis still remains on attributable numerical scores (but at least “process” too is being scored). The LIR seeks to assess intelligence via learning potential which leads one to wonder about the construct blurredness of the initial stance on the learning potential paradigm. Once this choice is made at the outset, the entire procedure is coloured with a sharp veneer of conventional methodology. Nevertheless, the approach can itself, as stated, induce change and hence elicit a new construct no matter how modified the original one might become. The target



population for the LIR is 7-8 year old children and this age is often considered crucial in the determination of future categorisation in terms of learning difficulties expressed resulting in (mis)classifications of sorts. Educational settings are sought to be rearranged based on the performance during assessment, a task almost never engaged in with static assessment initiatives. Resing (2000) refers to her model or approach as a procedure and this colours the understanding of the model utility and value.

Intelligence is defined as the ability to learn, utilisation of metacognitive awareness (executive control), the efficiency with which information or skill is learned, the degree to which novel items/procedures or concepts are acquired when given only limited instruction and the proclivity to transfer meaning to new situations. The construct of intelligence is thus changed but is done so through a process of accommodating conventional ideas in a process based approach. The “intelligence through potential” concern is a novel way of appeasing both sides of a two-dimensional continuum, if that is how it is perceived. As with many models in dynamic assessment, the need to fulfil the dictates of both approaches results in the acquiescence of both in forging a greater alliance for the greater good of assessment. Being too accommodating, however, can have its drawbacks. Academic intelligence is the more narrowly specified construct looked at in the LIR in terms of assessing for capacity and hence the following manifestation of the mix between approaches is evidenced: learning efficiency as speed of information processing within cognitive activities where cognitive strategies include selection, sequencing and modification which, it is endeavoured, will be transferred. An appealing feature of this approach (at least as far this author’s leanings go) is the appreciation of constraints within the larger system and by constraints is not meant the inhibiting of functioning. What is conveyed by this sentiment is that genetic heritage is both constrained and open to malleability (most likely meaning phenotype) and in reference to the transactional approach towards intellectual functioning the author cites this environmental-genetic interactive process.

In keeping with this tenet of development, the philosophy behind the need to dynamically assess takes cognisance of the limit to which potential is restricted (recall the often quoted statement that potential is not equally apportioned in those for whom this method functions well). In a reasoned plea, the author states that standardised assessment should form part of the package of assessment tools and in relation to the LIR itself, this would make sense due to its emphasis on inductive reasoning which forms a major part of the learning process within school related subjects. The LIR, is for this reason, correlated for predictive and concurrent validity with conventional IQ tests. If performance in school is the necessary requirement for success in life today, then it is hardly unreasonable to accept the need for such relations to be made evident. The problem arises when constructs are assumed to be made anew when in fact they are merely transformed. Vygotskian proximal developmental space is espoused by the number of hints required throughout the process but is scored based on the least amount of help required in the hope of progressive independent functioning which is precisely what is envisaged by the ZPD. A gradual accretion of mediated skill is transferred from assessor to learner in an environment stressing structured guidance for constructs utilised in mainstream testing in a manner typical of a pre-posttest set-up. As with other models in this approach, the mediatory phase is not scored as it is often not interpretable. This “fuzzy” stage of development is an odd time which is extremely difficult to model (physically and statistically) as it is often the case that predictions based on this phase are invariably skewed and incorrect. Is a construct even identifiable? The situation can be viewed as follows:

Pretest construct → Change construct → Posttest construct

Pre-assessment results with constructs being defined and cognitive and non-cognitive aspects accorded a level of functioning. This level represents a clear picture of functioning within the test set-up. Change is not induced (or is kept to a minimum). Predictive patterns can be ascertained with relative validity and reliability as variance accounted for in both current and future tests are co-varying. The tasks are similar in nature so it would be logical to infer this in the first place. Neither construct (in current and future assessment) nor individual functioning is challenged.

Fuzzy area of construct malleability. Cannot be utilised due to its unpredictability as construct. What kind of a construct is this? It can be construed as hypothetical, empirical (substantive) or statistical. What makes this so unpredictable? Is it the continuously changing variance accounted for? The instability of co-variation? Why is this construct so slippery? Change is being produced so one can say that it is nonlinear. Change is geometric. So, is it our measurement tools that are unable to adequately utilised and understand for the change? If change was exclusively evidenced via physiological renderings, then this exercise would be far easier to understand. But the mentalist and behavioural phenomena make this incredibly difficult; almost intractable. Is the construct so difficult to "know"?

Post assessment is based on a changed construct and a changed behavioural repertoire. Prediction based on the construct in future instances will most likely yield high correlated accounted-for variance. The change (if any) has already taken place. The correlations will not change, the scores however, in terms of their interval and ratio scales will evidence higher numerical results. Post assessment can thus be treated in very much the same way as pre-assessment. No change occurs (because it already has). Post assessment is usually found to predict better than pre-assessment. In other words, more variance is accounted for. Can one conclude then, that better accounted-for variance is due to change or improvement? Scores and variance are completely different issues; low scores still lead to good prediction just as with high scores. No matter what the score is, prediction has nothing to do with numerichised outcome. For instance, score "15" on item A on pretest and so on for another 10 items. The posttest evidences, for instance, very similar scores in terms of numeric scaling but varying across items in a different manner. Hence, score "18" for item A on posttest and so on for another 10 items. Suddenly the posttest results are more reliable indices of future results. The combined accounted-for variance is all that has differed. Poor performance on pretesting and poor performance on posttesting still delivers a change in prediction. How?

- Three individual constructs?
- Three related constructs?
- Accounted-for variance differs. Or does it?
- Co-variation leading to correlational significance: what does this mean in terms of the change undergone in the hypothetical construct? Moreover, does this construct become manifest in substantive reality?
- Predictive validities differ. Why? The construct has changed, but how is this reflected other than through the numerical increase in scores? Does any model address this? IRT change-based models come closest but even they do not provide these types of sought-after answers
- Oh dear.....

?

?

5.2.2.1 A(i) Ontology

Humans are, thankfully, viewed rationally. Genetic and environmental concerns allow for growth yet within a bounded system. Change is borne through a process wielding tools of mediation following on from Vygotsky. Cognition, information processing, learning capacity and intelligence are melded into one domain of functioning made manifest via inductive reasoning tasks such as verbal and visual exclusion. Brain and behaviour can change but the abandonment of mainstream assessment is not advocated. The human being is considered through a psycho-physical lens, hence the reliance on transactional perspectives as providing a platform from which to engage in assessment. Intelligence, as assessed for currently, is accommodated and even embraced. Change can be known but should be known through learning potential within intelligence conceptions as accepted in static conventions. A slogan could even be formulated for the LIR's ontological approach: 'potential through intelligence', where intelligence and potential co-occur and can be known. How this is known is not as quite easily achieved.

5.2.2.2 A(ii) Philosophy

Having grounded its ontology in knowing change, epistemologically the aim of the LIR would be to assess for such change in a manner befitting the model's aims of assessing learning capacity through intelligence constructs in a procedure imitating both conventional and dynamic assessment methods. Human's abilities are preordained by nature but outcomes are not necessarily linearly derived from initial predispositions. Progress and development can occur either in the predicted path or can alter course along regressive paths which need to be corrected and placed back on course. The concern for both inherent and environment is evident in the philosophical beliefs into human intelligence assessment. The portal of entrance in achieving this rectified trajectory is through the assessment and development via guidance of inductive reasoning. Standardised learning potential assessments are the middle way achieved in this effort. The LIR is administered individually thus aligning with idiographic ideals of assessment but is also valid and reliable with normative figures giving credence to the approach, at least as viewed from mainstream assessment circles. The data language is enveloped in standardised format but falls back on qualitative assessments of behavioural functioning in terms of guidance afforded by the assessor. Learning potential is operationalised through transfer and analyses of transfer tasks and this score is buffered by posttest scores and the lowest number of hints required for successful independent solution of problems. Research methods mimic those found within mainstream assessment. Time taken to answer items correctly is a feature common throughout dynamic assessment models and can be evidence of the "fuzzy" latent construct induced during training. Time taken refers to a rate at which progress is made which is most likely also not linear. It may well evidence exponential characteristics and in fact it would be something to follow up on in various models. Modelling of rate might well given an indication or at the very least a better indication of learning potential (see section 5.2.6).

5.2.2.3 B(i) Hypothetical terms

There are fewer organismic terms when compared to the number of mentalistic terms in the LIR model or procedure. The mentalistic terms refer to the constructs assessed for and include most of those listed under the scientific hypotheses section. Aspects such as "width" of ZPD, learning potential as evidenced through hints administered and performance on inductive reasoning tasks can be included here, although rightly they belong to scientific hypotheses and are reflective of more directive variables than dynamic variables as construed by Madsen. This is perhaps a reason why the model is so testable. Fewer hypothetical terms pervade the procedure.

5.2.2.4 B(ii) Scientific hypotheses

Ensnared in intelligence construct assessment and empirical analysis of these constructs is the anticipation of this procedure as exemplified through the combined model and approach as testable. Hypothetical terms are operationalised in this approach which is always a fitting characteristic especially for a framework such as that developed by Madsen. Of course, his framework was originally conceived for the application to motivation theories which by their nature lent themselves to more H-H conceived concepts. Biologically attuned motivation theories and mentalistic attuned theories resulted in an array of HQ results. Dynamic assessment's dual purpose as construct meaning-making and approach in addition to its activities on a two dimensional continuum, colours the HQ spectrum in a manner biased towards testable theories, models and approaches. Is this an echo of prevailing nomological contexts? Or is it a fervent attempt to move psychological assessment along a progressive front of novel construct definition? Most models assessed in this chapter are highly testable. Is this telling of the true story behind the movement? Or is it reflective of the pressure to conform? MLE and pure Feuersteinian mediation within dynamic assessment can result in both low and high HQ results. It all depends on the context.⁴

Theoretical (hypothetical) scientific hypotheses

- H (inductive reasoning) - H (verbal analogies)
- H (inductive reasoning) - H (visual exclusion)
- H (learning potential) - H (width of ZPD via minimum hints required)

Empirical (substantive) scientific hypotheses⁵

- H (Learning evidenced through verbal analogies) - S (encoding)
- H (Learning evidenced through verbal analogies) - S (inference)

⁴ This is an unfortunate yet often quoted statement when things cannot be explained within the social sciences or when results are too disparate to warrant a firm conclusion, the statement of "it all depends" perennially creeps in. Of course, this is also invoked within the natural sciences as entrenched in the nomological deductive framework but it is often (not always) eventually explained beyond the tentative misgiving of "it all depends".

⁵ See Resing (2000) for the specificities of the "unpacking" of the formulation for correctly processing and identifying analogies and exclusion tasks.



- H (Learning evidenced through verbal analogies) - S (encoding (ii))
- H (Learning evidenced through verbal analogies) - S (mapping and application)
- H (Learning evidenced through verbal analogies) - S (encoding (iii))
- H (Learning evidenced through verbal analogies) - S (comparison)
- H (Learning evidenced through verbal analogies) - S (response)
- H (Learning evidenced through verbal analogies) - S (justification)
- H (Learning evidenced through verbal analogies) - S (response (ii))
- H (Learning evidenced through visual exclusion) - S (encoding)
- H (Learning evidenced through visual exclusion) - S (inference)
- H (Learning evidenced through visual exclusion) - S (comparison)
- H (Learning evidenced through visual exclusion) - S (response (i))
- H (Learning evidenced through visual exclusion - S (extended feature analysis)
- H (Learning via metacognitive hints) - S (decreasing problem space)
- H (Learning via metacognitive hints) - S (problem recognition)
- H (Learning via metacognitive hints) - S (activation of existing knowledge)
- H (Learning via metacognitive hints) - S (strategy selection)
- H (Learning via metacognitive hints) - S (planning)
- H (Learning via metacognitive hints) - S (monitoring)
- H (Learning via metacognitive hints) - S (checking)
- H (Learning via metacognitive hints) - S (systematic response check)
- H (Learning via metacognitive hints) - S (learning from feedback)
- H (Learning as evidenced in scores) -R (number of hints for analogies to reach criterion)
- H (Learning as evidenced in scores) -R (number of hints for exclusion to reach criterion)
- H (Learning as evidenced in scores) -R (post analogies)
- H (Learning as evidenced in scores) -R (post exclusion)
- H (Learning as evidenced in scores) -R (number of hints given for analogies during training)
- H (Learning as evidenced in scores) -R (number of hints given for exclusion during training)
- H (Learning as evidenced in scores) -R (amount of time required for training of analogies)
- H (Learning as evidenced in scores) -R (amount of time required for training of exclusions)
- H (Learning as evidenced in scores) -R (metacognitive or task specific hints required during analogies training)
- H (Learning as evidenced in scores) -R (metacognitive or task specific hints required during exclusions training)
- H (Learning as evidenced in scores) -R (nature of justification given for answering in the manner answered in analogies - more qualitative)
- H (Learning as evidenced in scores) -R (nature of justification given for answering in the manner answered in exclusions - more qualitative)

The following HQ score is detailed:

$$\Sigma (H-S) = 23; \Sigma (H-R) = 12; \Sigma (H-H) = 3$$

Hence HQ for the LIR = $3/(23+12) = 0.08$. This hardly needs explaining given the nature of the procedure detailed above. There is a clear preponderance of operationalised variables assessed for. Does the paucity of H-H hypotheses indicate anything other than its empiricised version?

5.2.2.5 B(iii) Hypothesis system

The approach is testable mainly due its deductive explanatory approach towards the assessment of learning potential through intelligence. The rational follows logically from the acceptance of dual concern for static and dynamic protocols utilised within the procedure. If inductive reasoning underpins school related tasks as well as conventional IQ tests; and potential can be assessed through this framework in a manner identical to Vygotskian zone traversing, a compromise can be reached in terms of accommodating both frameworks within one larger system. The procedure is not overly abstract, as it represents process as opposed to content, the former more aligned to less abstruse hypotheses. The procedure does concentrate on a very narrow feature of intellective functioning but is also qualitatively bound to assist in attaining correct strategy use in order to solve items more successfully. The end goal is the independent functioning and solving of items which evidences the width of the ZPD. The reader is directed to figure 9 for an illustrative exposition of this notion and how it is conceptualised. The hypothesis system, as identified by Madsen seeks to mediate in between the data and meta-strata and does through a network of connecting empirical data and meta concerns. Hypothetical connections form between ZPD, cognitive training and efficiency, transactional nature of change, transfer width (directly connected to ZPD), inverse hint score (also width of ZPD) and the capacity to profit from incomplete learning.



5.2.2.6 C(i) Abstract data

The terminology included in the LIR does not remain unoperationalised for very long. Correlational and functional abstract data both form part of the data stratum. Although not entirely causal, the acknowledgement of genetic and environmental impingements results in correlational outcomes. But the role played by the procedure ensures that most variables are functionally related. Rate of change for instance can be viewed as both functional and causal.

5.2.2.7 C(ii) Concrete data

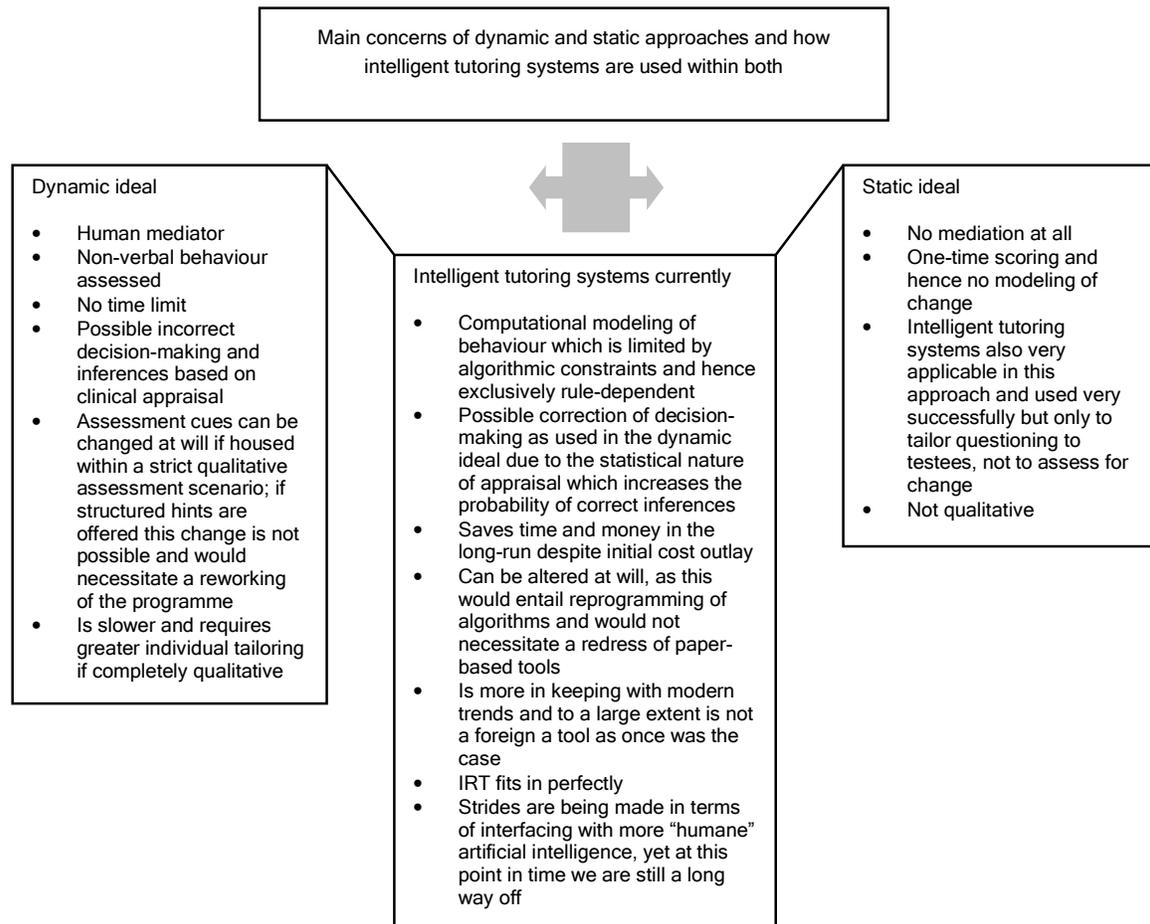
Both top-down and bottom-up renderings of the LIR within the framework are accommodated. Acknowledgement of biological functioning is a top-down concern which can be altered as perception with grounded data forming a bottom-up process so characteristic of how science is practised within the natural disciplines. Framed within theory though, certain hypothetical positions are hereby refuted or supported. Another top-down approach in the procedure is its reliance on mainstream assessment constructs emanating from cognitive findings.

5.2.2.8 C(iii) Prime considerations

Being firmly rooted in conventional methodology and construct borrowing, the LIR proceeds to market itself as both technique and test. Tasks have been taken over from other IQ tests and used in a process-based manner. No new construct is thus being offered. Psychometric analysis of items still takes precedence as well as the concern with the psychometric properties of the LIR regarding norms, reliability and validity. It is not surprising to find high Cronbach alphas for both the verbal analogies and visual exclusion tasks as these have merely been supplanted into a new context. Even though time can be used as indicator of potential, the duration of the test is still very limited which leads to the questioning of the degree of transfer possible, seeing as transfer is a main point of concern in the LIR. Construct validation is expected to be as high as it is. Of interest, in terms of new construct meaning-making, is the unique variance left over after accounted for variance has been explained. In this regard, it is stated that differences in both the content and problem solving tasks may result in the novel variance. Can one equate “learning potential” and “unique variance”? This is a statistical definition of a hypothetical construct yet to be substantiated. What exactly is meant by this? How can one logically equate two conceptually different notions of change when one is based within statistical variation and the other in physical reality? This has been discussed at length in chapter 4. Recall the fact that the latent construct can never be known and hence can never substantially materialise. The logical next step, it can be argued, is to verify its existence as statistical artefact which in this case happens to be unique variance unaccounted for in the co-variance model. Are we back to the beginning in terms of reconceptualising what it means to engage with dynamically assessed and dynamic components of potential? Will the construct forever be veiled?

5.2.3 Dynamic assessment for students with learning disabilities

Computer assisted technology is proving invaluable in settings where access to computers poses few or no problems. With minimal training, learners can engage with an intelligence tutoring system in a manner similar to that offered by face-to-face contact as is typified within traditional assessment settings. Successive levels of technological advance were briefly looked at in section 2.8.6 and how such technology has aided and will continue to do so in the future. As mentioned a number of times, South African dynamic assessment research is very often applied in secondary and tertiary education sectors but due to time and costs these applications are proving to be infeasible unless administered in computer assisted format. This lessens the time needed to administer the test, makes use of the already-existing infrastructure at tertiary institutions (most accredited higher education institutions in South Africa avail of computer laboratories) and assumes that student familiarity with this technology should not pose a problem as they will invariably need to know the technology for their studies. It is in this spirit, that Gerber's (2000) *Dynomath* is placed within the Madsenian framework. Another positive feature about the use of intelligent tutoring systems is their snug fit computationally, conceptually and practically with IRT modelling; the two go hand in glove so to speak. Having waited for almost forty years, the mathematical underpinning of IRT can finally flourish due to the requisite computational power at hand. *Dynomath*'s description does not entail IRT-based concerns as delineated in the literature but the utilisation of this modern test theory is obvious. Chapter 4 discussed this issue at length. The computational model employs dynamic assessment components, which, although not strictly dynamic in terms of human interaction does nevertheless comply with most dynamic tenets. Added to this is the fact that as time progresses the need to assess via technology is escalating. This model assesses domain specific tasks in comparison to some other models discussed in this chapter and is an instance of the very specified nature that an intervention can take. Below are delineated some concerns in terms of bridging the static-dynamic gap as applied to this computer model:



Typical notions of dynamic assessment rhetoric are maintained in a manner still manifestly interactive but are operationalised in the algorithms themselves. The conceptual reliance on Vygotsky's ZPD becomes visually available and one is in fact able to literally see the zone of next development as interspersed between items that can be answered with minimal or no help and items requiring a lot of assistance. Moreover, depending on the degree of specificity, this can literally be so finely tuned that ZPD's can be created for almost any function. As far as this author is aware, finer discrimination is far better suited to computational power than is the human mind and Gerber (2000) as much as admits this when he refers to the speed with which rules are implemented, assessed and compiled. Visual depiction of zones might at first glance seem counter-intuitive when assessed from an interactive and wholly qualitative human interactive process but as times change so does technology and this might really just be a case of practice catching up to technology. The theory, however, remains the same. Or does it? Is it possible that with future explorations into this field that computational modelling and artificial intelligence can somehow delve deeper into latent trait structure? IRT affords us the opportunity when assessing for item characteristics as well latent traits from a test theory point of view which is still very much grounded in the mathematical computations underpinning it. Is it possible that intelligent tutoring systems can see something else about the trait? IRT uses such systems as a base, so the question may not be entirely feasible to ask in the first place.

In the ensuing discussion, when the word “model” is used, it will refer to one of two things: the model behind the nature of assessment and the computational model as defined through algorithms. The difference will be noted as such. The philosophical model is predicated on searching for and delineating the zone of proximal development as originally conceived of by Vygotsky. The ZPD is thus defined numerically and accurately so, as it represents a summed score of individual areas of concern during the assessment of the specified domain. The specified domain is multiplication and is predicated on what is needed in order to complete multiplication sums, namely, retrieval and spatial mnemonics and is housed within both procedural and declarative performance. The information-processing paradigm is an apt launching pad as such domain specificity usually does not entail non-cognitive factors beyond the usual assortment of impulsivity and nervousness. Information processing as approach within intelligence assessment is useful as framework for relatively isolated types of activities such as multiplication. Physically, the system represents a space both visually and conceptually when multiplication arithmetic is given.

5.2.3.1 A(i) Ontology

Learning is conceived of as a dual process of assessment and adjustment. Instruction is emphasised within this model but from a dynamic vantage point. Cognition is still related to reality much the same way it always has been but a computational tool is now used as intervention which seeks to build a composite profile of performance based on a rule-inference system. Although cognisant of the broader functioning of the human within context, a narrow construal of one small cognitive activity does not warrant the extensive overview of complete functioning for this model and this is particularly emphasised by Gerber (2000, p.266) “our question aims at cognitive universals that may operate independent of social and cultural contexts”. This ontologically aligns the model with what is knowable in terms of specified cognitive tasks and epistemologically that it can be known via tailored pure cognitive assessment. In tandem with this sentiment expressed by Gerber is his support for the lack of human interaction in just such a scenario; “it is the ongoing social relationship between competent adults (or peers) and developing children that occasions appropriate moments of mediated learning. Nevertheless, it is reasonable to ask *how much of this process represents cognitive functioning that can disembodied*” (own emphasis) (p.266). Does this take the argument back to one of static vs. dynamic assessment modes? The author does not feel this to be the case. In such a narrow context, surely, a cognitive task can be “disembodied” without fear of loss of information attained by assessing for it in a less formal fashion.

Recall that the process is dynamic, as much as can be achieved given the constraints. One opposing argument could be, however, that static-based items are used all the time in a dynamic manner without recourse to changing the construct. Is this not the continuing debate throughout this thesis? This model does not create a new substantive construct and moreover, it as much states this. Training by way of aided instruction takes place. So can this method even be considered dynamic? Perhaps one can refer to such midway models as “fringe dynamic assessments”, which would then also make it “fringe static assessments”. In terms of human freedom as expressed by Madsen in connection with this stratum, human cognition is framed not only by a rule-based computer system but also by a rule-based arithmetic system. Hypothetical constructs nevertheless do play a major part as underpinning the ability to perform the task such as mentioned above (retrieval and spatial-mnemonics). Knowing how to do something and knowing what to do (procedural and declarative) play different roles in how arithmetic tasks are solved and this is spatially mapped as a theoretical space within which learners operate.

5.2.3.2 A(ii) Philosophy

Assessment of the ability to perform multiplication sums can be known and can be known extremely precisely. This can be deemed a considerable advantage of the model. The fact that it focuses so completely on a small sub-set of arithmetic skill allows for the hypothetical construct to become manifest much easier than with other domain-general tasks. Compare for just a moment generalised tasks such as mediated for during a Feuersteinian MLE session. Such constructs are nigh impossible to validate or define empirically due mainly to the scope which is covered. Harking back to chapter 3’s concerns over the need for psychology (read dynamic assessment as sub-field) to branch off in its own direction where such concerns can be readily dealt with in its own paradigm. Construct delineation and definition is far more accommodated within Gerber’s (2000) set-up. The author concedes that this issue has not been solved in this thesis but a tentative framework has been offered from which to view the issues. To venture an epistemological choice as to the model’s concern with reality and how this reality is knowable, instrumentalism could be a possible suggestion as, to paraphrase Madsen’s concern, truth is knowable only so far as it is pragmatically available. Likewise, a fine detailed construct can be isolated, mapped, assessed, even changed but can only extend as far as its own limits will allow. In other words, very little inference can be made on such narrowly defined constructs and this can be said to reveal the model’s greatest disadvantage but on the other hand, the model was not built for the purposes of generalising across cognitive tasks over different cultural contexts. Pragmatic truth is what is sought in this model after all. The more pragmatic the construct-seeking task the less convoluted the latent trait assessed for and vice versa. This aids in methodological workings of the construct definition and allows for “cleaner” interpretation of the empirical hypothesis as evidenced by the scores. A case can in fact be made for numerical assignation to these constructs although it is chiefly argued against in this thesis. Recall however, the consideration necessary for each case to be assessed on its merits. The need for numericising constructs is felt in certain applications even within dynamic assessment.

Interestingly, one can state that the nomothetic ideal is achieved through the idiographic ideal as “standardised individualism” is made possible. The research methodology employed is typical of that which can be said to have aided in psychology as a science (in as much as it can be referred to as a science). The method is replicable, scientifically deduced from known premises, entrenched in theory and verifiable. The process is valid and reliable, bar the notion of assisting the learner and this statement is made within the tenets of static assessment. The data language includes references to computational systems and mimics the research design. The nature of the data gathering methodology forms the design so to speak as the process revolves around these concepts necessary to carry out the procedure. In a way, it can be stated that the method dictates variables; i.e. use of computational systems forces the design to proceed in the manner it does. This is precisely the rationale behind delineation of the various strata that Madsen had in mind when he emphasised his concern for assessing these very issues behind models and theories. The intelligent tutoring system models three aspects in this system; the assessor (teacher); the learner and the domain tested but is obviously constrained in terms of how appropriate assessor responses can be, due to what Gerber (2000)



refers to as a lack of human agency. Modelling of human interaction is thus very limited, but this should not pose too much of a problem because, as stated above, the domain of enquiry is very narrow. Contingent computational responses can be viewed as the equivalent of human structured hinting or guided responses. Gerber (2000) acknowledges the fact that these systems can only satisfy the requirement of agency to a very limited degree. In this case, one can say that time indeed will tell what the future holds for these types of approaches or applications of testing.

5.2.3.3 B(i) Hypothetical terms

The role of hypothetical terms is particularly important within this model due to the nature of the context in which the assessment takes place; i.e. very specified domain. Seeking to operationalised hypothetical terms is achieved with relative ease because of the narrow state of empirical construct. Of the three hypothetical terms used by Madsen, constructive hypothetical terms would be the most fitting in terms of the model functioning. As stated in chapter 3, crude metaphors are often employed to detail and explain hypothetical functions and this is reflected in the approach towards multiplication for instance where the cognitive space is literally mapped onto the physical space. Cognitive functions such as memory can be reflected in the need to “carry over” a decimal to the next column or alternatively it could be stored in working memory without recourse to typing it in. The task of carrying over decimals is also a spatial task. Both process and structure terms are employed in a more concrete fashion than in some of the other models discussed in this chapter. “Dynamath is defined in terms of the processes required to compute or retrieve products of all simple multiplications involving numbers zero through nine” (Gerber, 2000, p.268) and refers to the retrieval dimension whilst the procedural dimension “refers to explicit or implicit mnemonic strategies to manage working memory demands during various stages or steps of problem solving” (p.268). The function of the hypothetical variable in this instance is one of directive activity which includes, almost exclusively, cognitive processes and structures.

5.2.3.4 B(ii) Scientific hypotheses

Scientific hypotheses are by nature testable and are necessary in order for a theory to advance within the nomological deductive model of how science is practised. The ever-present counter-argument to this notion is that there is as yet no consensus as to psychology’s place within the larger framework of what is considered a scientific approach. Humanistic models which tend to follow individually-based, relative and wholly qualitative approaches do not lend themselves to such verifiability and validity. Some approaches are simply not amenable to HQ determination in the first place. What would have been the case had the author of the study come across studies mostly reflective of the latter? There would be very few HQ’s generated and the incorrect application of a specific meta-theoretical framework. One can almost say that an *a priori* assumption was made based on the evidential literature which summarily pointed in the direction of testable models within dynamic assessment, including among others, Dynamath. The ontological classifications of scientific hypotheses can be said to follow an almost progressive development from existential to functional hypotheses as most such hypotheses proffer the tentative existence of a construct only later to be made manifest through function and structural terminology. Such as the analogous version of rendering cognitive “space” terms in terms of physical space.

Theoretical (hypothetical) scientific hypotheses

- H (information processing) - H (retrieval strategies based on procedural knowledge)
- H (information processing) - H (spatial mnemonics based on declarative knowledge)
- H (cognitive space) - H (working and permanent memory)
- H (physical space) - H (spatial representation of the problem)

Empirical (substantive) scientific hypotheses

- H (error detection) - S (physical manipulation of the cognitive space via visual prompts; relevant numbers and operation sign usage)
- H (error detection) - S (physical manipulation of the cognitive space via visual prompts; relevant numbers and operation sign usage in sequence)
- H (error detection) - S (physical manipulation of the cognitive space via visual prompts; change in layout)
- H (error detection) - S (physical manipulation of the cognitive space via visual prompts; movement of copies of numbers and operation signs from their original position in order to reduce the problem and response position)
- H (error detection) - S (physical manipulation of the cognitive space via visual prompts; as above but with the addition of filling in of correct response and movement of correct response back to original problem)
- H (errors of commission and omission) - R (nature of the processing fault underlying each type and category of response)
- H (errors of commission and omission) - R (application of the principle of ‘least prompt’; an attempt make the prompt as minimally intrusive as possible)
- H (errors of commission and omission) - R (analysis of relative prompt efficacy ‘scores’; how successful each prompt was in the correction of various errors)



- H (errors of commission and omission) - R (speed and accuracy)
- H (errors of commission and omission) - R (keystroke choices, response latencies, error making, self-correcting responses)

$$\Sigma (H-S) = 5; \Sigma (H-R) = 5; \Sigma (H-H) = 4$$

Hence HQ for *Dynomath* = $4/(5+5) = 0.4$. As can be seen from the above, this computer assisted model of dynamic assessment takes on a very concretised nature as the constructs are explicitly delineated within the narrow context of this domain-specific task. The fact that a software programme has been built and can thus be run limitless times, provides for a great opportunity for exact replication - a feature decried in the social sciences as hardly ever being invoked (chapter 3). The downside to this scenario is the obvious lack of more qualitative data derived from human contact sessions.

5.2.3.5 B(iii) Hypothesis system

Perhaps one of the more deductive models within this section, the explanatory system functions within a tightly knit framework of variable definition resulting in the ease with which nomological explanations can be tracked. "The program builds cells one at a time, in real time, as various student performance parameters are calculated. Assessment cells are made to vary systematically in the different cognitive demands they require for declarative and spatial-procedural knowledge" (Gerber, 2000, p.279). There is thus a direct link between the data and meta-strata concerns as the hypothesis system mediates between the two. The visual depiction of the number of prompts required illustrates the ZPD along two axes; deciles, which are constructed from responses on the entire spectrum of multiplication sums and category of problem difficulty of multiplication problems which could include single and multiple digits (and represents the actual problem space which is uniquely scaled to each learner and exemplified through retrieval difficulty). As the deciles construction will be unique to each individual, each profile will also look differently and thus present with a novel ZPD in terms of each cell's value. This makes the operationalisation of the ZPD incredibly specific and accurate. This is perhaps the best example of one of the most accurate depictions of this zone and represents "the invisible cognition underlying the visible behavioral record" (Gerber, 2000, p.290). The model thus chosen for this dynamic assessment is embedded within the computation model, so it can be stated that this is a model within a model.

There is no need to replicate the illustration here but it can be viewed in Gerber (2000, p.285). The ZPD can literally be broken down into fine detail where the exact nature of the multiplication problem can be identified in terms of the error made. This leads, however, to the inevitable question of the generalisability of such a domain-specific task. What can perhaps be transferred is the techniques involved in multiplication and if this is the original goal then far transfer or generalisability to broader domains is not necessarily feasible. One can possibly conclude then that there is a need for systematised, verifiable and accurate dynamic assessments and that such assessments take the form of domain-specific tasks. Can such models be readily accommodated in dynamic assessment within intelligence? It is unlikely that a substantial counter-argument can be leveled at this argument. Although the software is programmed and the algorithms followed are rule-based, one can hardly state that the nature of the data is difficult to grasp even though it is considered as abstract in Madsen's framework. The purpose of the model is thus to provide an explanation of the nature of how cognitively individuals learn and how learners might change their erroneous responses to multiplication sums and is less a physiological and mental model than an explanatory one.

5.2.3.6 C(i) Abstract data

The data can be considered abstract only in so far as the cells are the result of interaction effects between the different cognitive constructs utilised; namely retrieval strategies based on procedural knowledge and spatial mnemonics based on declarative knowledge. Also included are working and permanent memory and the spatial representation of the problem. These are systematically mapped onto three-dimensional space for the purposes of effective translation of the unique ZPD. The nature of the data is more concrete than abstract and the data present in more functional and descriptive ways than correlational.

5.2.3.7 C(ii) Concrete data

As stated above, the data are more concretised in this model than abstracted.

5.2.3.8 C(iii) Prime considerations

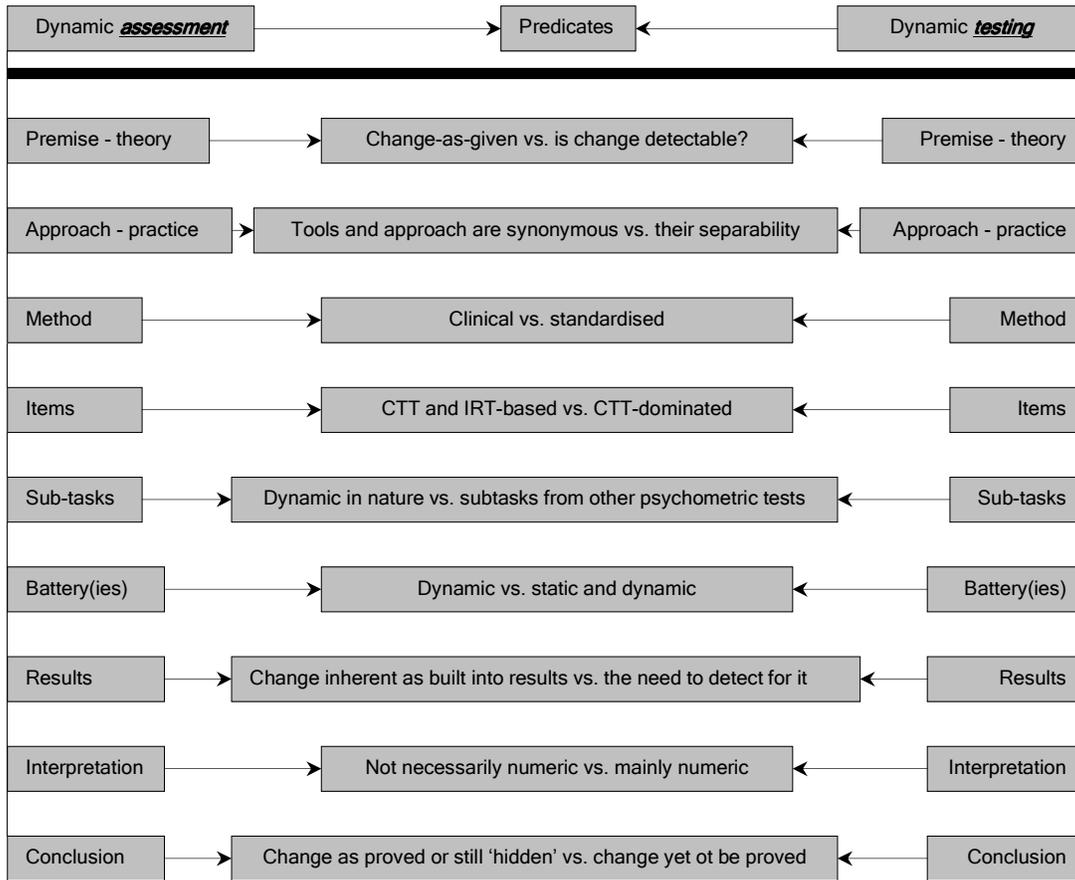
The meta-stratum ontology advocates the possibility and need to study cognitive universals which are not dependent on either cultural or social contexts. This immediately points to a standardised manner of assessment, which can prove both valid and reliable. Scores are ordinal-level representations of performance which illustrates the nature of statistical applications possible within such a model. Of particular note, is Gerber's reliance on rank-ordered results as test statistic when considering the nature of the measurement (or rather the limitations presented by ordinal measures). Interaction effects are also common when assessing for both declarative and procedural knowledge. Knowing what to do but not how and knowing how to do something but not what to do poses finer-grained problems which can nevertheless be dealt with in a manner befitting dynamic



assessment. Learners' responses are in turn treated as subprogrammes which vary in efficiency and precision and it is only the level at which the programme is set to tailor its approach that stops it from generating ever finer contingent responses after each problem is attempted. Hints at the outset are provided by default and as the programme progresses it becomes more individualized by "reprioritizing" the hint structure. Gerber (2000) refers to this as the level of granularity of the system. One of the more blunt statements in this section emanates from the issue of change score reliability but no attempt at a solution is offered. "How to conceptualize reliability of assessments under these circumstances [when students' performance is changed] remains problematic. These are matters of continuing concern" (Gerber, 2000, p.276). This does not seem to be overly problematic in this particular context. One of the main recommendations from this manner of dynamic assessment using intelligent tutoring systems is that as each item is iteratively produced (following on from responses to previous questions) the machine-learning approach, it is stated, can very well act as a repository or data mine for the generation of new rules that can be possibly utilised in future assessments. This is reminiscent of item response theory.

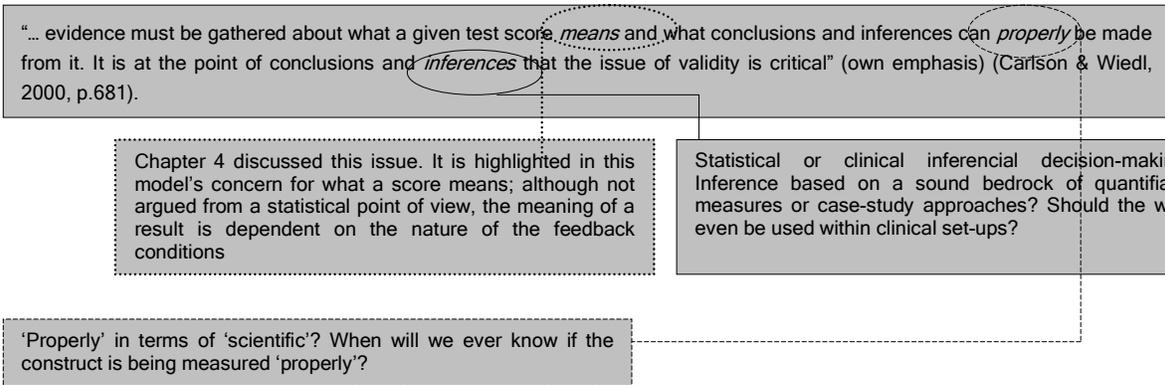
5.2.4 A model for dynamically assessing intelligence: differences in feedback

European contributions to dynamic assessment have been varied and heavily concentrated in efforts to establish for the domain of intelligence testing a semblance of feasible psychometric technique. In so doing, numerous approaches towards the dynamic assessment of individuals has emanated from the Netherlands and Germany in particular. The historical and political link between former East Germany and the old Soviet Union has been discussed in chapter 2. European work in dynamic assessment appears to offer the reader another glimpse into the field of assessment where the emphasis, or so it seems, lies more with dynamic assessment as technique towards testing than as the broad-brushed meta-concern. It is true that meta-concerns of various models do indeed encapsulate the dictum of change-as-given yet the fact that the method is referred to as dynamic testing and not assessment is one subtle indication of the concern with psychometric feasibility within the dynamic movement. The haggling over terminology may seem trite, but as always, an ever-present concern with governing meta-theory makes its presence felt even in terminology. Time, being as it is an equaliser of sorts, will most likely result in the gradual homogenisation of theories, tendencies and schools of thought to slowly accrue enough data on the subject matter thus effecting a less fractured test movement and bearing witness to a more streamlined technique calling itself dynamic. This remains to be seen. Dynamic assessment and dynamic testing can perhaps be summarised as follows as it pertain to various models' deployment of dynamic assessment and testing.



5.2.4.1 A(i) Ontology

How can an area of concern be studied if the said area of concern is not known? The job of theory is to proffer what can be assessed and the nature of the assessment seeks to find a way to effect just this; the 'how' of assessment (Campione, 1989). Dynamic assessment may not necessarily have moved the field of intelligence research measurably closer to a state of surety in construct delineation but availing the field of novel methodology has at least proven fruitful in terms of construct realignment. By varying the mode of feedback as intervention strategy, varying yields of intelligence assessments come to the fore. The system proposed is a dated one but nevertheless forms a model unto itself. This example if an instance of methodology highlighting hypothetical concerns. In changing formulation of enquiry into the hypothetical construct, one is in effect altering the construct as is typical of dynamic assessment methodology. One manner of attempting to assess for construct definition is by alternating the method of assessment in order to 'tweak' the outcome, in so doing the construct undergoes change. This issue is itself underscored by the varied concerns highlighted in this thesis thus far.



Comparing techniques within the feedback methodology allows researchers yet another angle at probing the elusive multifaceted diamond⁶ that is intelligence. Ontologically, it is assumed that intelligence as construct may not necessarily be available in a format that is methodologically accessible; but glimpses into some of its manifestations might be possible. Carlson and Wiedl (1992) posit that most research on intelligence is theory-driven where theory and measurement are treated as more than interchangeable and present as mutually instructive. This notion of 'mutually instructive' is questionable however in lieu of the aforementioned discussion in chapter 4 dealing with measurement theory within the social sciences. Does this 'mutually instructive' relationship carry with it an inherent tone of being beneficial? The answer cannot be affirmed as there is simply too much unaccounted for in social measurement for this to be the case. However, the authors take this as a premise upon which to leverage their taxonomy of feedback and investigate how this anchors research findings in dynamically assessing for intelligence. The authors view construct validity as originally endorsed by Cronbach and Meehl (1955) as being the unifying concept of validity but as has been noted in chapter 4, not all agree with this framework of validity endorsement. Recall the nomological net which provides the foundation for their version of construct validity - an essential tension perhaps clinically based and nomologically based interpretations of assessment? It is nevertheless their preferred framework from which to work. The model seeks to capitalise on what it considers to be its enhanced construct validation gained by filtering noncognitive performance-affecting factors from the assessment process by virtue of its feedback conditions. One could venture to state that this would logically seem to address the 'levelling of playing fields' for those individuals for whom noncognitive factors play a larger role in accounting for more variance in test scores. The motivation behind the need to have different feedback conditions is to explain the reasons behind successes and failures in different dynamic assessment assessments.

Basing their model on Vernon's (1962) broadening of Hebb's (1949) original classification of intelligence into categories A (biological intellectual endowment) and B (observable manifest behavioural intelligence) (see figure 15), intelligence C becomes the intelligence evidenced within intelligence tests. This is largely assumed to be resultant from genotype-phenotype interactions as intelligence tests are culturally-bound. It is intelligence C which is supposedly tapped in assessment and interventions and thus forms the link between intelligence as construct and the various dynamic assessment methods utilised to assess for it. "By making individual difference factors that affect performance evident, valuable information concerning potential interventions will be gained. This is especially significant for the special education and rehabilitation fields" (Carlson & Wiedl, 1992, p.169) and is echoed once again "comparison of performance observed under standard conditions with performance under conditions optimized according to hypotheses regarding parameters of administering the test can indicate, which factors may hinder the transmission of capability to performance" (Wiedl, May 2002). This model rests on an assumption of assessable intelligence but modifies the nature of such assessment by stating that the associated intelligence is only that which is observable in the form of tests specifically attuned to 'pick up' on such intelligent behaviour. This is unfortunately quite reminiscent of Boring's very dated definition of how intelligence is defined but it also highlights the progressive stance of being aware of the confines within which this type of intelligence is said to become manifest and which is measurable to a small degree. Intelligence B, the authors state can however be estimated via means open to dynamic assessment, especially when techniques can be changed to suit the nature of the task and level of intellectual functioning of the individual. The model is a reflective one of sorts as it seeks to modify the individual via modification of the test method through varying conditions (C₁-C₆). The modes of assessment feedback are detailed below (Carlson & Wiedl, 1978, p.560; 1992, pp.170-171).

- C₁ Standard procedure
- C₂ Verbalisation during and after solution: requires the child to describe the main stimulus pattern prior to searching for the correct answer and then, after a particular alternative is chosen, to explain why he/she made that choice
- C₃ Verbalisation after solution: involves the child describing the reasons for his/her choice after the choice is made
- C₄ Simple feedback: where the child is informed after the choice has been made whether or not it was correct
- C₅ Elaborated feedback: involves, in addition to simple feedback, an elaboration by the test administrator of the reasons why the chosen answer was correct or incorrect; the principles involved in the task are pointed out
- C₆ Elaborated feedback plus verbalisation during and after solution: is a combination of conditions C₂ and C₅; it involves verbalisation (description of the pattern to be completed) and after a choice has been made, includes the child's explanation for the reasons for the solution chosen and elaborated feedback by the tester, informing the child of the correctness of the response and explaining the principles involved in the task

Perhaps the most salient result to emanate from this type of research is that what may appear to be advantageous to individuals in terms of elaborated feedback and verbalisation for instance in fact does not necessarily work that way in practice. Due to developmental trajectories, different age groups react differently to the various techniques utilised to elicit correct behavioural outcomes. A strategy that may work for a ten year old may not be efficacious with a six year old, purely due to the developmental 'lag' evidenced at the younger age. What this is saying then, is that not only do varying degrees of feedback work in manners befitting different ages but that dynamically assessing for constructs too work in varying ways with age groups and static IQ

⁶ Thanks to George Murphy for this catchy phrase.

groups. Retarded individuals also perform along a varying continuum where different strategies have differential outcomes. From their work it has transpired that testing procedures involving verbalisation or elaborated feedback are more effective in increasing intelligence test performance. There is no one set policy guideline that can be followed when assessment for retarded individuals is considered and dynamic assessment is perfectly placed to accommodate for these differences. Its inherent mandate has as its core an understanding of unequal starting positions in terms of genetic endowment and ecological background. “we suggest that individual differences resulting in variation due to testing procedures exists *within* each designated group suggesting the necessity of applying appropriate testing techniques or procedures to uncover such variation” (Carlson & Wiedl, 1978, p.563).

The authors' model is more a practical implementation of an understanding of the limitations that govern the possible assessment of a watered-down construct as opposed to a deeply aligned philosophy espousing affiliations to either the materialist, spiritualist or dualist conceptions of humankind. However, Madsen's 'human freedom of action' meta-thesis is evident in the above table. Expressing interest in change-based performance and seeking to increase such performance via means other than typically found in static-based assessment attests to the tenacity with which this taxonomy of feedback is structured. Although as mentioned above this method is somewhat dated, its prescient concern with non-cognitive aspects involved in test performance is enlightening. Behavioural aspects such as impulsivity and anxiety account for a substantial amount of variance in test performance and increase in results (see section 2.8.7 where this was discussed). This idea comes across strongest in the work of Feuerstein and colleagues. Most dynamic assessment models owe allegiance to other earlier models or applications and hence overlap in ontological concern is usually evident. The increase in performance due to decreased anxiety for instance leads one to ponder how static-based intelligence test scores are negatively impacted upon especially within the younger population groups. The taxonomy of feedback proves worthy in the form of diagnostic tool where frequently correct answers are arrived at in an incorrect manner leading to at-times overestimations of results on tests. The diagnostic value of dynamic assessment in general plays forth in its role as clinical assessment technique.

Closely aligned to this approach towards differential assessment via different feedback mechanisms is the crucial difference between change in score brought about by learning during intervention and subsequently forgotten (poor transfer) and learning which does not take place at all (Wiedl, 2002); or so it seems upon first glance. The scenario looks as follows:

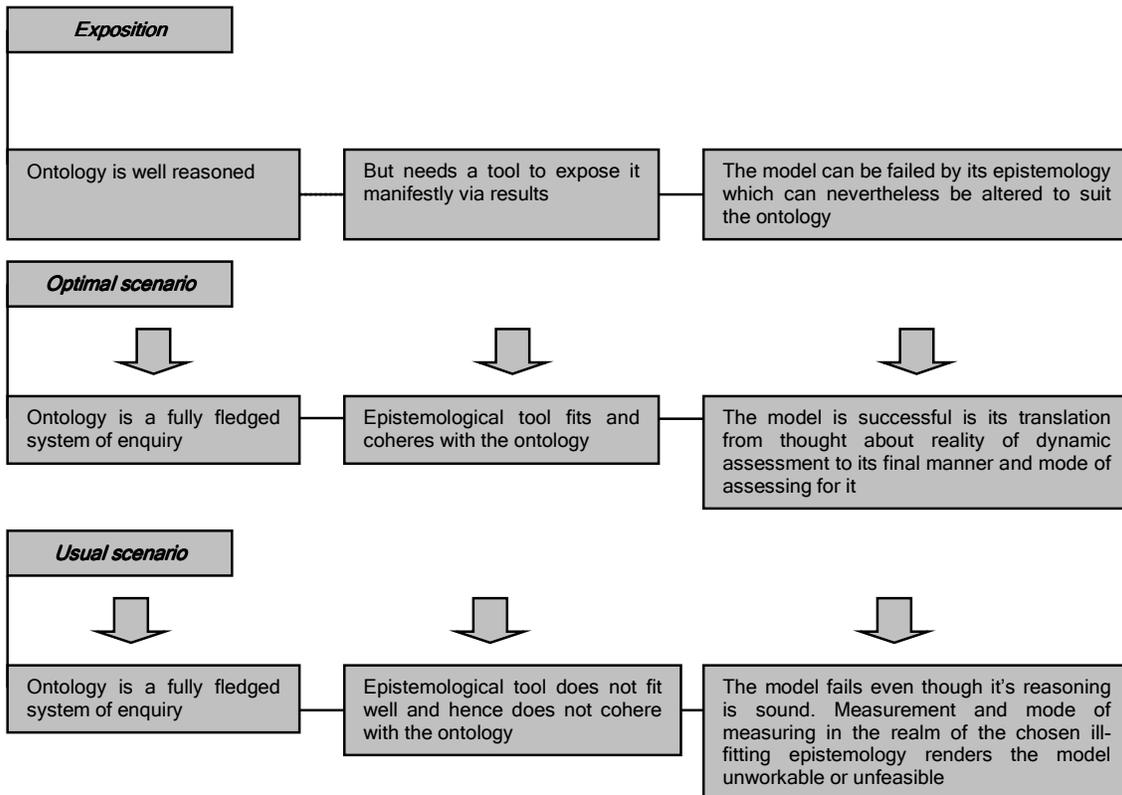
Static-based assessment	Dynamic assessment	Dynamic assessment with various feedback conditions
<p><i>High scores:</i></p> <ul style="list-style-type: none"> tentative evidence of high level of observable construct possible false construal of low level of observable construct (correct answer but wrong method used to reach answer) no explanation as to why this is the case 	<p><i>High change scores:</i></p> <ul style="list-style-type: none"> tentative evidence of learning having taken place transfer accounted for by change in observable construct no explanation as to why this is the case 	<p><i>High change scores:</i></p> <ul style="list-style-type: none"> evidence of learning having taken place - more assured than with generic dynamic assessment more accurate identification of which intervention process was more efficacious in bringing about change and transfer (near and far) explanation more forthcoming as to why this may be the case
<p><i>Low scores:</i></p> <ul style="list-style-type: none"> tentative evidence of low level of observable construct possible false construal of high level of observable construct (incorrect answer but partially correct method used to reach answer) no explanation as to why this is the case 	<p><i>Low change scores:</i></p> <ul style="list-style-type: none"> tentative evidence of little or no learning having taken place little or no transfer accounted for by little or no change in observable construct no explanation as to why this is the case 	<p><i>Low change scores:</i></p> <ul style="list-style-type: none"> evidence of little or no learning having taken place - more assured than with generic dynamic assessment more accurate identification as to why certain interventions were not as efficacious as others in various sample groups
<p><i>Learning:</i></p> <ul style="list-style-type: none"> not applicable 	<p><i>Learning:</i></p> <ul style="list-style-type: none"> equated with 'modification' 	<p><i>Learning:</i></p> <ul style="list-style-type: none"> identified as resulting from genuine modification or resulting from context-dependent intervention referred to as 'compensation'

Compensation:

- key ingredients necessary for performance are missing from the individual who then compensates for this deficiency by substitution - known as prosthetic compensation
- capabilities are latent but blocked for other reasons and then become activated - known as catalytic compensation⁷
- differs in nature from modification
 - changes in certain sub-populations such as retarded individuals may not be detectable in learning situations but may be detectable in situations catering for supportive facilities such as compensation
 - individuals incorrectly identified as non-learners may in fact be mislabelled as such. Given the right feedback, these individuals may evidence learning (only to be identified as such through strategies found in the right-hand column)
 - prediction of future behaviour can only be made in lieu of the specific type of intervention (feedback) used (referred to by Wiedl as 'ecological validity')
- "The differentiation of different modes of intervention thus is a necessary requirement for a conceptual framework or model of [dynamic testing]" (Wiedl, May 2002) because "for a variety of reasons [a] student's performance may or may not reflect what we think it does and the conclusions we draw. Accordingly, the inferences we make may be inaccurate and misleading and we have a validity problem" (Carlson & Wiedl, 2000, p.681). This 'validity problem' is the hypothetical-empirical construct debacle which consumes much of this thesis's debates.

5.2.4.2 A(ii) Philosophy

Change is inherent in any measurement of a process. This assumption is not to be found within static-based assessment. Ontologically, this model not only recognises that change is inherent but seeks to further delimit it by eliciting it via various feedback mechanisms. The model's ontology, one can state, presupposes the epistemology: finer-grained assessment of change elicitation can be effected via practical strategies in modes of assessment. Carlson and Wiedl's concerns surrounding the validity of dynamic assessment hinges around core epistemological issues. How to assess for that which is ontologically feasible. Epistemology as tool for ontological concerns needs to be cognisant of how results are effected and the results, as discussed in chapter 4, are redolent of the tool used to get to them. In other words, results are only as good as the tool utilised. This is seen as follows:



⁷ It really does seem to be the case that psychology is mired in natural science terminological frameworks. However, the terms are apt in what they seek to convey.

Understanding the differential effect of assessment practices on sub-samples within dynamic assessment populations allows for finer granularity of change-based measures. The model takes into account, assesses for and mediates change within varying levels of performance, personality and task characteristics. The use of verbalisation within assessment scenarios which has proven to be invaluable as aid in reducing impulsive behaviour is almost identical to the LLT model (section 5.2.1). The usual manifestation in terms of improved test performance results from the increased focus on task strategy within the test set-up where individuals are forced to understand why they take the decisions they do. This is a direct link to metacognitive strategy use which is one attempt at bridging the ontological and epistemological gap: change can be induced and assessed (ontology) and one such tool is to engage with higher order thinking skills which trickle back to the individual's engagement with their own resources and knowledge thereof (epistemology). This offers insight into the model's conceptualisation of the nature of the existence of reality and how we can know it as such and which the author proffers as instrumentalist. An objective truth exists but is only knowable in so far as the instrument used can attest to this which is where the wrangle surrounding fit between ontology and epistemology enters the discourse. The fact that thought processes are seemingly restructured confirms the notion of thoughts being amenable to such change in the first place. This is the ontological premise. The epistemological tool used to account for this in an instrument is metacognitive strategy use via data thesis concerns such as verbalisation.

Regarding the model's epistemological theses and its relation to the ideals of science practice, the authors' main contention throughout their discussion on the model is the nature of validity which by definition is encompassed in a nomological framework. Predictive, criterion, content and construct validity can only be assessed in a nomological framework as such a framework is inherently structured to ensure lawfulness. So, although positioned to assess for within an idiographic ideal, nomothetic concerns are well aired. As is typical of most models assessed for in this section, both nomological and idiographic models are utilised in a complimentary fashion. This is a given in most instances as dynamic assessment models need to straddle the abyss of psychometric feasibility and individual focus; an issue woven throughout this thesis. The authors support the notion that both theory and pragmatics are closely aligned and mutually beneficial but the extent to which static-based assessments feed back information from assessments into future assessment is questioned. Static based tests are assessed for robustness and psychometric properties whereas dynamic assessment tests and accompanying methodology is also assessed for robustness and psychometric properties. Administration results are of importance to the former whereas administration methods as well as results are of importance to the latter. In a simplistic rendering then, one can state that the latter goes 'further' than the former and can thus be said to offer more in terms of veracity of assessment. Another important statement from the authors of the model is that within the nomological framework certain conclusions emanate from studies evidencing that predictive validity for instance differs across sub-populations groups which echoes the notion of intelligence and learning potential being considered as non-unitary in nature. This poses an interesting argument:

- **Argument 1**
 - Intelligence is similar to learning potential or is learning potential
 - Intelligence is unitary
 - Therefore learning potential is unitary
- **Argument 2**
 - Intelligence is similar to learning potential or is learning potential
 - Intelligence is not unitary
 - Therefore learning potential is not unitary
- **Argument 3**
 - Intelligence is not similar to learning potential as the former is a construct whereas the latter is a method
 - Argument stops here
- **Issues from arguments 1 and 2**
 - Learning potential has been packaged by most models as a method towards the extraction of 'extra' information from individuals
 - As a method various techniques can be utilised to extract this information
 - Some models (although decidedly in the minority) have sought to view learning potential as a construct
 - As a construct learning potential has still only been envisioned as a unitary concept with very little literature discussing it as non-unitary
 - What does this mean for future endeavours within dynamic assessment as multimodal construct?

5.2.4.3 B(i) Hypothetical terms

The model is made of all three Madsenian hypothetical terminology types which serve to bring coherence to the system of observables. From operational definition status to one of hypothetical link, the latter seeks to bridge the data aspect of the former with the larger meta stratum governing the system. From initial concept to raw data concern, the hypothetical terminology is often surrounded by more meaning than that which is evidenced at the data level (Madsen's 'surplus meaning'). This section detailing hypothetical terms cannot ignore the theoretical model on intelligence envisaged by the authors and will follow after the assessment model's exposition. The following hypothetical terms as per the model for assessment are delimited below:



- a. Mentalistic hypothetical terms (H_m)
 - i. Personal factors - general of categorical structures; heuristic structures; componential activities
 - ii. Diagnostic approaches - modification; compensation; restructuring of thought processes
- b. Organismic hypothetical terms (H_o)
 - i. Personal factors - reflective processes; impulsive processes; systematisation
 - ii. Diagnostics approaches - inhibition; overt verbalisation; thinking out loud; impulsive
- c. Constructive hypothetical terms (H_c)
 - i. Task characteristics - differential testing approaches (more grounded and hence more reflective of a data stratum aspect than a hypothetical term but can fit in here as the mode of testing which affects relationships between predictor and criterion variables); catalytic compensation; prosthetic compensation; dual coding of information - recall that constructive hypothetical terms also provide the model with crude analogised versions of how the explanatory system works

What is perhaps most highlighted in the above is the nature of the validity of the model in terms of searching for a unidimensional construct epitomising validity which is upheld by the sentiment expressed by the authors “the search for a unitary coefficient for traditional psychometric measures or for learning tests is elusive” (Carlson & Wiedl, 2000, p.690). The authors’ model of intelligence encompasses hypothetical terminology and aids in prying apart what is meant by the construct of intelligence and features heavily in their attempts to determine the validity of dynamic assessment as a whole. The authors seek to address intelligence in all its manifestations and to do this requires a broad-based model of intelligence. The model’s mandate is to at once:

- (1) Be generalisable
- (2) Be integrative
- (3) Take cognisance of culture
- (4) Enhance cognitive education as well as informing assessment practices in general

Interestingly, the model acknowledges across-the-board contributions of biological intelligence correlates, g -based understandings of intelligence, generalised cognitive structures (which are akin to the tools utilised within dynamic assessments) and specific cognitive structures which are linked to the generalised cognitive structures. Recall in previous discussions that the more encompassing the model the less likely it is to be testable rendering itself less robust and in the HQ system it will undoubtedly receive a higher than average score (less testable). The further one moves away from generalised structures to specificities the more likely it will be that the construct validity will improve which it does. This is reminiscent of the move towards curriculum-based dynamic assessment. This however entails a move away from generalised cognitive structures. Is there a middle ground? Working memory (as evidenced in other models in this chapter) forms a central aspect of research into static-based and dynamic assessments and is linked to g by way of variance accounted for. The intelligence model employed by the authors can be said to present as structure terms functioning in a directive capacity which is not unexpected as most intelligence models are by their nature inherently composed of structure terms. The model encompasses biological correlates of intelligence at its core, followed by g -type markers such as neuronal efficiency, followed by generalised structures (Vygotskian tools) and specific structures all housed within the particular cultural context in which intelligence functions. Of note in their discussion on biological bases of intelligence, the authors note that it is unlikely that such biologically primary abilities will feature much in dynamic assessments due to the unrefined nature at which dynamic assessment currently operates; i.e. assessments and interventions are a far cry from measuring neuronal efficiency and the statement should be understood within this context. However this need not obviate the need for dynamic assessment even at this level in the future.

5.2.4.4 B(ii) Scientific hypotheses

Theoretical (hypothetical) scientific hypotheses

- H (cognition) - H (basic mental abilities)
- H (cognition) - H (procedural characteristics - planning and strategic behaviour)
- H (cognition) - H (orientation variables including task-specific orientation and test anxiety)

Empirical (substantive) scientific hypotheses

- a H (levels of task requirements) - S (material for practical activity)
- a H (levels of task requirements) - S (material for indirect perception)
- a H (levels of task requirements) - S (material for linguistic-conceptual knowledge)
- a H (diagnostic approaches) - R (modification - implicit/explicit)
- a H (diagnostic approaches) - R (modification - predetermined/self-determined)
- a H (diagnostic approaches) - R (compensation - prosthetic)
- a H (diagnostic approaches) - R (compensation - catalytic)

a H (diagnostic approaches) - R (inhibition - complicating factors)

$$\Sigma (H-S) = 3; \Sigma (H-R) = 5; \Sigma (H-H) = 3$$

Hence HQ for the Carlson and Wiedl's model = $3/(3+5) = 0.375$ which results in the model's testability.

5.2.4.5 B(iii) Hypothesis system

Mediating and bridging the gap between the data and meta-strata lies the hypothesis system which in the case of this model is exemplified by its own model or framework which serves as conceptual schema for analysing test performance and change based on varying feedback protocols. The model houses the scientific hypotheses seen above but also serves as a coherent model for the approach as a whole. Madsen views the hypothesis system as consisting of deductive explanatory systems as well as model explanations within the nomothetic explanatory system and seeing as the authors base their model on natural science dictates to a certain extent, it is surmised that their conceptual schema falls under the rubric of nomothetic explanation as opposed to hermeneutical and idiographic systems (although an argument can be made for the inclusion of the latter notion as tailored feedback exists to ensure greatest efficacy of change-based evaluations). Although not precisely the same, Carlson and Wiedl's (1992; 2000) schema is reminiscent of Feuerstein's (2002) concern for the testing experience in totality. The latter highlights eight issues, included among others are basic assumptions regarding the nature of intelligence, the nature of changes made during assessment or intervention and the structure of tasks utilised to assess for modifiability. In similar vein, Carlson and Wiedl state that they do not consider test performance to be necessarily indicative of latent ability (which has already been noted in chapter 4) but that performance should rather be conceptualised "as a result of a dynamic interaction between the individual, the test materials, and the test situation" (1992, p.181). The conceptual scheme is delineated below (Carlson & Wiedl, 1992, p.181; 2000, p.683). A model, state Carlson and Wiedl (1992), should meet three criteria, namely, they should:

- offer an explanation of how various solutions' strategies towards solutions vary according to the test tasks
- be complex enough to explain how both cognitive and non-cognitive factors contribute to test performance
- avail of multiple suggestions as to the best interventions to be used in order to optimise cognitive functioning (this of course depends on the strategy chosen for feedback for instance)

I Task characteristics	
<i>Levels</i>	<ul style="list-style-type: none"> • practical activity • direct perception • indirect perception • linguistic conceptual knowledge
II Personal factors	
<i>Structures</i>	<ul style="list-style-type: none"> • epistemic • heuristic
<i>Components</i>	<ul style="list-style-type: none"> • basic abilities (relationships, seriation, classification) • procedural characteristics (planning, flexibility, self-regulation, activation) • orientation variables (task-specific orientation, self-concept, situation conceptualisation)
<i>Processes</i>	<ul style="list-style-type: none"> • analysis • synthesis
III Diagnostic approaches	
<i>Modification</i>	<ul style="list-style-type: none"> • implicit/explicit • predetermined /self-determined
<i>Compensation</i>	<ul style="list-style-type: none"> • prosthetic • catalytic
<i>Inhibition</i>	<ul style="list-style-type: none"> • complicating factors

5.2.4.6 C(i) Abstract data

The intelligence model utilised for explanatory purposes offers information pertinent to the abstract data stratum which become concretised measurements. Biological correlates, general cognitive and specific cognitive structures are instances of this type of data. This level is where the data are collected within the framework of the conceptual scheme for test performance and change.

5.2.4.7 C(ii) Concrete data

Construct validation hinges around accepted statistical practice methods but traverse a greater distance in terms of bringing to dynamic assessment more consistency of approach and end-result. Although not wishing to engage in constructionist views of assessment, the author highlights the support that Carlson and Wiedl give to sentiments surrounding the use of tests and how all

facets of the testing procedure impact on eventual outcome and how they do not merely reflect their utility value in terms of results only. Concrete data play a vital role in their validation model and is explored below in terms of its “losing” and “gaining” information depending on the level at which constructs are assessed and expanded upon; theoretical and empirical.

5.2.4.8 C(iii) Prime considerations

This section is prefaced by the following citation which is warranted on the basis of bolstering the main concern within this thesis; that of measurement feasibility:

“Dynamic assessors are particularly interested in constructs such as intelligence, learning potential, motivation, cognitive modifiability and the like. Unfortunately many constructs in the social sciences in general and psychology in particular are problematic¹ as a variety of operational definitions are often used for the same or very similar theoretical construct². Although ambiguity can be reduced when theoretical constructs are operationally defined³, measurement alone leads to its own ambiguities of explanation and prediction if it is independent of theoretical context⁴ (Carlson & Wiedl, 2000, p.706).

1. This whole notion has been discussed at length in chapter 3 especially section 3.4
2. The concepts “intelligence” and “dynamic assessment” being a case in point. ‘Working memory’ is of central concern to intelligence measurement as is ‘inductive reasoning’. Whether the placement of the two can be put alongside an equals sign remains to be validated, although much research does point in this direction. The most problematic would be the confusion of the concepts ‘intelligence’ and ‘dynamic assessment’ as the two are not functionally equivalent. Notice the authors’ cognisance of the difference between the substantive and theoretical construct which attests to the ideal realm of the theoretical concept but acknowledges that the bridge between this realm and reality as is assessed for can be quite distinct
3. Operational definition takes away from all that is encompassed in the notion of the original theoretical notion and although may be able to retain a substantial amount of what is meant to be conveyed by the theory will never be able to fully represent what is inherent in the theoretical idea or construct. This is the nature of measurement and is discussed at length in chapter 4. In terms of the authors’ model explicating the validation process within dynamic assessment, the two activities prior to the derivation of the theoretical construct salute the psychological enterprise but also highlight its ‘scientific standing’ within the greater sphere of science; namely observation and previous research and theory. These two activities inform the theoretical construct and in keeping with Madsen’s data stratum-meta-stratum model, both levels interact to yield information necessary and pertinent to the governing meta-theory and the data gathered during assessment. The validation model is not only a model showcasing validity issues but tackles the very mechanism which *is* the scientific enterprise⁸
4. This sentiment hardly needs more embellishment. Dynamic assessment practitioners are saddled with a two-fold problem of assessing for change in an area riddled with lack of consistency, even though strides are being made.

The validation model encompassing structural validity judgements is housed within a nomological net as stated by the authors which lends credence to the testability quotient received above (even though the validation model itself did not receive an HQ score). Seeking generalisability from their model, the authors acknowledge the roles played by path analysis, structural models and IRT as aiding in the linkages between theoretical and empirical constructs within validity issues. Referring to what they call “consequential validity” the authors also make mention of the need for the utility value of a measure which is a move away from a concern with basic research to one of practical value. They view this concern as paramount in a world which sorely necessitates alternative modes of enquiry especially in areas such as assessment. “With its focus on applicability, the ultimate goal for dynamic assessment is consequential validation. This requires a clear and scientifically defensible justification for the constructs developed, testing procedures employed and assessment devices used” (Carlson & Wiedl, 2000, p.708). In fact reference is made to dynamic assessment as a science warranting a technology of its own. The authors’ ideas concerning validation within a nomological net are revealing of the sentiments conveyed by Cronbach and Meehl (1955) which has been controversial in some quarters (see chapter 3 and 4).

5.2.5 Dynamic assessment of the level of internalisation of elementary school children’s problem-solving activity

As the title suggests, this dynamic process reflects an approach and not a theoretically developed model regarding construct definition. As was cautioned at the start of this chapter, dynamic approaches are not amenable to HQ determination in the same way that models via empirical results are. There are far fewer H-S, H-R variables and as such conclusions will be overwhelmingly biased in favour of H-H variable consideration. This is not necessarily a “bad” thing just not practicable nor is it

⁸ The model is not duplicated here as it merely highlights aspect number 3, but the reader can view it in Carlson and Wiedl, 2000, p.707.



comparable with other HQ's. However just because some dynamic assessment procedures are approaches and not construct definitions does not in any way result in their not being scrutinised within this framework. After all, this is just the point about dynamic assessment: it's status as model and approach. It spans the boundary of construct enhancing mechanism to approach within other construct-making or construct enhancing endeavours. Such approaches then, cannot be ignored due to their failure to fit within the model.

The approach taken by the authors Karpov and Gindis (2000) represents many researchers' concern with dynamically altering static-based assessment techniques in order for such tests to reveal hidden aspects of functioning. It is not always the case that dynamic assessment picks up what static based assessment cannot in terms of underestimating ability but it can also underestimate what has been overestimated by static based assessment. Working diagnostically is after all one of the dynamic assessment mandates and in this regard it usually performs well. The premise within this approach is the underlying importance of three levels of analogical reasoning within problem solving. Analogical reasoning as is well known within mainstream intelligence assessment is a construct which is located in many such intelligence batteries. Applying analogical reasoning across domains yields information about a child's level of performance in Piagetian terms and in order to aid in cross-domain application of algorithms learned in one domain to the next, Vygotskian notions of levels of next development are invoked as support for the contention that movement across domains is possible given the right circumstances beyond those offered by mainstream assessment techniques. Typically children solve problems firstly in the visual motor domain, then in the visual imagery domain and lastly in the visual symbolic domain or level. Approaching the solution of tasks in a dynamic manner allows for better understanding of how and why children perform poorly on statically presented formats.

5.2.5.1 A(i) Ontology

The human is viewed through the lens of obtainable achievement at least as it can be assessed for in the typically understood fashion of intelligent functioning. Other test batteries and models have been premised on inductive reasoning and problem solving via working memory for instance. The construct itself, analogical reasoning, is not challenged, the nature of its change however is what is recorded. The construct itself exists *a priori* but the facility with which it can be utilised can change given the right circumstances. Such change is evidenced through the alteration of qualitative manifest behavioural attributes. The rule-based system of analogical reasoning can be taught and assessed which of course brings into question the veracity of the construct as evidence of innate unchangeable intelligence in the first place. The authors invoke evidence suggesting the linear move through varying levels or domains of analogical reasoning which it is assumed is sequentially based and can in a manner attest to the level of next development as understood within the Vygotskian framework. Two core concerns within this model of assessment include the highest level of understanding that is achieved and the highest level at which a new problem is solved given the understanding of the algorithm learned and applied. Solving analogical problems at the visual motor level, visual imagery level and lastly at the visual symbolic level illustrates the Piagetian influence and taking this into consideration it is hardly surprising then that those children who evidenced the greatest success with transfer tasks were those who had mastered analogical reasoning at the symbolic level. The dynamic process utilised in this study was centred around a learning situation thereby becoming fused with the learning paradigm so telling of dynamic assessment. The assessment takes the form of a typical pre-posttest scenario with dynamic assessment as interspersed between the two. Each level of analogical reasoning is assessed for and dynamically re-administered if not understood and in a way follows the usual pattern seen in these types of tests; i.e. if step 1 is answered correctly, proceed to step 2; if not repeat step 1 and so on.

The approach emphasises qualitative variables such as the identification of non-cognitive skills that are not picked up by conventional assessments. Behavioural aspects include self-regulation, impulsivity, distractibility, difficulty in following directions, self regulation and can be housed under the umbrella term of meta-cognition - an important non-cognitive attribute discussed in chapter 2. Such non-cognitive attributes regulate what can be sourced in mainstream assessment batteries but cannot be identified as such. In other words, mainstream assessments can highlight poor results but do not highlight the reasons therefore which, when dynamically assessed for, can be identified. Vygotskian notions of shared responsibility for learning in the early stages and later relegated to the learner also play a major role in how this assessment proceeds. The ontological responsibility thus, can be said to reside within the accepted tenet that poor performance usually has reasons behind it other than the conventional reasons attributed to low level IQ and the like. Probing deeper in the matter often results in the identification of such non-cognitive and qualitatively ascertained behaviour patterns.

5.2.5.2 A(ii) Philosophy

The approach towards the individual as understood within a learning environment attests to the social concern in assessment. The fact that assessors are co-opted as teachers counts in the favour of the approach's idiographic emphasis. Cognition is seen as developing within a socially based reality and not as being manifested in a sterile testing situation. This is proven to be the case where mainstream assessment yields poor performance without identifying the reasons behind such performance. By engaging the learner in the process and still keeping to standards as required by mainstream assessment tools and also encompassing constructs still proven to be robust (analogical reasoning through various levels) the dynamic assessment component offers explanations as to why performance is poor in the first place and seeks to redress these problem areas;



something anathema to mainstream assessment. The data language is specific to an approach-based manner of assessment as opposed to a test battery application which reviews behaviour as primary determinant of performance and not necessarily construct application. Ontologically speaking, it is assumed that previously hidden behaviour can be identified and epistemologically that it can be brought to the fore via behaviour analyses which is elicited through a process of gradual learning of algorithms. In this case it is the solving of analogical problems through different levels of difficulty without abandoning currently accepted notions of what it means to assess for aspects of intelligent functioning. The approach is a clear instance of how the prevailing model directly impacts on the outcome. Assuming change is possible within the confines of what can realistically be expected, behaviour modification becomes a highlighted issue which in the case-studies the authors cite, results in upward modifications of results. Shared meaning-making and steady apportionment of responsibility from teacher (assessor) to learner (testee) underlies the tie to Vygotskian ideas on collaborative attempts at entering the next level of development. Private speech is also recommended as strategy for furthering the development goals of the learner.

By simply changing the assessment criteria and environment, completely altered pictures start to emerge. Likewise it is not always the case (although by far the most apparent) that static based assessments underestimate certain skills as they can in fact identify certain skills as normal when in fact dynamic assessment can show that this is not the case. In other words it can at times account for why mainstream assessments yield normal values but are unable to account for poor school-based performance. Hence the diagnostic utility of dynamic assessments. Nomothetic ideals are nevertheless upheld in this approach emphasising the individual. This is quite representative of dynamic assessment models as a whole with the constant interplay between individual diagnosis and remediation tailored to idiographic ideals, at least this is what is usually aimed for in the research design. This idiographic pull can be seen in the still prevalent data language utilised. For instance, this is evident in the differentiation between three supposedly successive levels of analogical reasoning ability onto which a linear trajectory of progress is mapped in the hope of obtaining quantifiable data so as to either vindicate or dismantle (as the case may be) presuppositions ingrained in both static and dynamic assessment methods, or in this case, approach. It is warranted at this stage to highlight the differences and commonalities between dynamic assessment as construct-changing method and dynamic assessment as approach which can endeavour to change the construct being measured. Note that for the sake of the argument, it is assumed that whatever the construct encompasses, it can be measured with an instrument which can detect measurable entities.

A

Dynamic assessment as offering constructs (new or derived concepts)

In which dynamic assessment is characterised by its novel contribution to innovative construct creation which can be accomplished in one of two ways (or a combination):

- i. Taking an already-existing mainstream construct (typically the approach taken in by far the majority of cases) and adapting it via change-based assessment techniques which are either outdated (pre-post test scenario) or modern in conception (as offered by modern test theory)
- ii. Creating a novel construct (not often seen) of which there is a landmark instance; “learning potential”, however
 - a. Learning potential can be characterised as a new construct or
 - b. It can be considered a manipulated statistic garnered from change-based assessment techniques which leads us to ask:
 1. Which is it? And if the two are different, how do we control for it?
 2. Creating a novel construct allows for measurement as with any other known construct. This allows for it to be defined and treated as a separate variable and is thus bequeathed empirical status via the hypothesised construct

B

Dynamic assessment as offering method (approach)

This division (B) is similar to the bullet (b) in the first division (A) as it can also be understood to represent a manipulated statistic but the emphasis is on approach more so than what can be accomplished with a number. Usually, but not exclusively, this division concerns itself with studies revolving around qualitative behavioural changes that can be altered when administered static-constructs in a dynamic fashion. No new construct is being used although it is often assumed that due to the change in behaviour a new construct becomes manifest. This is possibly erroneous. Administering methodologically altered constructs does not necessarily imply intrinsically altered constructs. This is open to debate and the last word, it is assured, has not been voiced. Nevertheless, this division can serve as point of departure for more intensive discussions around this issue. The approach as discussed in this example (5.2.5) is just one instance of many approaches pervading the literature, especially articles. Chapters in edited volumes and books offer what can be considered division A-like conditions but articles necessarily of shorter length prefer the characteristics of division B. Is this due to how science is practised? Or how social sciences seek conformity to the nomological-deductive method (see chapter 3) or is it due to a concerted effort to move the discipline along the lines of what is considered to be the ‘best’? One cannot move away from these preoccupations regardless of the palatability of the arguments

This distinction is almost invariably never explicitly stated (if ever) even though it may be implicitly implied. The reigning blurred conception of dynamic assessment as technique or construct-innovator is a part of the larger problem surrounding this domain within intelligence research. This is the author’s thesis given the models discussed in this chapter. Is dynamic assessment contributing anything new by way of construct creation or is it yielding to science a methodology which can quite easily tie in with the currently accepted intelligence field? Intelligence research can itself be construed as offering a two-fold purpose within psychology; a technique (static-based) and novel constructs (working memory, inductive reasoning and so on; although these are now quite dated, a “true” discovery of innate characteristic features cannot by definition be dated unless the entire scheme and purview of intelligence research is completely overhauled). One must take cognisance of the fact that we can only divide mainstream assessment into similar categories such as techniques and novel constructs because of the existence of dynamic assessment in the first place. There would never have been a category “static” had dynamic assessment not offered an alternative! In a manner then, this could be a false distinction to make but it has been made for us; an *a posteriori* categorisation which is perhaps unfair to mainstream assessment.

5.2.5.3 B(i) Hypothetical terms

Keeping the above in mind, it can be seen how implicit assumptions mandate the methodology followed. In changing the testing situation to one of dynamic assessment, the approach towards assessment is altered but not necessarily the construct. Madsen employs hypothetical terms as mode of explaining or bringing to light those terms which cannot easily be identified. The approach of the authors in this regard highlights a number of organismic hypothetical terms more so than mentalistic and constructive terms although these too are employed. Process terms fit in with the abundance of organismic hypothetical terms and provide tentative explanations of the changes undergone by the individuals. For instance, terms such as impulsivity, high distractibility, difficulty following directions, self-regulation, self-planning, self-evaluating, self-monitoring and self-checking are indicative of organismic hypothetical terms covered in the model. Process terms would indicate the involvement of the assessor in the mediating environment and include monitoring, intervention, joint activity, regulating, verbal tools, external speech and shared performance. These terms indicate the goals towards which the interventions strive. More often than not, directive variables exceed the number of dynamic variables (within the hypothetical term stratum) within dynamic assessment models as the case usually presents itself in such a way that conventional terminology is used as it pertains to mainstream intelligence constructs. However, when approaches are studied as in this case, more dynamic variables come to the fore. Such variable usage refers to less tangible constructs (if they can even be considered as such) which are not as amenable to measurement as are directive terms.

5.2.5.4 B(ii) Scientific hypotheses

Scientifically attuned hypotheses can be made testable due to their nature as concrete instantiations of hypothesised variable as given above. Meaning and depth (quality of meaning) is partially lost when translated to testable hypotheses but such is the nature of measurement when not all variables can be similarly measured. Existential hypotheses include the move along a progression of analogical reasoning abilities starting with visual-motor problems and culminating in visual-imagery problems. Concretising these constructs into sub-tests allows for functional hypotheses as explanations of mechanisms involved in the relations between such variables. The manner in which one traverses the three levels of analogical reasoning assumes a linear progression towards mastery of the problems. By involving a dynamic assessment approach the levels can traversed more successfully or least can provide a deeper understanding of the processes involved in this progression. It is anticipated that the HQ score for this approach will be among the higher scores in general purely because of the number of H-H hypotheses involved in relation to H-S and H-S hypotheses.

Theoretical (hypothetical) scientific hypotheses

- H (learning potential/intelligence) - H (analogical reasoning as exemplified through the visual motor reasoning modality)
- H (learning potential/intelligence) - H (analogical reasoning as exemplified through visual imagery reasoning modality)
- H (learning potential/intelligence) - H (analogical reasoning as exemplified through visual symbolic reasoning modality)
- H (learning potential for analogical reasoning as deduced via) - H (self-regulation; metacognitive skill) which can be broken down into the following sub-categories:
 - H (learning potential) - H(impulsivity)
 - H (learning potential) - H(high distractibility)
 - H (learning potential) - H(difficulty with following directions)
 - H (learning potential) - H(self-planning)
 - H (learning potential) - H(self-monitoring)
 - H (learning potential) - H(self-checking)
 - H (learning potential) - H(self-evaluating)

Note that due to the sub-categorisation of the self-regulation hypothesis, only half a mark will be accorded the minor H-H hypotheses

Empirical (substantive) scientific hypotheses

- H (learning potential) - S (teaching of algorithm, evaluation of performance, performance at different levels of analogical reasoning)
- H (learning potential) - S (shared performance, external speech, monitoring and remedial intervention)
- H (learning potential) - R (three performance levels assessed; visual-motor, visual-imagery and visual-symbolic)
- H (learning potential) - R (changes in self-regulation)

$$\Sigma (H-S) = 2; \Sigma (H-R) = 2; \Sigma (H-H) = 7.5$$

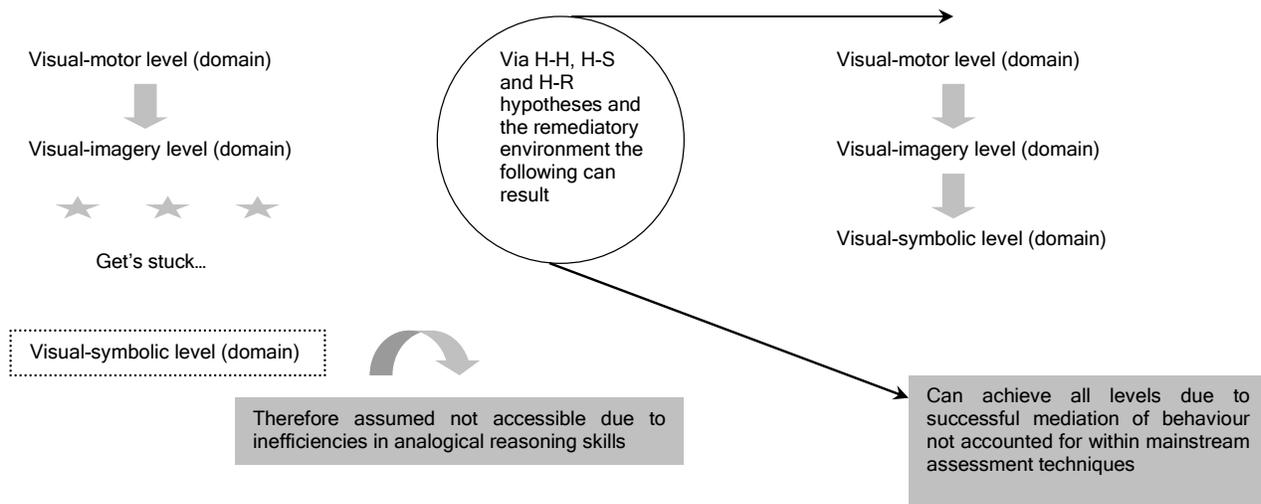
Hence HQ for the approach towards assessing for analogical reasoning = $7.5/(2+2) = 1.88$ evidencing that the further one moves away from conventional assessment practices the less likely the model is testable. Compare this HQ result to those HQ results emanating from empirically based groundings. The increased difficulty faced by such models (approaches) in terms of verification and/or falsification leads one back to the tenets derived from and for scientific models as practised within science (both natural and social) (see chapter 3). This does not bode well for such approaches in terms of the framework espoused by such concerns which is not to say that such approaches need be relegated to such frameworks; hence the continued plea for these areas to remove themselves from the mainstream framework and become useful within another framework. One can tentatively state that the more testable the model, the more conventional it is and the less testable the less mainstream the model. Moreover, the approach based manner or division as discussed above lends itself more to such untestability of approach than the division espousing construct creation. A new construct is by nature more amenable to testing than is an approach which can at times be considered amorphous. Thus, this *approach* is not as testable as other *models* (note the difference).

5.2.5.5 B(iii) Hypothesis system

The model does an admirable job in accounting for its explanatory system in terms of utilising mainstream analogical reasoning as foundation. Theory proposes the development of how the process unfolds which adds credence to the model. Initial constructs are theory-based and validated in empirical studies after which the constructs are imported into the model or approach. The deductive argument might proceed as follows:

- Analogical reasoning across three modalities follows a progression of developmental feats which
- if properly learned results in the progress from one mode to another but
- this is sometimes hindered by complicating factors such as non-cognitive behavioural variables which
- can for the most part be detected by dynamic assessment approaches and can result in
- improved indicators of performance within the mainstream realm of intelligence assessment, thus
- offering more by way of understanding what and why attributes have resulted in under-estimates of performance or purely underperformance per se

A possible model depiction can be offered as follows:



5.2.5.6 C(i) Abstract data

The data are explained away in mediatory terms, especially terminology emphasising meta-cognition, a plethora of Vygotskian influenced notions and ideas as well as structured aims throughout the intervention and proceeds along a classical pre-test and intervention mode. The dependent and independent variables in this instance are exemplified through functional relations and not correlations.

5.2.5.7 C(ii) Concrete data

Concrete data are more ephemeral but become concretised through the manner of initial choice of variables to be assessed for. Assessing constructs as traditionally accepted but endeavouring for change allows two things in this model: the nature of intangible behaviours to contribute to the result after mediation and allowing for these behaviours to become measurable via the process of analogical reasoning ability increasing over time. The ability as such is not perhaps altered according to strict



classical interpretations of g for instance, but what is being controlled is the non-cognitive aspects which militate against the manifestation of the correct levels of cognitive functioning as omnipresent yet not manifest. The approach elicits full functioning by skirting around what happens to be blocking it at the time before intervention. It reveals what has been hitherto concealed. The main conclusion from this argument then is that a dynamic intervention (with no new construct of its own) can offer an unlocking mechanism by breaking away what hinders better performance; in this case behavioural attributes. It does not offer anything new by way of novel construct. Only an approach. This loops back to the figure above where this is differentiated.

5.2.5.8 C(iii) Prime considerations

The model is communicated through a process and so one infers its attributes from what transpires during interventions. The use of classical test statistics is kept to a minimum but in a way whether or not such tests are utilised is almost of no concern as they do not in anyway impinge on the process itself. The original constructs assessed for are no doubt infused within statistical workings but as discussed above the main focal point is not the creation of new constructs nor the assessment of these constructs but how change of assessment can yield better performance. Single studies employ statistical levels of chance to designate which learners have indeed performed at higher levels in a significant manner and of those obtaining certain levels (highest level obtained) who among them are able to successfully negotiate novel tasks at the achieved level (highest level obtained through successful completion of a sub-test). Statistical data are obtained “after the fact” and as such provide no new information beyond that which is already known. As one reads the literature in this regard, it becomes increasingly evident that statistical data illustrating significant changes or not do not add value to the approach as the approach centers on behavioural changes brought about by changes in metacognitive strategies. However, in defense, one could state that prior to dynamic assessment a score resulted in x and at post-intervention the score resulted in y . X differed from y in a significant fashion thereby leading one down a path of elimination. If y is significantly better than x and the same test strategy is used it must be due to the intervention which targeted behavioural features. Statistical calculations can avail themselves of the use of statistical analyses but perhaps one should re-look the level at which the data are analysed (incidentally, the authors do utilise non-parametric test statistics).

5.2.6 The learning potential test for ethnic minorities (LEM)

The Flynn effect, although not mentioned by Hessels (2000) can be tentatively said to be working among the ethnic minorities in the Netherlands where successive generations of ethnic minorities have evidenced an increase in terms of scores on conventional assessments yet there remains a substantial gap between indigenous and ethnic groups across broad measures. As with many instances across the world, misclassification is problematic due to the supposed low level of functioning at which many students are assumed to perform due to incorrect classification as cognitively deficient. The cycle is self-perpetuating as these individuals perform poorly specifically because of the low expectancy within the school environs. What is most evident about the rationale behind the test's development is the similarity of construct being assessed for, namely a better estimate of general intelligence. This is an important point to highlight, as it is intelligence that is being assessed for and not learning potential. Hence the construct remains the same but the purpose is to provide a more reasonable estimate of intelligence via the endeavour of assessing dynamically. The “learning potential test” can thus be misleading in terms of its initial rationale. Perhaps this is yet another instance of the cause of confusion within dynamic assessment in the broader intelligence field. For, if the test is said to assess “learning potential” but in fact measures “intelligence” then one is lead to believe that the two constructs are synonymous which they may not be.⁹ An expected retort could well be that a better estimate of intelligence is garnered via dynamically assessing for it in other words methodology is invoked to assess for a similar construct and that the method is not the construct. This is conceded. Note the use of dynamic assessment in some instances as method and in others as construct: can it function as both? This is problematic and is an issue which has not been solidly thought through resulting in statements and labelling of tests and constructs as opposed to having kept these concerns as separate issues altogether. Standardising feedback offers a method of comparability even though the feedback is non-verbal in nature. The test is lacking in qualitative error-detection as is evident in other similarly structured batteries such as the ACIL and LLT discussed above and there is no hierarchical arrangement of hints and probes. There are also no time limits imposed on the testee.

The premise upon which this test is based is the measurable construct of “inductive reasoning” yet another popular mainstream intelligence construct. No pure dynamic construct has been used in tests thus far, at least not as far as this author is aware. This brings one back firmly into the territory of the existence of pure dynamic construct - the learning potential paradigm is a method more so than a new construct but this message does not always come across directly and often one is left wondering about this very aspect. Dynamic assessment as method cannot replace “intelligence” as construct but it can however replace intelligence testing and herein lies the crux! Unfortunately, and rather odd, is Hessels' construct validation process typically seen in many

⁹ Note that no-one knows for sure. That is partly the reason given by the author in her contention of confusion of terminology within this domain. Murphy (2002) noted this as a point of concern in South African research in dynamic assessment where batteries partly referred to as “potential” were in fact static-based conceptualisations of intelligence.

such instances; correlations between IQ scores and the LEM. It is odd because Hessels is cited in section 4.4.2.3. as supporting the notion of not substantiating the hypothetical construct with similarly undefined constructs, although one could argue that inductive reasoning as construct is merely being transferred to a new context in which it is assessed for in a dynamic manner. As is common with other test batteries which assess dynamically, the LEM is able to highlight what Hessels refers to as “false negatives” those who are falsely categorised as low intellectual performers when in fact they are underestimated by mainstream assessments.

5.2.6.1 A(i) Ontology

The guiding philosophy within this model as exemplified through the battery is the acknowledgement that change can aid in re-classification of previously misclassified learners who happen to be grouped as ethnic minorities. The philosophy is not entrenched in maintaining the mistaken notion that ethnic minorities are all unfairly discriminated against merely because they happen to find themselves within another culture, nor does it suggest that the label assumes *a priori* that deficiency within standardised learning situations represents maladaptive inherent traits. The test format goes to great lengths in accommodating cultural (or what is perceived to be cultural differences) by gesturing the nature of incorrect answers. The author nevertheless appreciates the current climate and its pervasive influence regarding the suitability of using static-based intelligence constructs as this is the core of the LEM as assessed for dynamically.

5.2.6.2 A(ii) Philosophy

There is both nomological and idiographic concern even though the test is directed at a specific sub-population. Cognitive functioning is assumed a universal process which is simply not elicited in the correct fashion within standardised conditions. People, whether of minority groupings or not are assumed to possess differing abilities regardless of prevailing circumstance. The choice of inductive reasoning as construct and its subsequent manifestation within the various items assessed show allegiance to mainstream assessment tenets in terms of choice of sub-tests. The data language represents common parlance and is interpretable within both static and dynamic type set-ups.

5.2.6.3 B(i) Hypothetical terms

The LEM as with other tests serves a diagnostic function and is thus practically aligned in its determination of this goal. The author cannot locate hypothetical terms beyond those detailed below, namely learning potential and the H-S terms which can be classified under the rubric of directive variables.

5.2.6.4 B(ii) Scientific hypotheses

Returning again to the litmus test of testability within the Madsenian framework it is immediately apparent that substantive hypotheses dominate thus yielding an outcome all too familiar within research oriented models. The unique feature about the response based hypothesis is the nature through which behavioural outcomes are elicited, namely through non-verbal feedback and gestures.

Theoretical (hypothetical) scientific hypotheses

- H (inductive reasoning) - H (intelligence)
- H (change in inductive reasoning) - H (potential)

Empirical (substantive) scientific hypotheses

- H (inductive reasoning) - S (classification)
- H (inductive reasoning) - S (word-object association recognition)
- H (inductive reasoning) - S (word-object association naming)
- H (inductive reasoning) - S (number series)
- H (inductive reasoning) - S (syllable recall)
- H (inductive reasoning) - S (figurative analogies)
- H (inductive reasoning) - S (verbal abilities)
- H (inductive reasoning) -S (memory span) - akin to “working memory” utilised as governing hypothetical construct in the Swanson-CPT
- H (potential) - R (change as based on the particular phase where assistance is needed; different quantifiable phases which are dealt with exclusively in a non-verbal manner)

$$\Sigma (H-S) = 8; \Sigma (H-R) = 1; \Sigma (H-H) = 2$$

Hence HQ for the S-CPT = $2/(8+1) = 0.22$ making this a testable model. A trend which can already be seen is the preponderance of H-S/H-S hypotheses and the comparative scarcity of H-H constructs which in the social sciences domain is quite interesting as it shows that the field is veering towards a nomological-based framework within which to work whereas a large section of the social sciences is given to non-testable theories and models. All the more reason to heed the call made throughout this thesis for dynamic assessment within intelligence to either split off or completely divorce itself from the larger domain of psychology and discover for itself its own rich niche within which to work or to further integrate in a manner befitting the dominant trends in the social sciences. Constantly attempting to bridge the gap in a delicate balancing act only inhibits further progress.

5.2.6.5 B(iii) Hypothesis system

The closest that the author can come to describing the deductive state of the model is by viewing it through the lens of standardised testing situations as this is what it most closely represents. If inductive reasoning yields a reasonable index of intelligent functioning, then more leeway can be accommodated in terms of doing so dynamically in the hope of yielding an equally reasonable index of learning potential. The model is more empirically constructed and in a manner of speaking can be said to fulfil its dynamic agenda on the basis of initial standardised concern which is known to consist of empirical indicators of intelligent functioning.

5.2.6.6 C(i) Abstract data

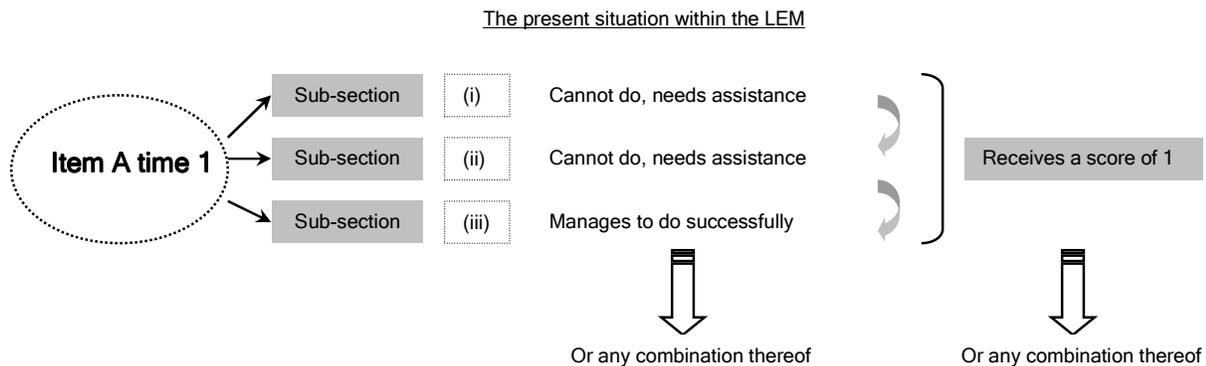
The abstract data theses are already accounted for in the model's description. Recall the abstract data's relevance in terms of its enmeshment with empirical relations.

5.2.6.7 C(ii) Concrete data

As encountered during the non-verbal feedback and relating specifically to the level at which assistance is required. These are quantifiable.

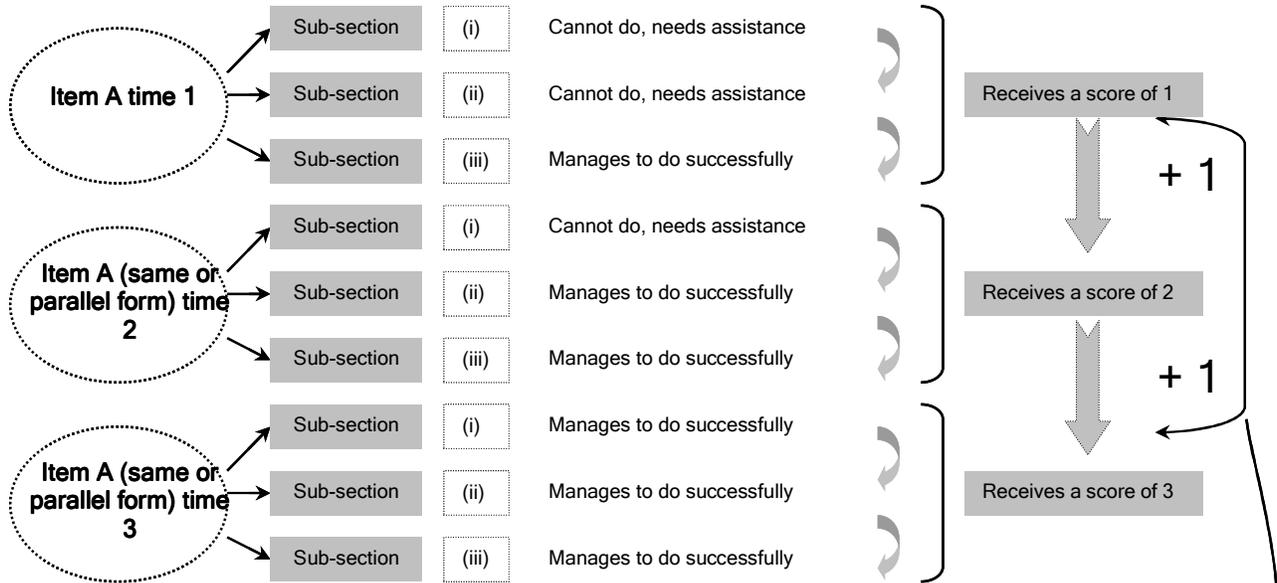
5.2.6.8 C(iii) Prime considerations

Hessels (2000) skilfully avoids the pitfall of change assessment by skirting around it altogether. The LEM offers no pretest or training as interspersed between the conventional pre and posttest scenario. The quantifiable attribute occurs during very particular phases of training where numbers with different designated values are attributed to performance. What would have added value to the LEM in terms of its change attributions is the rate of change taking place. The following illustrates what is meant by this.



Given the present situation then, the score is indicative of the learning potential at one point in time only. Yes, it may well yield a potential score of sorts, resulting from a process-based assessment style. But what of future attempts and these respective patterns of scores? What might they yield in terms of the *rate* of change which is yet another perhaps finer discriminator and indicator of change? A closer look at how rates can be assessed is explored below.

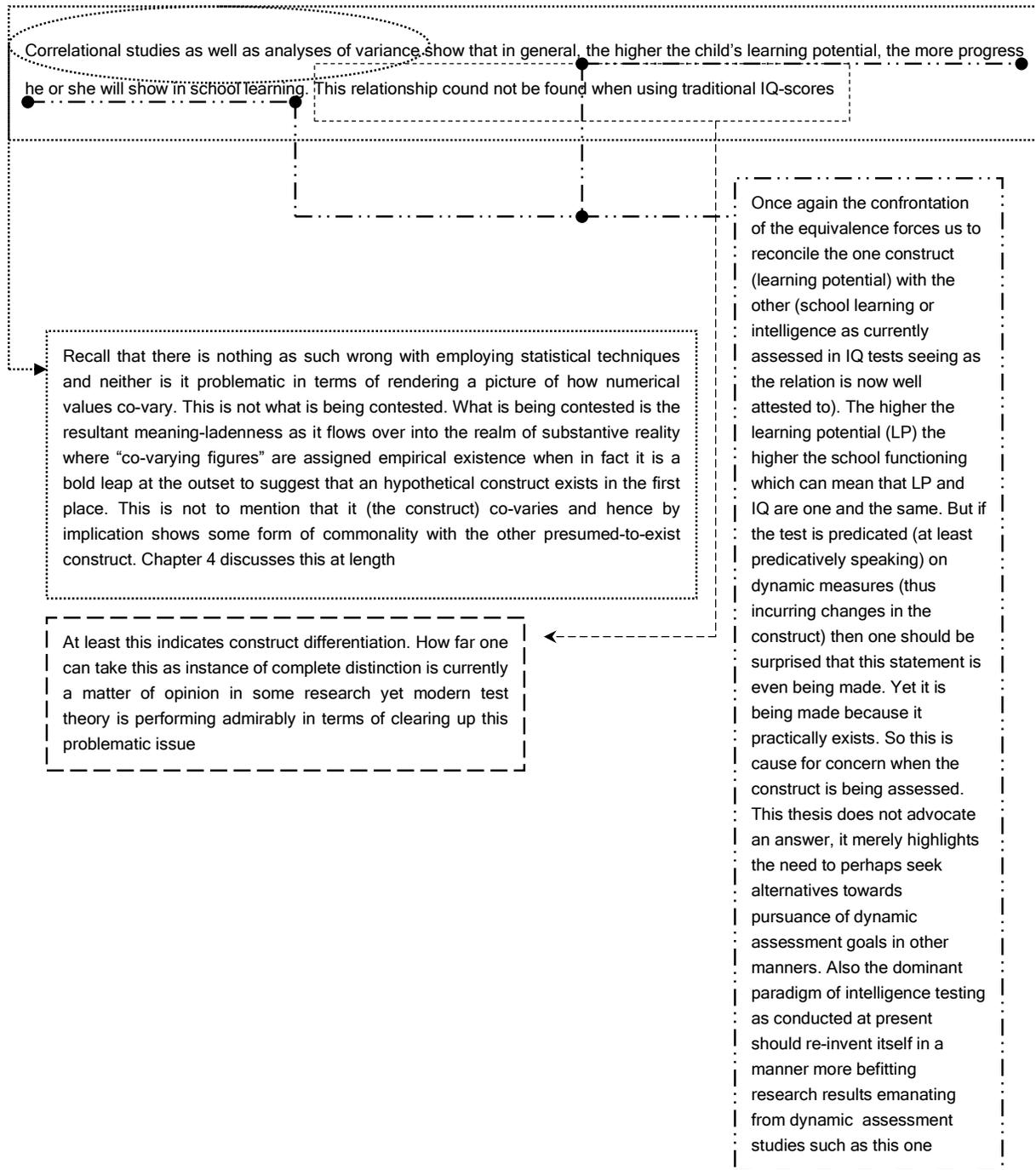
A possible situation within the LEM



That change occurs is apparent but the rate at which this change takes place is of diagnostic value as well and offers a new metric and possibly represents an underlying new latent trait to watch out for. Over time, the change has been (in this particular example) consistent with improvement. Various scenarios may present during which the reverse may happen evidencing something else beyond learning taking place (such as luck for instance). At present only Item A at time 1 is utilised with its specified scores and pattern of results

Recall that not only does this take into account progressive levels but increasing difficulties of these items. This would suit IRT particularly well

The more assistance that is needed the lower the score thus higher scores evidence greater learning potential. What can be added to this with the above in mind, is that the more consistent the rate of change the greater the learning potential. Cronbach alphas are very high resulting from the homogeneity of items; they can be said to be assessing for inductive reasoning. Of interest to this model is the battery's concern with predictive validity utilising another more dynamically attuned type of assessment (teachers' ratings of achievement which is being increasingly utilised as measure and incorporated as variable into predictive studies) thus allowing, for once, prediction based on dynamic predicates. In this way, the construct has changed which is really very much the point in the first place. The nature of the construct becomes highlighted once again when Hessels (2000, p. 127) states that



5.2.7 Swanson-cognitive processing test (S-CPT)

This test is premised upon a construct supposedly defining a key aspect within intelligent functioning; namely information processing and in particular working memory (WM). Working and researching in the field of intelligence is truly an adventure as so many divergent opinions are held by an equally diverse erudite populace whose main concern is that their understanding of what intelligence is will be placed as central to their evaluative attempts. The premise of tests/batteries/models is what influences all else to come in the process of test conception, construction and eventual deployment. No matter how well the process ensues, one is left with the core around which all else revolves and if this core is understood to be firmly entrenched in scientific methods of research then it is assumed that the battery upholds this view throughout. As is now very evident, different views concerning intelligence dictate different stances taken by different researchers. This links back in bold fashion to the discussions that have taken place throughout this thesis. Nevertheless, this battery as with others will now be looked at in similar fashion utilising the same framework. In keeping with modern trends the need to construct a test which is workable in mainstream construals is made clear from the outset of this model's views surrounding the importance of ensuring validity via mainstream instances such as academic and language tasks, problem solving and mathematical ability. The hypothetical construct of "working memory" is not to be confused with the concept of short-term memory and as has already been discussed in chapter 2, working memory and g are related. To what extent then, will dynamic assessment of working memory yield indices of change if global ability is taken to remain stable? Once again, another contentious debate rears its untamed head. Swanson (2000, p.72) is quick to substantiate his conception of the linkage by stating "given that WM is an appropriate construct for a standardized test of dynamic assessment (DA), the test attempts to provide an *approximate index of processing potential*" (original emphasis). Note that Swanson now indicates an entirely new hypothetical construct; that of processing potential and not learning potential. But if the rationale behind this is that "processing" and "learning" are one and the same then his argument can stand (for the time being). In essence the author is not entirely convinced of the veracity of his claim to merely "dynamicise" constructs simply because it is considered timely to do so. Can one really employ a construct and add the word "dynamic" to it and assume that the latent construct now reveals something entirely different? This is, firstly, not scientifically plausible, and moreover results in completely different conceptions for what is now being assessed. The counter argument to this is that dynamic assessment is not a test itself, but a method of assessment. Yet modern test theory shows particular concern with how latent constructs become manifest when there was originally none to be found. Changing an item (not to speak of a construct) by means of changing the assessment process is very much a radical change in the nature of what is being assessed.

The S-CPT pays particular attention to the problematic aspect of transference or maintainability. How well will derived potential be maintained and sustained? As highlighted in chapter 2, g -based assessments do not attest to transference very well if at all, with many programmes having failed to maintain far transference of learned behaviour. What does this mean for potential behaviour? Potential is ascertained via a testing-the-limits approach in a standardised manner. What in fact occurs within this model is the cuing of correct strategies to be used when completing items. Training for better strategy use and attempting to locate potential is quite a leap indeed. The efforts at individuating the procedure seem, at least on the surface, to be superficial attempts at making the process look dynamic, by way of highlighting the individual-in-context when in fact this might not be necessarily the case. The test's commendable features include diagnostic information (a positive aspect and one that is clearly recognisable as indicating some concern with dynamically assessing in broad terms) by assessing how information is forgotten. The manner of assessment via probing and analysis of errors made is reminiscent of a number of dynamic tools surveyed thus far. The nature of errors made, indicate something beyond the scant information derived from the fact that an error per se was made. Swanson acknowledges the difficulty surrounding the measurement of a hypothetical construct, namely "cognitive modifiability" couched in mainstream terms and acquiesces to the common understanding of this index attesting to potential; i.e. the fewer hints required the greater the potential. This strikes a deep chord of discontent within the author however. The fewer the hints required can also attest to the greater ability in general of the testee, or greater intelligence. Thus one is in fact indirectly stating that potential is intelligence; which one does not want to do in the first place. The argument given in support of this contention, is that it is not the number of hints per se which is indicative of potential but the rate of decreasing hints required. This also links back to a common metric, if it can be referred to as such, within dynamic assessment; latency or time taken to complete items or time taken to complete the entire test. The author invokes many intelligence researchers' comments on the link between speed and intelligence and adaptability and intelligence. Can one make the link to potential as well? If $a = b$ and $a = c$, then surely $b = c$ ¹⁰? Something to ponder.

¹⁰ The author awaits the flood of criticism that may well come her way regarding the "flawed use of logical methods" within the domain of a soft science. "The tools are not correct" they may add, and so her argument may well be suffused within irrelevant characterisations of what it means to practice a science. This of course leads one to consider the veracity of the claim that dynamic assessment within intelligence can be considered a science.



Seven scores are captured within the testing situation, all supposedly given to indicating “cognitive modifiability” and include:

- The initial score - an estimate of processing ability, which is stated as being similarly interpreted to an IQ score or as Swanson states, Spearman’s g
- The gain score - the highest score obtainable given the hints and probes (after accommodating for initial differences on scores due to initial uneven familiarity with items). When testees are buffered under ideal conditions, this score can be said to be more representative of “true” ability
- The score attributable to the number of prompts required - referred to as instructional efficiency
- The maintenance score reflecting the level at which information is processed without recourse to more prompting
- A difference score - literally the difference between the gain and initial scores which Swanson refers to as the Vygotskian notion of proximal development (is this not too simplistic a rendering of what was initially conveyed, not to mention that the ZPD was not really emphasised in Vygotsky’s writings; see chapter 2)
- The processing stability score which reflects the difference between initial and maintenance scores. These PDI scores run into some measurement trouble though. Consider the issues raised in chapter 4 concerning quantification and what it means to assign a number (or numeral) to an attribute or score or construct. The metric involved becomes meaningless if one is unable to attest to the quantity of the construct being assessed. A difference of “2” in pre-post test scenario and the difference of “4” cannot be said to be to be representative of interval or ratio measures at all, yet the scores are treated accordingly at just such scales. The entire discipline of psychology is at fault when it comes to this issue and not only the S-CPT. Amazingly Swanson acknowledges this but does not present a convincing case as to how best to proceed on this basis of the inequality of scoring, especially as ratios become meaningless. Hence, all the more need to reaffirm psychology’s present positioning along the lines of quantification. This is a point in case and is just one battery that has fallen into this quantification trap. Lastly,
- The strategy efficiency score resulting from testees’ declarative knowledge of processing prior to assessment

These scores are reflective of more traditional understandings of what is meant by “change” and does not seem to fit the bill of dynamically assessed conceptions of what change encompasses. The test has been standardised according to various criteria. The test has evidenced factor loadings on various key dimensions utilising factor analyses and has been shown to correlate with standardised measures of working memory as understood to exist as construct within mainstream assessments. Can one truly attest to the S-CPT’s dynamic status? The use of seven manipulated scores is as yet a far cry from identifying latent constructs. Construct-related validity may well be evident in terms of its comparability to traditional measures, but perhaps the time has come to validate dynamic assessment tests (not methods, but models) with other dynamic assessments. Hopefully the first such test in a linkage will not have been validated by static-based measures. One might well be able to provide a unique baseline measure if this direction is pursued. As is to be expected the traditional scores as stated above evidence good correlations with constructs such as “intelligence” and “achievement” but evidence poor correlations with change-based scores. This is to be expected and has become rather mundane in terms of what new information is gleaned from the test, other than the fact that it is not measuring traditional constructs but rather something else. Is that “something lese” potential or merely deviations from the norm? Hopefully future data resulting from correlations will not litter the field with “negative” findings. This is already known. We need to move onwards toward techniques which will enable explanations as to what in fact these negative correlations bespeak of their hidden constructs. Unlike some dynamic assessment techniques which are more dynamically attuned, the S-CPT evidences high Cronbach alpha’s which is not necessarily a good thing! Is the test truly accommodating change? The test does however come into its own when its diagnostic utility is sourced thus enabling the differentiation between true learning disabled students and those incorrectly classified as such. However, does a test battery need to be “dynamic” in order to do this? There is nothing wrong with employing static-based measures to do the job if that is their function. Why the need to pursue dynamic assessment as alternative in order to do what can already be done?

Problematic for this thesis, the S-CPT nevertheless prides itself on the following:

- Its focus on cognitive activities as opposed to mental abilities (information processing vs. information possession) which are two constructs which can be said to reflect different aspects of possibly one unifying underlying construct, namely g . This is perhaps an unfair criticism lodged at the S-CPT as it can be lodged at any number of other batteries accounting for very much the same thing. There is no consensus about the positioning of two constructs (assuming for the moment that are not unitary to begin with). That the tasks generalise across school-type subjects hardly makes the case more compelling
- The claim that the S-CPT offers an alternative measure of potential to traditional IQ is not necessarily well founded given that “measures of potential” can often be any one of a multitude of measures. The S-CPT may well yield an alternative measure per se, but whether it reflects potential as such is questionable. The S-CPT measures are framed within mainstream ideas of statistical renderings of factors supposedly evidencing constructs in a manner which can supposedly be dynamically assessed

- The third claim is more realistic in outlook and can stand up to scrutiny; that of the tests' link between assessment and instruction. The engagement of administrator is obviously radically different to the usual process of tester-testee engagement.

5.2.7.1 A(i) Ontology

The nature of how information from the world is processed is of main concern to this model. Its initial endeavours in the field of intelligence assessment have suited its ambit well, but the move towards process-based assessment of what the model can quite comfortably accommodate in its static form is perhaps less well-founded. The human being as agent of change can hardly be said to be the central concern here. That change can later be brought about is affirmed but not well argued and construct blurriness does not aid in its strategies of measuring dynamically either. Being wholly dependent upon psychometrically derived cognitive activities makes for a superficial attempt at potential assessment however well intentioned it may be. The underlying assumption is not that change can be assessed for but that the processing of information can be assessed for and in all likelihood the model does this admirably well having been substantially normed and tested for reliability and validity in its traditional form. An important comment within this thesis that is still upheld is that static-based conceptions of testing need not be done away with nor need they be sustained merely because they attest to reliable and valid indices. The need for such batteries/models to be continually utilised is confirmed by the veracity of their claims especially as it pertains to modern-day assessment and constructs assessed for in industrialised nations. The need to fulfil the dictates of dynamic assessment because information processing needs to be assessed for in a manner befitting the detection of potential is not sound and neither is it convincing. There is nothing scientifically wrong with keeping to static based measures of a construct which is traditionally associated with such measurement.

5.2.7.2 A(ii) Philosophy

The epistemological rationale underlying this test which can be gleaned from what has been written, is the statistically derived notion that working memory lies at the root of other batteries' concern with information processing. This construct is important to Swanson, and as with all other researchers he is entitled to seek the truth of his claims. If the logic behind the statement is to be held consistent, one would then have to question the need to change from static assessment to one of dynamically based modes of assessment. The logic can be said to proceed as follows:

1. Swanson's predilection for the information processing approach as playing a major role as construct in test batteries thus far encountered within the intelligence literature
2. His subsequent awareness of the core feature of information processing - that of working memory
3. Leading to the undisputed notion that working memory is indeed imperative in processing information as it features heavily in discriminations between learning disability testees and normal functioning testees
4. Which it is presumed can be derived via dynamic assessment methods thus perhaps unintentionally doing a disservice to the domain of dynamic assessment per se
5. And due to the reliance on static based measures and static based approaches in validation and reliability detection is unable to adequately function within a learning potential paradigm

There is thus nothing convincingly philosophical regarding the tests' claim to assess truly dynamically - only superficially so. The test has its own governing philosophy yes, but not one that can be labelled as "change-based" in its conception of the human being. The concern for idiographic relevance outweighs its concern for nomothetic relevance. This is a decided disadvantage as far as dynamic purists go but should not surprise those more concerned with norms and generalisability on a large scale. There is thus nothing wrong with the test as such, as it pertains to standard assessment and is lauded as a fruitful effort at placing working memory at the core of its approach (especially given the emphasis that the author places on *g!*) Once again, is *g*-based testing incommensurate with a dynamic approach given the method's own philosophy? The S-CPT is decidedly influenced by its data language and the use of its methodology. One need only review the seven types of scores evidenced in the test battery to understand the philosophy behind it. From the outset, the model behind the battery (as can be best gleaned) warrants its methods. Despite superficial attempts at eliciting potential, the model does not truly assess for this construct (see section C(iii) below).

5.2.7.3 B(i) Hypothetical terms

Swanson's S-CPT is perfectly positioned to fulfil Madsen's criteria of ontological referential hypothetical terms, including mentalistic, organismic and constructive hypothetical terms. Recall Madsen's notion of hypothetical terminology serving as link between empirical and hypothetical aspects as well as serving to bring coherence to a system which may at times seem chaotic (and what research area is more fraught with chaos in terms of construct labelling than dynamic assessment and intelligence research? In fact the entire soft science approach characterises this state of affairs, but this has already been detailed in chapter 3). A candidate in the offing, representative of a mentalistic term, would in this case be "processing potential" which we now realise to be Swanson's way of legitimising a traditionally understood concept of information processing within a learning

potential paradigm. “Working memory” is a potential candidate term representing an organismic term although it can be argued that this term is reflective of the construct itself and not the person possessing this construct. Another candidate could well be “learning disability” as this is more in keeping with the framework’s criteria. Constructive hypothetical terms in this instance could encompass both process and structure terms as is illustrated by the various subtests within the S-CPT. The subtests perform the function of making manifest the latent construct to be measured and include the eleven subtests, rhyming, visual matrix, auditory digit sequence, mapping and directions, story retelling, picture sequence, phrase sequence, spatial organisation, semantic organisation, semantic categorisation and nonverbal sequencing. These terms reflect the hypothetical constructs and are most fitting to directive variables according to the Madsenian framework because they reflect controlling and regulatory effects and include cognitive processes and structures which is precisely what the S-CPT is premised upon (cognitive activities as opposed to mental abilities).

5.2.7.4 B(ii) Scientific hypotheses

Existential hypotheses proffer the existence of the hypothetical hypotheses offered above. In other words they seek to concretise the terminology and functional hypotheses seek to explore their effects which in this instance are managed by the subtests. The testability criterion is paramount in placing this test alongside the others that have transpired above. The S-CPT HQ determination follows.

Theoretical (hypothetical) scientific hypotheses

- H (intelligence) - H(information processing)
- H (intelligence) - H (working memory)
- H (potential) - H (processing potential)

Empirical (substantive) scientific hypotheses

- H (information processing as exemplified through WM) - S (rhyming)
- H (information processing as exemplified through WM) - S (visual matrix)
- H (information processing as exemplified through WM) - S (auditory digit sequence)
- H (information processing as exemplified through WM) - S (mapping and directions)
- H (information processing as exemplified through WM) - S (story retelling)
- H (information processing as exemplified through WM) - S (picture sequence)
- H (information processing as exemplified through WM) - S (phrase sequence)
- H (information processing as exemplified through WM) - S (spatial organisation)
- H (information processing as exemplified through WM) - S (semantic association)
- H (information processing as exemplified through WM) - S (semantic categorisation)
- H (information processing as exemplified through WM) - S (nonverbal sequencing)
- H (trainability) -R (initial score)
- H (trainability) - R (gain score)
- H (trainability) - R (instructional efficiency score)
- H (trainability) - R (maintenance score)
- H (trainability) - R (difference score)
- H (trainability) - R (processing stability score)
- H (trainability) - R (strategy efficiency score)

$$\Sigma (H-S) = 11; \Sigma (H-R) = 7; \Sigma (H-H) = 3.$$

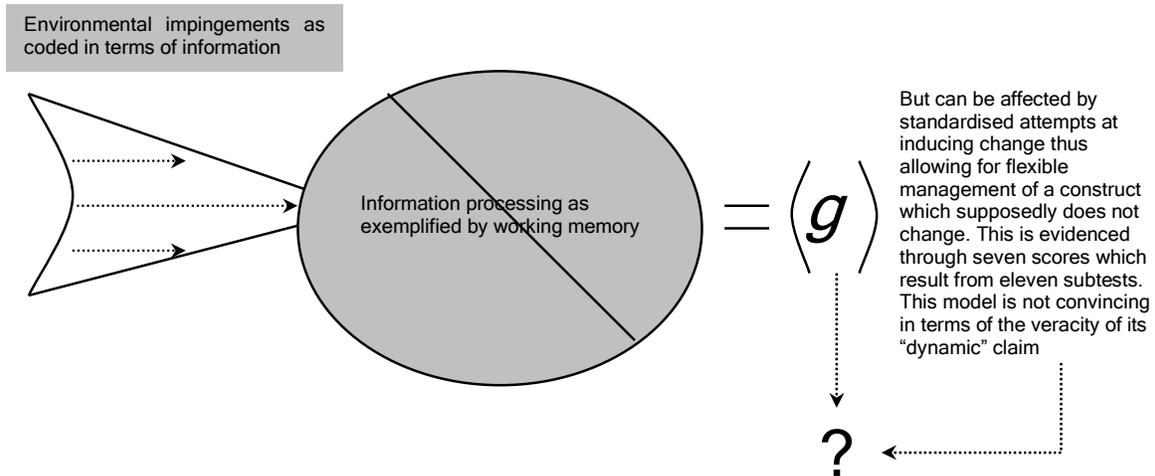
Hence HQ for the S-CPT = $3/(11+7) = 0.16$ which is not surprising. It is remarkable how such a simplistic rendering of a model by means of the HQ can be predicted by the types of terminology utilised by the authors in their models. The S-CPT is indeed a testable model.

5.2.7.5 B(iii) Hypothesis system

Swanson does make use of a nomological deductive argument in support of this model and it follows the following logic:

- Information processing is the approach from which all else is derived as it pertains to intelligent functioning (premise)
- Working memory has been highlighted in most information processing models as being of particular significance (premise)
- Hence, Swanson’s model should follow suit and establish for itself a similar core construct around which to work and extrapolate (conclusion)

- Note that the approach itself is not defended as this is situated in the realm of untestable hypotheses so even though the model is bounded by traditional deductive reasoning, its origins are borne from inductive logic which is known to be unfalsifiable (see chapter 3)
- The model does not take the form of a two or three dimensional depiction but the following can be illustrated as perhaps indicative of what it might look like:



The model when depicted as above does not seem to offer convincing arguments for the change in construct of working memory from one of statically designated term to one evidencing change. The manner of assessment can be said to be dynamic but not the constructs themselves which is precisely where one of the major issues within the dynamic assessment lies. Swanson states “examples of probes (DA)...” (2000, p.75) which is tantamount to equating dynamic assessment to probing which in fact is not entirely true nor very representative of dynamic assessment philosophy. Truly, one cannot say that this model brings us any closer towards a viable solution for placing the learning potential paradigm within intelligence assessment. However, as it stands, the model does very well within the ambit of pure static-based conceptions of invoking working memory as necessary construct in intelligence research.

5.2.7.6 C(i) Abstract data

Abstract data within the Madsenian framework consists of terms that are descriptive but not hypothetical in nature. Information processing is functionally related to working memory with a correlational relation being tacitly assumed to exist between working memory and g . The hidden relation between this correlation relation and its subsequent functional relation to “processing potential” is a leap of faith. The data gathered from factor loadings as well as from correlations with other batteries assessing for working memory yield supportive evidence in favour of the construct’s importance within intelligence assessment (given the premise that it indeed is so) but cannot be logically deduced to conclude that the same is true for change as evidenced from processing potential.

5.2.7.7 C(ii) Concrete data

Manifest behaviour is the data level at which researchers can be the most clear in terms of increments or decrements in observable changes. The need for less intensive prompts or hints is quantifiable. The entire system devised within the S-CPT is a vindication of quantification and in the case of change is a simplistic linear rendering of a construct assumed *a priori* to exist as such. In other words change is evidenced through the manipulation of various scores in a straightforward manner. Near transfer is tested but the need for far transfer and maintenance across domains is not assessed.

5.2.7.8 C(iii) Prime considerations

Correlations are the driving vehicle in the original decision as to the choice of utilising working memory as construct most indicative of information processing as it has been associated with and co-varied consistently across numerous test batteries. Qualitative modes of assessment are utilised in terms of probing based on the type of error made but for the most part the scoring of change and base level of functioning is assessed in a classical test theory set up where scores are subtracted from one another with no indication as to the nature of change in the underlying construct which we know to change as evidenced in item response theory models of change-based assessment. The test is thus based on performers and not test item information and thus allows nothing to be inferred from specific items and how they function under conditions where change has supposedly



occurred. Swanson states that there is as yet no agreed-upon manner of assessing for cognitive modifiability, which may be true but to resort to standardised methods is unfortunate. The following measures are detailed as:

The facts

- i. Initial score which is equivalent to an IQ score
- ii. Gain score obtained via probing which represents the highest score possible under such conditions
- iii. Probe score - the number of probes necessary to achieve the gain score
- iv. Maintenance score is the degree to which the gain score can be maintained without further probes
- v. Processing difference score which is the gain score minus the initial score which as Swanson states measures the Vygotskian notion of ZPD which we know not to be the case
- vi. The degree to which the score can be maintained minus the initial score represents stability of processing
- vii. A final score which is not considered further (strategy efficiency which is more qualitative)

With commentary on the each of the above

- i. The assumption being that assigning a number to a behavioural response becomes measurable. This manner of equating to weakly defined intelligence measures does not aid in its dynamic status. Recall that simply stating that something is measurable does not necessarily make it so. This score is in reality being treated as an extensive measure whereas it is in fact an intensive measure not capable of concatenation and is premised on derived correlations which means that no unique attempt is being made to carve a new defining feature for the construct of change but remains dependent on the old notion. The problem that enters later on is how order-preserving the scores really are. If change as unacknowledged (in this instance) underlying construct is not stable (i.e. not reliable as index as classical test theory would have it) then how can one assume that the initial score is any way meaningfully related to the gain score? Can the one rightfully be subtracted from the other in the hope that the same scale is usable? Also recall that psychological constructs are not interval scaleable due to their intensive status but they can become measurable if derived additively from conjoint measures taking the process back to its basic axioms
- ii. This type of crude measure harks back to Bereiter's concern over the measurement of change but when taken in tandem with the probe score can be said to reflect two conjoint factors which make possible the measurement of change as manifest
- iii. A crude measure as well but once again can become measurably meaningful if taken in tandem with the gain score
- iv. Reflecting ostensibly "how well the testee has learned" which perhaps comes closest in this model in determining potential which in the author's opinion is a more adequate indicator than
- v. What can be reflected by a simple linear subtraction of initial from gain score
- vi. How relevant is this score? The processing stability score represents a *rate* of maintenance. Can one accommodate two types of measures in this way?

In essence one can almost comment on the near perfect CTT tenets invoked and preserved within this test battery which should, to all intents and purposes not be near this level of accuracy given its dynamic status. This does not mean that dynamic assessment batteries per se are neither reliable nor valid, as they are within the ambit of modern test theory. This test functions within a CTT environ and comfortably within a static-based scenario as is clearly evidenced in its normative and practical empirical result. This is not a criticism of the test as such but it is a trenchant criticism lodged from a learning potential approach. There is nothing to indicate that the test does not function, it obviously performs reliably in terms of detecting and differentiating learning disabled students from those who are not learning disabled but have been mistakenly labeled as such. The question surrounding this battery is its claim to dynamic assessment status.

5.2.8 Feuerstein's Learning propensity assessment device (LPAD)

The Learning Propensity Assessment Device (LPAD) can be considered the "mother of dynamic assessment batteries". Due to the considerable number of models and theories within the dynamic assessment field, many of which have taken ideas, concepts, tools and general bearings from this model, the model explores very similar notions of assessment as do many other test batteries. The LPAD was an early conceptualised, systematised, tested and published dynamic assessment battery in the field and was a pioneering effort at combating static-based notions of intelligence as trait as opposed to intelligence being construed as a state. Trait indicates a stable construct and is an *a priori* biased heuristic deployed by many assessors without much thought on the matter. Intelligence as state allows process-based inference of definition. Most of the models discussed in this chapter evidence similar ideas and notions to that of the LPAD, but the major distinguishing feature of the LPAD was that it was the first model to be utilised in the manner described above. The LPAD is the longest lived model within the dynamic

assessment paradigm and has received attention from scholars across the globe undergoing necessary transformations throughout the years.¹¹ It is the model against which most other models and theories derive construct meaning and from which most form some sort of attachment whether in theory or practice. Simply put, a discussion on dynamic assessment hardly ever omits references to the LPAD.

Its venerable status in the field is due to two main reasons; the novelty associated with the underlying theory of structural cognitive modifiability (SCM) working in tandem with mediated learning experience (MLE) which is exemplified and made manifest via the tools of the LPAD and the length of time it has been resident in the literature. Due to its age, it has evolved to become more refined and more importantly entrenched in theory. The LPAD is a particularly good example of a theoretically fostered assessment approach and battery. This will be seen when Madsen's HQ is calculated for the battery. It has, however, been the recipient of much criticism but has continued to be utilised unabated in areas where the model and technique are warranted and feasible. One of its most innovative contributions to dynamic assessment as paradigm (if it indeed can be considered as such) is its concern for novel construct meaning-making at a time in intelligence history where immutable constructs such as intelligence was a universal given. No longer a novel idea, the LPAD sourced its novel constructs from individuals evidencing lack of the substance referred to as intelligence. In other words, in order to offer the realm of intelligence assessment and research a method by which to progress, the model's theory necessitated a return to individuals who did not possess such "substance" in the first place; or at least did not possess much of "it".

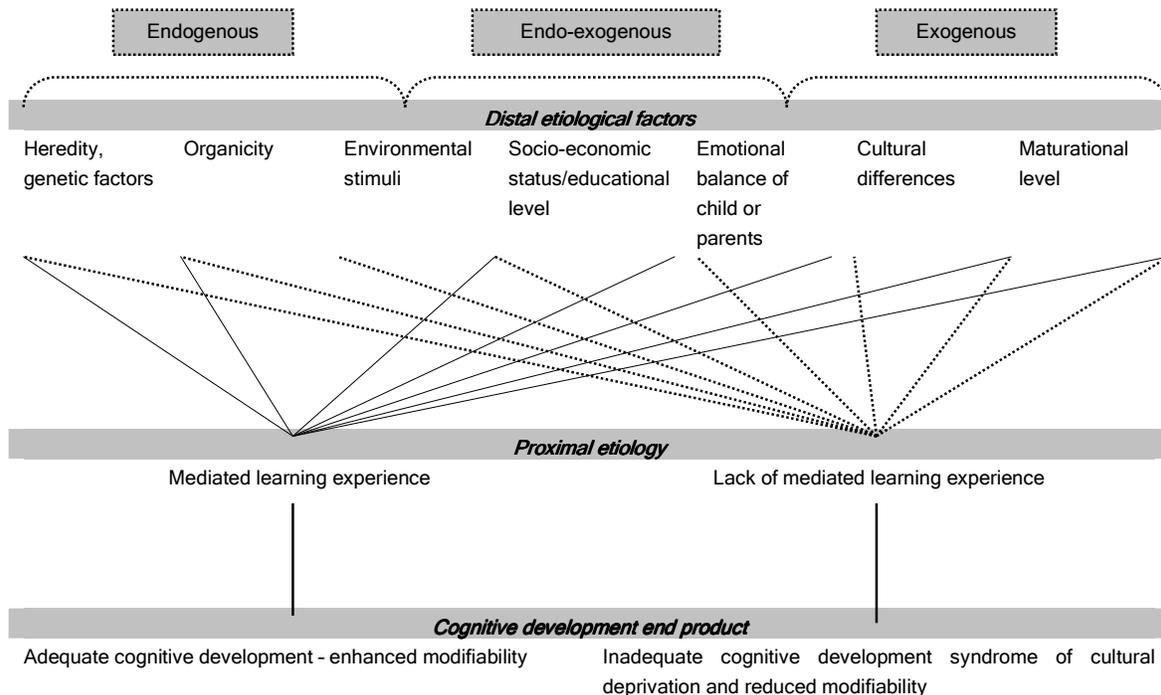
Moreover, the LPAD's nascent beginnings were resultant from a surge in practical problems surrounding mass testing of orphaned and displaced individuals. Would an assessment package such as the LPAD have come into existence if there was no pressing practical need? Its pragmatic beginnings align it with a school of thought where capability and intelligence needs to be assessed in a manner as expedient but also as fair as possible. The rationale for Binet's choice of testing as well as Vygotsky's dilemma of having to construct measures of assessment predicated on the notion of equality are consistent in rationale underlying the development of the LPAD. All three researchers were asked or compelled by institutions to come up with novel ways of fairly assessing individuals in a manner which did not disadvantage them and also in a way that could be utilised within the then-current framework of assessment practice. The LPAD battery/model/theory is also perhaps the most studied, utilised and critiqued dynamic assessment battery to date. This brief discussion can hardly do justice to the LPAD and the intention here is most certainly not to pry apart every aspect of the theoretically wide-ranging model but to place the model as practically as possible within the Madsenian framework so that it too can be compared to other models within the field. This exercise is thereby constrained and is thus limiting in its conclusions. Recall that the metatheoretical framework is itself bound by its own constraints. This very brief section then, only comments on the LPAD as it pertains to this manner of assessment and according it a place within the broader framework. It goes without saying (and is applicable to every one of the models discussed in this section) that there are certainly many facets of the model which will not receive attention as they should rightly receive if analysed according to other frameworks. The discussion is couched within the LPAD as model and as will be evidenced below, is presented as highly testable in Madsen's HQ determination. This may strike the reader as odd especially due to the clinical nature of the intervention programmes usually associated with the LPAD. But it is the philosophical model which is discussed as it pertains to the test battery itself which results in an unusually high HQ score. The reasoning for this will become clear.

5.2.8.1 A(i) Ontology

Central to Madsen's comparative meta-theoretical taxonomy is the constant interplay between meta-concerns and data theses. The dictum of theory informing practice and practice informing theory is clearly evident in the framework. Data gathered from below invariably seeps through to meta-strata which in the case of intellectual assessments of immigrants to Western cultures necessitated a return to theory due to plausibly incorrect notions of what to assess for in populations evidencing vastly divergent circumstances from the norm group upon which static-based assessment were predicated. The LPAD is truly a valid exponent of the reigning times during which it was conceived as its existence is explained by cultural attitudes towards assessment and assessment of minorities (in whatever manner minorities are construed). The broader political ideology and practices resultant from such ideologies created a niche for the development of alternative methods of assessments for individuals supposedly lacking substantial 'areas of a trait'. The need to assess equitably culminated in novel construct meaning-making which became leveraged on a state-based understanding of intelligence and not a trait-based approach. The move away from trait to state is no longer looked upon with awe but is, in some circles, considered a given. This was not the case over fifty years ago, at least not within mainstream settings.

¹¹ The LPAD was originally referred to as the Learning Potential Assessment Device. 'Potential' is defined as something which is possible but not yet realised; capable of existence but not yet existing. 'Propensity' is defined as an innate inclination or a tendency. The latter seems to evoke a concern for the already-existing state of changeability whereas the former indicates that changeability is not forthcoming till it becomes manifest. The latter also seems to imbibe meaning of expectant latency of changeability. This is how the author views it.

Analysis of the LPAD in terms of ontology is particularly easy to accomplish as Feuerstein and colleagues (Feuerstein, Feuerstein, Falik & Rand, 2002) explicitly posit their understanding of the human being which can be viewed below. There is no need to search for implicit assumptions surrounding ontological issues within the model as these assumptions are highlighted as fuelling and guiding the assessment process. Given the LPAD's impetus in not assuming stability of intelligence as trait, the determinants of functioning are indeed very closely aligned to modern model counterparts which likewise stress the importance of considering biological, environmental and other factors and how these mould intellectual and social behaviour. As with Vygotsky who was also keenly aware of the biological basis of functioning, Feuerstein et al. (2002) are not blind to current thinking and trends within mainstream intelligence assessment. Current conceptions of the individual as coalescent product of genetic and environmental impingements are similar to most models pervading intelligence research. The one dominant critique to be lodged against this model is its broad scope encompassing, as it does, a vast array of aspects. The model, in addition, is a qualitatively based form of assessment.



The individual is viewed as a free agent who is not under the control of innate capacity but able to negotiate through development in a manner befitting adaptive agents. Madsen viewed the conception of man from three positions; biological, social and humanistic and utilised this spectrum as dimension along which to identify theoretical choice of placement. The LPAD assumes that the individual can be explicitly assessed along each dimension. Psychophysical theory has to be inferred from the LPAD but due to its pragmatic agenda, it can be assumed that brain is mind (behavioural repertoire is brain made manifest). Madsen's human freedom of action concern is clearly upheld as affirmative. The model exemplifies growth and change from within and without (proximal and distal properties) but also emphasises that both distal and proximal influences can be dangerous in outcome if such influences are simply not present or fraught with negative factors. Feuerstein's preoccupation with such impinging variables is easily traced to ideas emanating from war-torn countries where thousands of displaced individuals had to be accommodated in terms of assessment and further education. Such concerns are not relegated to war-torn countries but also to families where both proximal and distal influences are, for a variety of reasons, not present or at least not adequately mediated to the child. Human freedom of action permeates the theory and it comes across that affirmation of change is a given. In others words, as with most models within dynamic assessment but specifically those more attuned to qualitative assessment, such freedom to change is an *a priori* meta-concern governing the model as seen above.

5.2.8.2 A(ii) Philosophy

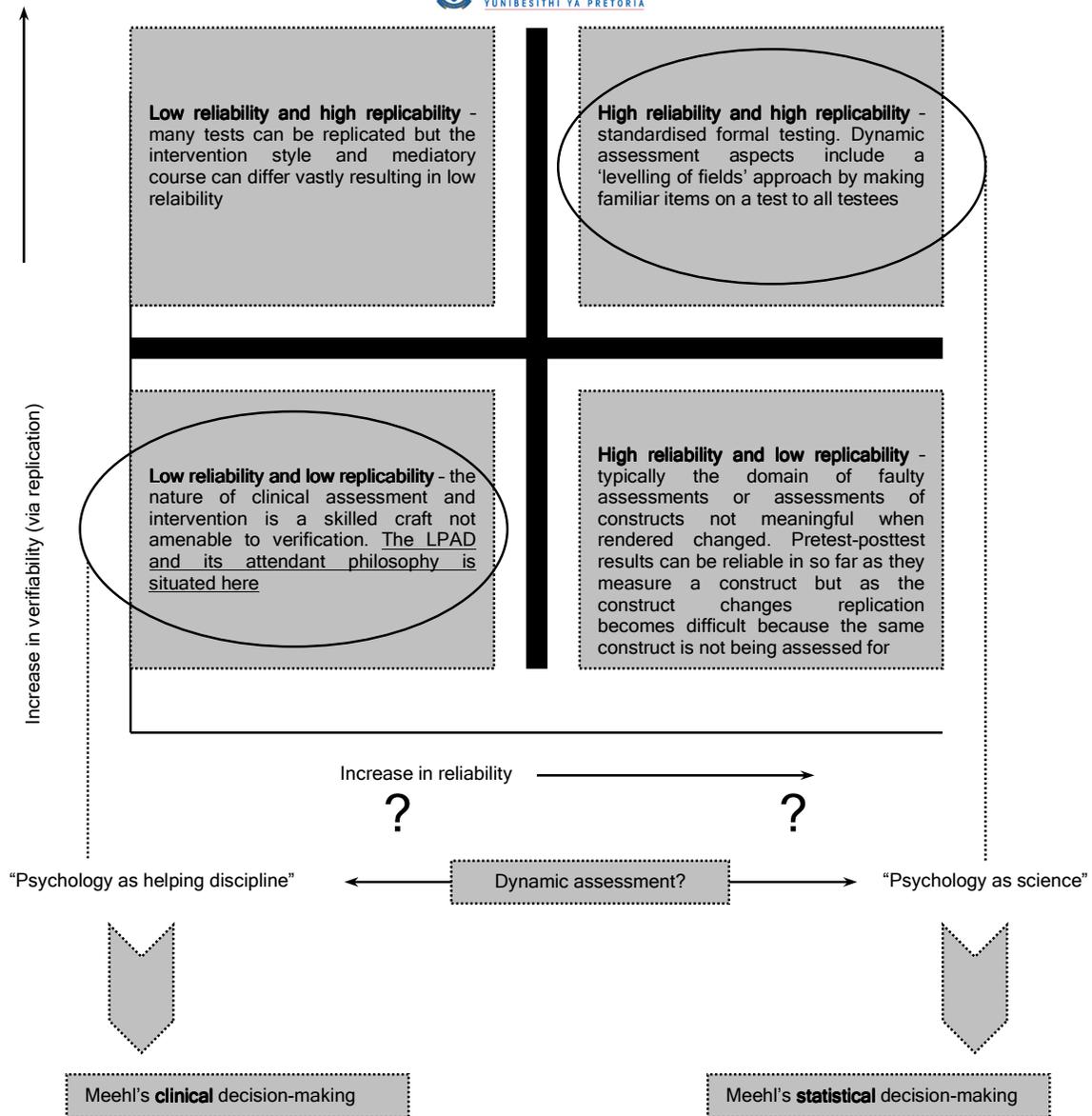
Ontological descriptors showcase what can be known and in this instance, as with most models in this area, it is assumed that potential and change can be known and elicited. Given this implicit stance it remains the task of concretised theory (model) to search for ways and means of doing justice to the mandate set forth by the theory's ontology. Different models have different methods for doing this but essentially travel along a similar path towards their end goal. The notions underlying the LPAD and the actual method utilised throughout the LPAD's deployment are synchronised in understanding the task necessitated by the intervention in dynamic assessment scenarios. For instance, theory built around the LPAD tool is infused with structural



cognitive modifiability (SCM) and mediated learning experience (MLE) and both theory and tool co-ordinate the effort to make change not only theoretically feasible but practicable. The goal for Feuersteinian assessment is to mediate and remediate and not necessarily to assess. Assessment is a core function of remediation but the emphasis is very clearly on process vs. product. The LPAD is an extension of MLE and SCM theory and it is not workable to discuss the LPAD without reference to these two theoretical contributions. Ontologically, potential can be realised because it is assumed to exist in some manner or form, not necessarily as a construct (typified in mainstream assessment as a measurable entity) but as a process of engagement through which it can appear as construct (change construct). Epistemologically, potential needs to be reified to some point in reality and equating potential with a reified construct cannot be further removed from the LPAD's mandate; so to circumvent the notion of reified construct (as this is what the LPAD in essence seeks to avoid) process assessment and intervention become the hallmark of this type of assessment/intervention strategy. It can be argued that true LPAD programmes are rather more like interventions than assessments.

The LPAD can be described as a humanist approach towards assessment. Its motivational thrust seeps from a concern with the individual and how best to engage in progressive steps forwards toward a position of attainment within the larger environment be it educational, career wise or coping in life in general. All dynamic assessment initiatives, in some form or another, seek to remediate and are thus concerned with the individual as opposed to the individual's scores on a test. The LPAD's strong theoretical base evidences a concern for a scientific approach towards assessment which although purporting to assess in as holistic a fashion as possible, does so via a stringent theoretical framework. Theory belongs to the realm of nomothetic ideal although it can be argued that it can apply only to idiographic ideals if needs be. The LPAD's tasks are theoretically grounded in a nomological network but the application is based on idiographic set-ups. Criticisms levelled at the LPAD result mainly from the latter's perceived impracticalities as assessment and intervention sessions are lengthy and often not feasible within constrained situations. The mediation necessitated by individuals is similar in notion to the therapeutic interventions that many of these individuals require and as with most therapies, time is needed to ensure the success of intervention. The philosophy behind the development of the LPAD and the origins of theory underlying the assessment intervention can be easily accommodated in a clinical set-up which it invariably reflects.

The idiographic goal of individualised assessment and attention characterises the LPAD but also sets it apart from more standardised assessment interventions. Feuerstein and colleagues are keenly aware of the clinical approach towards their dynamic assessment and maintain that in order to reach the goal set forth by their philosophical thinking on the need to adequately assess, a long-term solution is the most viable in terms of ultimate success. On this issue they are likely to be more correct, if only because longer time spent on remediation will increase the chances of future retention and further solidification of strategies learned during intervention. This notion can however be contested and is contested especially when long-term studies seeking to modify intellect show decreased likelihood of intervention efficacy as time progresses. Feuersteinian dynamic assessment is synonymous with long-term intervention style and clinical approach. This highlights the issue of clinical and statistical decision-making in psychology and typifies the controversy eloquently debated by Paul Meehl and discussed in chapter 3. To what extent can judgements and predictions be made about individual functioning if assessment and intervention occurs within a setting not entirely geared towards verifiability? This issue is perhaps most pertinent to the discussion on the utility of the LPAD within the domain of dynamic assessment. If dynamic assessment is construed as a continuum of assessment ranging from very reliable and verifiable on the left to unreliable and not verifiable on the right, the LPAD could quite easily be situated on the extreme right. What has been critiqued above should not in fact be considered as such as it has been argued that psychology as "helping discipline" can and should quite comfortably continue as such but without the attendant scientific dictates being promulgated as being followed. The reader is referred to figure 29 where the originating schism in psychology as science is illustrated. Keeping this figure in mind, the following is offered as further explanation regarding the LPAD rationale in particular. The methodological theses bespeak of idiographic research methods and data language even though much of the early conceptualising for the LPAD took place in a climate less tolerant of steerage away from mainstream understandings of assessment practice.



5.2.8.3 B(i) Hypothetical terms

Feuerstein's LPAD and dynamic approach is peppered with hypothetical terms which seek to provide coherence to a system for which no alternative viable explanations exist. Part of the criticism lodged at the LPAD's approach towards assessment is that there is no system in place for verifiability and even worse, no place for falsifiability of concepts within the network of terminology. This is a particularly good example of 'psychology-as-helping-discipline' and not of 'psychology-as-science'. Many features of the pure clinical approach cannot be rendered scientific for the very reason that the approach does not belong in the area of psychology calling itself scientific. One cannot critique a helping discipline especially if its motivating force is to aid in as humanly a fashion as possible individuals who have not received adequate mediation for whatever reason. How can the decisions based on the LPAD be more 'statistically' attuned? Herein lies the dilemma, as it has been emphasised in chapter 4 that no amount of statistical rendering of results can make it so. Are the two concerns really diametrically opposed in the first instance?

The LPAD primarily assess for deficient cognitive functions and ascribes to a number of mentalistic, organismic and constructive hypothetical terms. In so doing it seeks to remediate upon such deficiencies. This is very evident in Feuerstein's model of what constitutes a mental act via a three-fold process of input, elaboration and output which impinges on his views on the domain of mental operations and specified impairments, all of which have been assigned H-R and H-S terms according to Madsen's HQ system. These can accordingly be divided as follows:

- H_m - mentalistic hypothetical terms



- Lack of, or impaired receptive verbal tools (input phase)
- Lack of, or impaired spatial orientation and lack of stable systems of reference pertaining to the organisation of space (input phase)
- Lack of, or impaired temporal concepts (input phase)
- Lack of, or impaired conservation of constancies of factors such as size, shape, quantity and colour (input phase)
- Lack of, or deficient need for precision and accuracy in data gathering (input phase)
- Lack of capacity for considering two or more sources of information at once (input phase)
- Inadequacy in the perception of the existence and definition of an actual problem (elaborational phase)
- Inability to select relevant vs. irrelevant cues in defining a problem (elaborational phase)
- Narrowness of the mental field (elaborational phase)
- Lack of, or impaired interiorisation (elaborational phase)
- Lack of, or impaired inferential, hypothetical thinking (elaborational phase)
- Lack of, or impaired strategies for hypothesis testing (elaborational phase)
- Lack of, or impaired planning behaviour (elaborational phase)
- Non-elaboration of certain cognitive categories because the verbal concepts are not a part of the individual's repertoire on a receptive level (elaborational phase)
- Lack of, or impaired verbal or other tools for communicating adequately elaborated responses (output phase)
- Lack of, or impaired need for precision and accuracy in communicating responses (output phase)
- Deficiencies in visual transport (output phase)
- H_o - organismic hypothetical terms
 - Blurred and sweeping perception (input phase)
 - Unplanned, impulsive and unsystematic exploratory behaviour (input phase)
 - Lack of spontaneous comparative behaviour of the limitation of its application by a restricted need system (elaborational phase)
 - Episodic grasp of reality (elaborational phase)
 - Lack of, or impaired need for pursuing logical evidence (elaborational phase)
 - Egocentric communicational modalities (output phase)
 - Difficulty in projecting virtual relationships (output phase)
 - Blocking (output phase)
 - Trial and error responses (output phase)
 - Impulsive, acting-out behaviour (output phase)
- H_c - constructive hypothetical terms
 - the input-elaboration-output model of a mental act can in its entirety be considered a constructive hypothetical system as it is, according to Madsen, an analogised version of an explanatory system

It is ironic that an approach towards assessment exemplified in its most clinical form attests to such strict delineation of terms and process set out in a framework detailing the aspects of behaviour and mental processes in relatively objective form. It is evident from the above that such hypothetical terminology exists within an ontological network made manifest via a process of assessment. Quite rightly, one could refer to these hypothetical terms as hypothetical constructs. Perhaps this idea was resident within Madsen's original conceptualisation of the hypothetical stratum. It can also be promoted that the issue surrounding hypothetical and empirical constructs can be paralleled in the hypothetical and data strata. If this is so, Madsen's meta-theoretical framework was indeed prescient.

5.2.8.4 B(ii) Scientific hypotheses

The LPAD is unique among test batteries surveyed in this chapter in terms of its theoretical foundation which offers much by way of model exposition and theory underpinning. The usual format for the discussion on scientific hypotheses entails detailed delineation of what constitutes hypothetical and empirical variables as utilised within the model. However, as the LPAD is constructed upon MLE and SCM theory as well as being deployed via its own tasks in a dynamic manner, this section will need to encompass three separate analyses, which can and should be read and understood as pertaining to the entire rationale behind the LPAD. Section (i) focuses on the MLE theory and its associated variables, section (ii) on SCM theory and its associated variables and lastly section (iii) on the LPAD as assessment instrument which makes manifest much of what underlies it theoretically. Section (iv) seeks to amalgamate the three sections' results. One criticism which can be justifiably lodged at the constructs highlighted by Feuerstein is the at-times vague conceptions of what in fact they represent. Inclusion into Madsen's HQ system usually requires that a construct be defined as judiciously and as parsimoniously as possible. A number of Feuerstein's constructs are somewhat less definitive but are extrapolated upon in his renderings of these constructs particularly as they become manifest during the LPAD assessment and mediatory intervention programme.

Section (i) - Scientific hypotheses for Mediated Learning Experience (MLE)

Theoretical (hypothetical) scientific hypotheses

- H (intelligence) - H (is a trait not a state as exemplified through potential made manifest)

Empirical (substantive) scientific hypotheses

- H (mediated learning experience) - S (intentionality-reciprocity) (universal parameters of mediation)
- H (mediated learning experience) - S (transcendence) (universal parameters of mediation)
- H (mediated learning experience) - S (mediation of meaning) (universal parameters of mediation)
- H (mediated learning experience) - S (mediation of the feeling of competence) (situational parameters reinforcing and elaborating MLE)
- H (mediated learning experience) - S (mediation of regulation and control of behaviour) (situational parameters reinforcing and elaborating MLE)
- H (mediated learning experience) - S (mediation of sharing behaviour/individuation and psychological differentiation)(situational parameters reinforcing and elaborating MLE)
- H (mediated learning experience) - S (mediation of goal seeking, goal setting and goal achieving behaviour) (situational parameters reinforcing and elaborating MLE)
- H (mediated learning experience) - S (mediation of challenge: the search for novelty and complexity) (situational parameters reinforcing and elaborating MLE)
- H (mediated learning experience) - S (mediation of awareness of the human being as a changing entity) (situational parameters reinforcing and elaborating MLE)
- H (mediated learning experience) - S (mediation of search for an optimistic alternative) (situational parameters reinforcing and elaborating MLE)
- H (mediated learning experience) - S (mediation of the feeling of belonging)(situational parameters reinforcing and elaborating MLE)
- H (dynamic assessment) - R (retention/permanence)(criteria according to which change is evaluated)
- H (dynamic assessment) - R (resistance)(criteria according to which change is evaluated)
- H (dynamic assessment) - R (flexibility/adaptability)(criteria according to which change is evaluated)
- H (dynamic assessment) - R (generalisability/transformability)(criteria according to which change is evaluated)

$$\Sigma (H-S) = 11; \Sigma (H-R) = 4 ; \Sigma (H-H) = 1$$

$$\text{Hence HQ for the theoretical underpinning MLE} = 1/(11+4) = 0.06$$

Section (ii) - Scientific hypotheses for Structural Cognitive Modifiability (SCM)

Theoretical (hypothetical) scientific hypotheses

- H (structural cognitive modifiability) - H (change)

Empirical (substantive) scientific hypotheses

- H (cognitive map and task dimensions) - S (content)
- H (cognitive map and task dimensions) - S (modality)
- H (cognitive map and task dimensions) - S (phase)
- H (cognitive map and task dimensions) - S (operations)
- H (cognitive map and task dimensions) - S (level of complexity)
- H (cognitive map and task dimensions) - S (level of abstraction)
- H (cognitive map and task dimensions) - S (level of efficiency)
- H (cognitive functions in the mental act) - R (blurred perception) (input phase)
- H (cognitive functions in the mental act) - R (impulsive exploratory behaviour) (input phase)
- H (cognitive functions in the mental act) - R (impaired receptive verbal tools) (input phase)
- H (cognitive functions in the mental act) - R (impaired spatial orientation) (input phase)
- H (cognitive functions in the mental act) - R (impaired temporal concepts) (input phase)
- H (cognitive functions in the mental act) - R (impaired conservation of constancies) (input phase)
- H (cognitive functions in the mental act) - R (deficient need for precision) (input phase)
- H (cognitive functions in the mental act) - R (lack of capacity for considering two or more sources of information simultaneously) (input phase)



- H (cognitive functions in the mental act) - R (inadequacy in the perception of the existence and definition of an actual problem) (elaborational phase)
- H (cognitive functions in the mental act) - R (inability to select relevant vs. irrelevant cues in defining a problem) (elaborational phase)
- H (cognitive functions in the mental act) - R (lack of spontaneous comparative behaviour) (elaborational phase)
- H (cognitive functions in the mental act) - R (narrowness of the mental field) (elaborational phase)
- H (cognitive functions in the mental act) - R (episodic grasp of reality) (elaborational phase)
- H (cognitive functions in the mental act) - R (impaired need for pursuing logical evidence) (elaborational phase)
- H (cognitive functions in the mental act) - R (impaired interiorisation) (elaborational phase)
- H (cognitive functions in the mental act) - R (impaired inferential, hypothetical thinking) (elaborational phase)
- H (cognitive functions in the mental act) - R (impaired strategies for hypothesis testing) (elaborational phase)
- H (cognitive functions in the mental act) - R (impaired planning behaviour) (elaborational phase)
- H (cognitive functions in the mental act) - R (non-elaboration of certain cognitive categories) (elaborational phase)
- H (cognitive functions in the mental act) - R (egocentric communicational modalities) (output phase)
- H (cognitive functions in the mental act) - R (difficulty in projecting virtual relationships) (output phase)
- H (cognitive functions in the mental act) - R (blocking) (output phase)
- H (cognitive functions in the mental act) - R (trial and error responses) (output phase)
- H (cognitive functions in the mental act) - R (impaired verbal or other tools for communicating adequately elaborated responses) (output phase)
- H (cognitive functions in the mental act) - R (impaired need for precision and accuracy in communicating responses) (output phase)
- H (cognitive functions in the mental act) - R (deficiencies in visual transport) (output phase)
- H (cognitive functions in the mental act) - R (impulsive acting-out behaviour) (output phase)
- H (cognitive functions in the mental act) - R (affective motivational factors)

$$\Sigma (H-S) = 7; \Sigma (H-R) = 28; \Sigma (H-H) = 1$$

$$\text{Hence HQ for SCM} = 1/(7+28) = 0.02$$

Section (iii) - Scientific hypotheses for the LPAD

Theoretical (hypothetical) scientific hypotheses

- H (learning potential) - H (change as state not trait)
- H (learning potential) - H (modifiability)

Empirical (substantive) scientific hypotheses

- H (cognitive dimensions) - S (degree of novelty and complexity of the task)
- H (cognitive dimensions) - S (language or modality of presentation)
- H (cognitive dimensions) - S (mental operations required to solve a given problem)
- H (cognitive dimensions profile) - R (grasping of underlying principle)
- H (cognitive dimensions profile) - R (amount and nature of investment required to teach the principle)
- H (cognitive dimensions profile) - R (extent to which principle is applied in problem solution)
- H (cognitive dimensions profile) - R (modality choice and preferences for these)
- H (cognitive dimensions profile) - R (effects of remediation)

$$\Sigma (H-S) = 3; \Sigma (H-R) = 5; \Sigma (H-H) = 2$$

$$\text{Hence HQ for the LPAD model} = 2/(3+5) = 0.25$$

Section (iv) Amalgamation of scientific hypotheses

In sum:

$$\Sigma (H-S) = 11; \Sigma (H-R) = 4; \Sigma (H-H) = 1$$

$$\text{Hence HQ for the theoretical underpinning MLE} = 1/(11+4) = 0.06$$

$$\Sigma (H-S) = 7; \Sigma (H-R) = 28; \Sigma (H-H) = 1$$

$$\text{Hence HQ for SCM model} = 1/(7+28) = 0.02$$



$$\Sigma (H-S) = 3; \Sigma (H-R) = 5; \Sigma (H-H) = 2$$

$$\text{Hence HQ for the LPAD model} = 2/(3+5) = 0.25$$

If all three HQ scores are computed on the usual basis of HQ determination the following is obtained:

$$\Sigma (H-H) = 4$$

$$\Sigma (H-S) = 21$$

$$\Sigma (H-R) = 37$$

$$\text{Hence HQ for the complete LPAD and its underlying rationale} = 4/(37+21) = 0.06$$

Of note is the LPAD HQ which is the highest among all three HQ scores and whilst it is not in itself a high score, it does present as interesting result. The LPAD is a physical manifestation of the many varied hypothetical and empirical constructs permeating the LPAD rationale. Being essentially a replicable model, it should to all intents and purposes reflect the lowest HQ score as the constructs dealt with are empirical in nature. The LPAD as such then, when analysed according to Madsen's model, presents with fewer empirical constructs as opposed to the underlying theory upon which it is based. MLE theory presents with the most H-S hypothetical constructs possibly suggesting that as a theory it seeks information from the individual on more dimensions than SCM and the LPAD. SCM seeks far more than either the LPAD and MLE in terms of responses from mediation and assessment. The main reason being the structured nature of the cognitive mental act; being elaborated to encompass three stages of cognition according to Feuerstein. The most striking aspect about this analysis is the very low HQ result which indicates how testable the approach really is. This is odd to say the least, especially given the criticisms lodged at the approach since its inception, about its purported untestability and unverifiability (as well as its unfalsifiability). This leads to the asking of three questions: is the Madsenian model flawed? is the author-determination of HQ flawed? or have criticisms been far too harsh?

- Madsen's model is workable to a point which is why it was attenuated to incorporate a number of other pertinent aspects to this thesis. In essence fault cannot be found with the model (assuming that it is an agreed-upon model to deploy in this situation which it is)
- The determination of what goes into the model is based on the author's understanding and interpretation of the model characteristics. However, as set out initially by Madsen, the model does well to guide the researchers into what goes where in the model. Needless to say, other researchers may well include constructs not considered by this author as necessary and may well swap around other constructs
- Criticism comes in varied forms some legitimate and other less reasoned and more impassioned. The former manner of criticism views the LPAD as essentially not robust enough to sustain its own credibility within the psychometric tradition. Feuerstein has argued at length that this is the very point; theirs is not a tool to be ensconced within the psychometric intelligence-as-trait paradigm but is firmly lodged within the clinical method of assessment. This does, of course, have its own drawbacks as was illustrated above in section 5.2.8.2.

What is one to make of this paradoxical state of affairs? It could be said that the theoretical underpinning evidences one of the strongest foundations of dynamic assessment models to date as it has had more time to become better established and researched. The LPAD approach was conceived of and developed within a specific dynamic assessment framework and can be said to have devised for itself its own dynamic constructs. One possible answer to the paradox is that the constructs identified above are almost all process-based constructs and not product-based constructs thus making it exceedingly difficult to score objectively. This is perhaps the crux of most criticisms levelled against the utilisation of the LPAD. Having said this, the HQ remains high indicating the testability of the LPAD as instrument and theoretical model. Is there a future for a standardised LPAD? This brings us back the discussion highlighted in chapters 2, 3 and 4.

5.2.8.5 B(iii) Hypothesis system

Madsen's scheme of deductively derived explanatory systems as exemplified through logical deductive arguments seems a little ill-at-ease within the LPAD assessment scenario. A dichotomy is once again looming upon this debate. Can the rationale behind the LPAD be said to be illogical and not deductive? As it supposedly presents as clinical it may be odd to state that it is also a logically derived deductive explanatory system in which the hypothesis system seeks to bridge the meta-stratum with the data concerns. There does not seem to be anything wrong per se with the explanatory account of what the LPAD has to offer nor is there anything amiss in terms of model coherence. What seems to be the issue is the one-on-one clinical set-up. Resources and time constraints aside, the model should be evaluated along the lines of scientific credibility and according to the system outlaid by Feuerstein, the link between hypotheses and variables is established. Although Feuerstein's three dimensional cylinder



model will not be reproduced here,¹² the model typifies what Madsen had in mind when he detailed model explanations as serving the function of an explanatory system, which in this case takes the form of a schematic cylinder. The degree of abstraction inherent in the model is no less and no more abstract than any other model encountered within the assessment and intelligence literature. The purpose of the model is three-fold; to describe, explain and govern meta-theoretically. It is ironic that the LPAD methodology, having been conceived with masses of people in mind, becomes a clinical tool for use in very personalised settings but then again, this is perhaps the whole point in the development of the assessment/intervention/remediation approach. Although it was developed according to nomological dictates, it makes itself present in a very hermeneutical and idiographic manner. This area, where the two 'opposing' ideas 'clash' is possibly one reason why the assessment tool has been both a recipient of criticism and praise. It is a striking example of an instrument accommodating both psychometric intelligence research and clinical therapeutic work. Depending on which side of the proverbial fence one happens to be seated, the method and assessment can either be a boon or can be considered as hampering the enterprise of assessment as a whole. The theory of mediated learning experience is predicated on wide-scale nomological findings which indicate that inadequate or lack of mediation, learning or experience can seriously hamper development. MLE shares a number of core concerns with Vygotskian thinking in terms of child development which is a co-constructed process of development necessitating the 'filtering' of information from the environment by an elder or more capable peer. Feuerstein et al (2002, p.91) sum up the essence of their approach towards the assessment of individuals by stating that "most studies exploring types of changes in learning performance at [a] time were limited to the manipulation of the stimuli rather than to the manipulation of the individual".

The hypothesis systems glue together the scientific hypotheses delineated above and does so primarily through the model exposition. It is clearly evidenced that H-R and H-S constructs dominate the hypothesis system which lends credence to the claim that it is not the product which is of import but the process through which the product is obtained. Initial quantification can be said to support the need for some measure of baseline functioning (without which one truly is left in the dark) if only to systematise the remediation endeavour. The quantification of assessment is truly a means to an end and in no way purports to act as an end unto itself. "Human intelligence, within the theory of SCM, is characterised by the option, possibility, and propensity to become meaningfully changed by experience ... modifiability does not just affect the content and skills of the individual, but the structures themselves that are responsible for further acquisition" (Feuerstein et al, 2002, p.101). The overarching concern is the facility with which change can be effected via mediatory interventions. The governing meta-concern is modifiability and change enhancement which is theorised to be accessible via cognitive processing of information. The Feuersteinian manner of bringing about change is conducted from a mediation stand point and hence its clinical nature whereas pure exposure to tasks without the requisite mediatory attempts at correcting deficient cognitive functioning cannot be said to be the same thing. Two different constructs are being assessed for. Dynamic assessment, as is by now obvious, manifests in shades of 'dynamism' and the LPAD can be considered as one of the most therapeutic in nature. This is characterised primarily by the novel role of assessor as mediator.

5.2.8.6 C(i) Abstract data

Often, abstract and concrete data overlap in terms of definition of the construct they purport to measure. Clinically, the behaviour as presented during intervention is assessed for in both subjective and objective ways. The subjectivity with which some aspects of behaviour could be viewed may be cause for concern from certain quarters. Data that can be considered as abstract in terms of Madsen's conceptualisation, namely, highly abstract but containing no hypothetical terminology, include the domain of mental operations which are made manifest through the LPAD tasks and extend across the mental act: input, elaboration and output phases. The interpretation of what constitutes an act is partly credited to the skill of the assessor or mediator, which, when given a score becomes part of the concrete data stratum. The philosophy surrounding this seemingly innocuous move from clinical observation to quantification has been discussed at length in chapter 4. Feuersteinian dynamic assessment is perhaps the model where this debate and discussion comes to a head. One is left asking the one question which plagues almost all of social science disciplines: can clinical assessments be quantified and analysed as such?

5.2.8.7 C(ii) Concrete data

As with any tool within psychology the LPAD has been assessed in a variety of settings and has much to attest to its psychometric status as robust tool especially in group settings and within particular sample settings such as attention deficit disorder sufferers, autistic children, developmentally delayed children and so on. Concrete data builds up theory from the bottom informing the meta-stratum in a two-way process. Other than the realm of therapy interventions which sits more comfortably in the 'psychology as helping discipline' arena, evaluative attempts in general are assessed in terms of veracity of outcome which, by implication, needs to be measurable. The LPAD to some degree as with most other batteries acquiesces to the need for numericisation of observables and details its results in similar descriptive and inferential fashion.

¹² But can be found in Feuerstein et al. (2002, p.164).

5.2.8.8 C(iii) Prime considerations

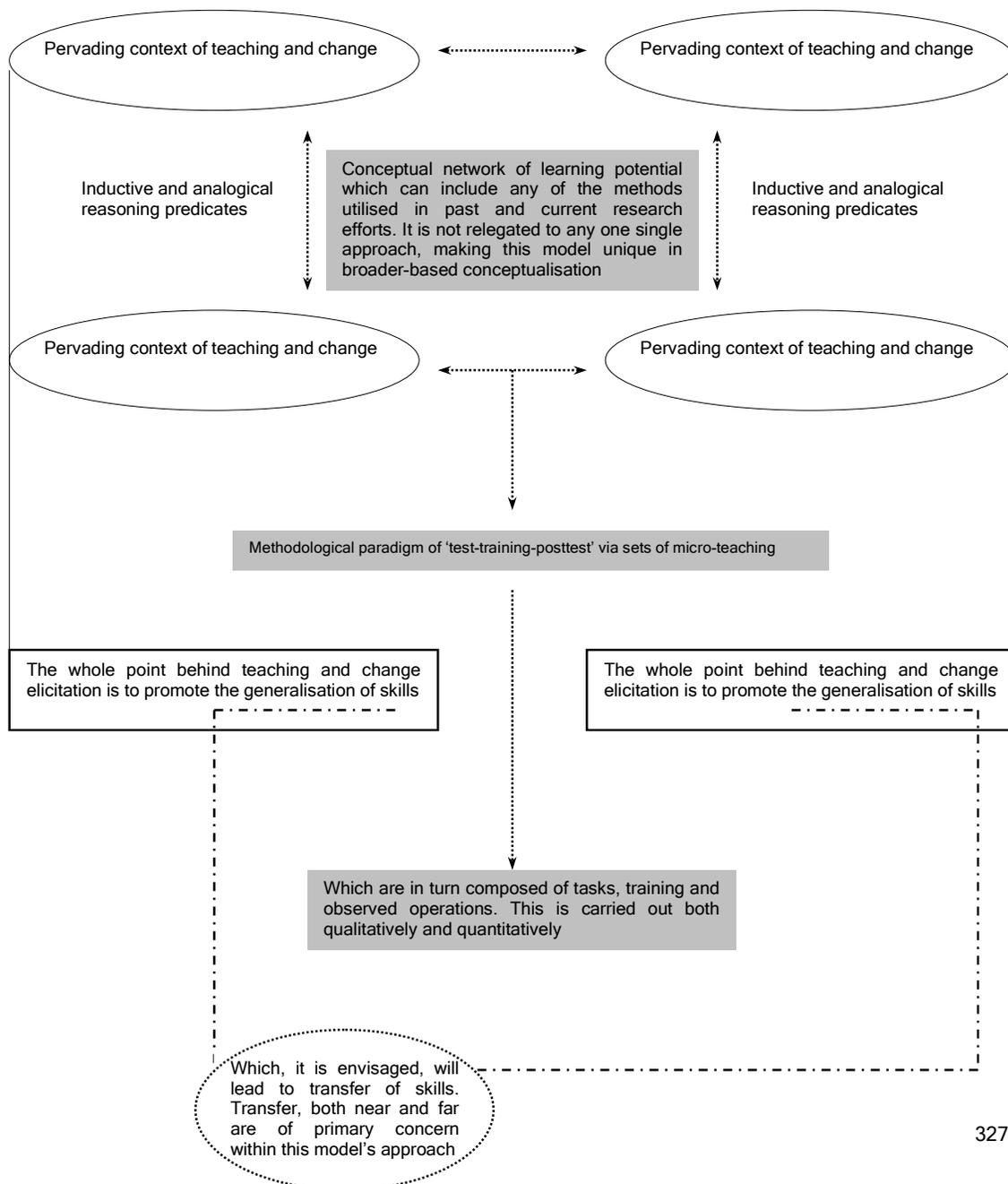
Much of what has been said above in terms of concrete and abstract data can be included in this section.

5.2.9 The assessment of learning potential: the EPA instrument

The EPA's framework is set-out in what can only be described as elaborate ensconcement within multitudinous references.

- Referring to Vygotskian inspired notions of the two-fold nature of teaching and change
- Feuersteinian instrumental enrichment as conduit for eliciting change and
- Budoffian research design similarity predicated upon a foundation of inductive reasoning as assessed for by the Ravens, the model epitomises the method of cross-boundary research.
- Utilising an information processing influenced model depicting experimental models of optimum strategy choice within the spatially and verbally oriented matrices and employing (what can now be considered dated) research on frequent error analysis,

Fernandez-Ballesteros and Calero's (2000) model is hardly unidimensional in nature. The following depiction is perhaps a better representation of what the model is all about:





Priding itself on its offer of both qualitative and quantitative information, the EPA secures for itself evidence of both near and far transfer applicability, although this is contested in section 5.2.9.8 below. It has also secured for itself a well-rounded theoretical foundation in terms of its approach towards inducing change.

5.2.9.1 A(i) Ontology

Humans are emphatically viewed as change-based in the model. The assumption of transference attests to the nature of malleability via learning. Important aspects considered within this change-based model of intellectual functioning include the rationale behind the choice of items used, the types of training that are most suitable to the modes employed and the choice of a subsequent criterion of learning. The governing reason behind training in the first place is to induce change, a construct in need of definition in most dynamically assessed tests. The authors of this model are explicit in their efforts to determine the nature of change brought about and highlight three aspects, namely, evidence that training is effective, durable and transferable to other learning contexts. The manner of going about assessing for these concerns is easier said than done. Ontologically, humans are endowed with the capacity for change. It is the duty of the assessor to find this construct and produce it or its affiliated measures.

5.2.9.2 A(ii) Philosophy

The EPA manages to balance the need for representation and idiographic concern. The individual is paramount in this model but cannot be dislodged from the larger entity within which learning takes place. That learning takes place at all is an assumption running throughout the method and battery employed although cognisance is taken of the extent to which such change can be effected. The categorisation of learners into gainers, non-gainers and high scorers is an old notion dating to Budoff's initial research classification system. The model is also predicated on notions surrounding the nature of change as stable, consistent, specific and significant which thus reflects back on the underlying philosophical treatment of the individual. As such change is induced during the learning situation and micro teaching-learning situations become the context in which assessment takes place. Learning potential and change are not synonymous constructs in this model as "learning potential could be understood as a conceptual network" (Fernandez-Ballesteros, 2000, p.294). This demarcation of learning potential as domain status as opposed to singular construct is unique among the models discussed in this chapter. Learning potential becomes a methodology employed to elicit change in change-inducing scenarios and not a singular measurable entity, although as manifest construct, this is precisely what becomes of it at the end (posttest minus pretest score). The EPA is a tool and change is brought about via the training programme, so it can be stated that it is through training that change is measured and not through the task itself. This highlights the model's concern with meaning-making constructs which in this instance is methodologically induced through teaching and not brought about via novel construct creation. The model lies conceptually close to the nature of change-enablement as understood through Vygotskian and Feuersteinian dynamic assessment. The Feuersteinian programme is detailed and lengthy and the EPA (also utilising the Raven's) can in a manner of speaking be referred to as a mini-instrumental enrichment programme.

5.2.9.3 B(i) Hypothetical terms

Perhaps the most identifiable mentalistic term embraced within the model that is not identifiable as a term per se is the conceptual network represented by learning potential and its methodological implications for assessment scenarios. The concern within this model is not so much the type of mediatory feedback as much as the fact that change is ineluctable in the first place. Change is garnered from approaches towards assessing for it. Organismic terms are found within the qualitative assessment and are more cognitive in nature referring to cognitive processes taking or not taking place as the case may be and are therefore directive in function, however more behavioural manifestations are contained such as impulse control. As has been noted thus far, the more standardised and generaliseable the model the more replicable it is. This results in terminology which is more grounded in extant reality to the extent that it can be measured more specifically but runs the risk of measuring a construct that has been borrowed from prior research. Meaning-making novel constructs are eschewed in favour of consistency but this does not necessarily indicate the impossibility of measuring meaning-making novel constructs. This has and can be accomplished but is far more lengthy a process and accomplishes far less in terms of identifiable measurements as end product.

5.2.9.4 B(ii) Scientific hypotheses

The EPA is an assessment tool making use of a specific method of intellectual assessment via matrices testing for inductive reasoning. Various sub-componential matrix functions are delineated and assessed for within the inductive reasoning assessment. Hall-mark intelligence functioning is manifested through the facility of transference although the authors ensure that such assumption of transference is in fact just that: an assumption. The nature of the task lends itself to adjudication and replication (an aspect reiterated within the fundamentals of what is considered a science in general) due to the specific constructs assessed for in each matrix. The assessment is also standardised ensuring similarity of procedure yet is tailored to each individual in terms of immediate feedback regarding the subsequent choice of strategy, suggestions for possible solutions and feedback about the correct choice of answer if incorrect answers are provided. Qualitative yet standardised analyses of



answers are provided and include error analysis, analysis of information, answer modality and the appropriateness of the answer. The following hypotheses can be tentatively identified:

Theoretical (hypothetical) scientific hypotheses

- H (inductive reasoning) - H (intelligence)
- H (learning potential) - H (change; simplistic dated model of posttest minus pretest score. Learners are however categorised into one of three groups based on novel use of descriptive statistics)

Empirical (substantive) scientific hypotheses

- H (learning potential) - S (cues)
- H (learning potential) - S (feedback)
- H (inductive reasoning) - S (gestaltic; completion and superimposition)
- H (inductive reasoning) - S (gestaltic and analytic; constancy and closure)
- H (inductive reasoning) - S (2x2 analogies; constancy, closure and orientation)
- H (inductive reasoning) - S (3x3 matrices; increase, decrease, movement and contraction)
- H (inductive reasoning) - S (3x3 matrices with combination of elements; as above with combination)
- H (inductive reasoning) - S (3x3 matrices; addition, subtraction and combination of above)
- H (learning potential - improvement in performance) - R (generation of strategies)
- H (learning potential - improvement in performance) - R (approach to task)
- H (learning potential - improvement in performance) - R (self-regulation)
- H (learning potential - improvement in performance) - R (error type; incomplete correlates, erroneous reasoning principle, confluence of ideas, repetition and persistence)
- H (learning potential - improvement in performance) - R (analysis of the information; impulsive or non-impulsive)
- H (learning potential - improvement in performance) - R (answer modality; graphic, gestural, verbal and anticipatory)
- H (learning potential - improvement in performance) - R (appropriateness of the answer; correct answer, spontaneous correction, correction after feedback, correction after training and incorrect answer)

$$\Sigma (H-S) = 8; \Sigma (H-R) = 7; \Sigma (H-H) = 2.$$

Hence HQ for the S-CPT = $2/(7 + 8) = 0.13$. The EPA is highly testable and given the above hypotheses allocation is hardly surprising. The EPA is a sound instance of how both qualitative and quantitative measures can be accommodated within a stringent testable model of change. Although not representative of typical Feuersteinian format, it negotiates a path between standardised technique and individual focus. Replication via standardisation and individual tailoring via one-to-one focus on response analysis is couched within a framework of micro-teaching for change. The validity of assessment for change as a linear subtraction of pre from posttest score however remains debateable.

5.2.9.5 B(iii) Hypothesis system

The intermediary hypothesis system within this model is more in keeping with traditional nomothetic appeals to science practice and as with many such models emphasises the need to consider robust psychometric properties thus ensuring its survival within the larger domain of intelligence assessment. Recall chapter 3's discussion on how science progresses according to dictates not always controllable by the scientific community. The EPA capitalises on prior research on the Raven's, information processing models researched by Hunt and Sternberg, Vygotskian mediation and Feuersteinian models of enrichment. One can almost say that it has a finger in many dynamic assessment pies. The need to ascribe to so many theoretical models could be questioned on the basis of parsimony (see chapter 3). This is not a criticism, merely a note regarding the need to encompass every aspect within dynamic assessment in one model or approach. The criticisms lodged at social science research do include aspects such as these but the model followed by natural science methodology dictates expediency when it comes to terminology and theory exposition. The simplest rendering of the facts are invariably the most alluring from a pragmatic stance but the need to follow in the footsteps of natural science methodology has itself been questioned. However, this model is premised upon standardised enumeration of sub-componential constructs and lends itself to the typical nomological network within which theoretical models are ensconced so the question then remains as to where it should be placed. If the model presents as a seamless fit into conventional deductive models of hypotheses testing then what happens to the social science trajectory carved out as a separate domain of "other" science (assuming that the word 'science' can be used)?

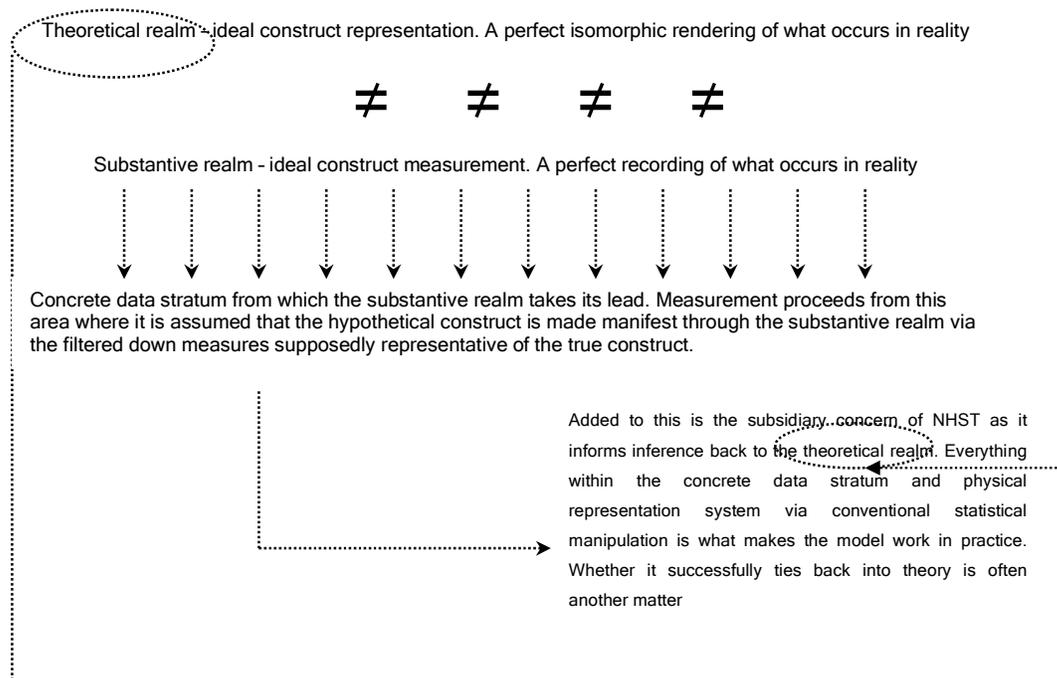
5.2.9.6 C(i) Abstract data

The core function of a concretised model such as the EPA is to provide data which is verifiable, replicable and standardised. Abstract data links raw or concrete data to descriptive hypotheses but does not necessarily carry any hypothetical terminology. The nature of the Raven's being non-verbal for instance, describes the type of data that is captured but does not in itself

describe hypothetical terms relating to the nature of non-verbal intelligence. Other abstract data terms include significant cues, feedback, visual processing and generation of strategies. The data generated hinges around the nature of the training programme as this is where most of the behaviour (mental and otherwise) is observed. Data are captured via a dialogue format which is “oriented to the learning process and aimed at generalization; it is based on shaping, provision of reinforcement and immediate feedback, elicitation of verbalization and detailed analysis of the strategies followed for problem solution established item by item” (Fernandez-Ballesteros & Calero, 2000, p.300). The data obtained are descriptive yet is not confounded by hypothetical terminology and is representative more of functional relations as explicated within the Madsenian framework. Reliability and validity estimates are garnered from concrete data results.

5.2.9.7 C(ii) Concrete data

The numerous studies conducted with the Ravens in an attempt to modify it for dynamic assessment attests to the ease of use within culturally and socioeconomically disadvantaged settings. The division of end-status performers into one of three groupings allows for predictive and comparative studies to be carried out with other static assessment batteries. Via the usual NHST statistical measures, training has proven to be effective and lasting in some instances although transfer of training has yet to be solidly argued for. The realm of evidence presides in substantive reality and as has been discussed thus far, the tentative bridge between empirical and hypothetical reality is at times very controversial in nature due to much speculation surrounding the nature of the constructs. Concrete data forms the foundation of this evidential realm and even though the data stratum forms the “lowest” section on the meta-theoretical ladder, it is this very rung from which much is deduced. Statistical deductions follow on from concrete data only, hence the need to discuss at length in chapter 4 issues surrounding prime considerations within statistics, specifically NHST. The leap from theoretical construct to empirical construct is itself problematic. The leap from concrete result to inference is similarly problematic. Infused within this mix is the nature of qualitative analysis (from which dynamic assessment in most forms will not get away) which merely adds to fuzzy results. The situation is depicted below and is not specific to the EPA.



There is thus not only a link but a direct link between concrete data and hypothetical construct. Along this chain are myriad other concerns which have not yet been adequately dealt with. Causal inference from possibly incorrect use of statistical methods based on data which hopefully ties back into the theoretical realm does not make for a reliable science (assuming that is what assessment strives to be)



5.2.9.8 C(iii) Prime considerations

The most obvious contender for a draw-back, if it can be stated as such, is the model's reliance on a simplistic rendering of a change score which has been highlighted and discussed at length in chapter four. The subsequent classification, however, of performers into Budoffian categories is a procedure which considers mean gain score and how this relates to one standard deviation in the pretest session. Based on this criterion, the cut-off score is utilised in determining gainer status. In order to secure veracity of prediction though, another measure emanating from epidemiological studies is utilised; namely, sensitivity and specificity. In other words how well does the gain score identify those who can truly be said to have significantly changed as opposed to those who have not significantly changed. Correlating gain scores with these two indicators as well as the degree of positive and negative predictive value ensures its use as valid criterion of change. It has evidenced predictive validity of improvement in IQ scores after Feuersteinian instrumental enrichment. This notion is still difficult to conceptualise and as is the case with the Flynn effect, increases in supposedly stable constructs makes for an uneasy academic stance on the nature of the construct "IQ". The authors stress the generalisability of various components within the model that need to be taken cognisance of; training, the tasks, observed operations and participants. The discussion in chapter four (section 4.4.2.2) regarding generalisability theory highlighted the role of resolving error into multiple simultaneous 'error-bearing components' and the nature of finer levels of discrimination that results from this. The idea behind such generalising is the need to extrapolate beyond that which immediately confronts the assessor. How well the EPA is able to generalise across and to other task applications (transfer), population groups (elderly, culturally disadvantaged), cognitive domain and various modalities of training is of critical importance to this model and is an aspect not as fully considered as with other models discussed in this chapter. In other words, how stable are gain scores? It is common knowledge that most mediatory interventions do not result in far transfer primarily due to either domain specificity of task (which then logically implies the lack of task similarity underlying other broader tasks) or stability of intelligence measures across the lifespan. The authors are at great pains to provide for the EPA some measure of assurance that constructs assessed for and changed (note the fundamental difference here between static and dynamic agendas) are generalisable across samples and domains. This is illustrated by numerous studies attesting to such attempts, some of which work better than others, or at the very least provide better estimates. Recall the argument for heritability (h^2) and the greater variance accounted for in results over time leading to a genetic explanation for future performance. Likewise the short-term effects are not necessarily durable in their current form. Advancing the notion of predictive validity in IQ score improvement then does not make sense. The argument can be viewed thus:



1

- EPA seeks far transfer capabilities and far transfer reliability

2

- This would mean that far transfer is possible, which, depending on the literature that is read, is assumed highly unlikely. As time progresses the less likely it becomes that trained skills are retained. Variance accounted for is usually due to innate intellectual functioning. Hence, mediated skill enhancement triggers what can only be referred to as superficial skill acquisition. This argument also hinges around the stability of IQ measures which is now known not to necessarily reflect the construct it seeks to measure. To use the terms 'IQ' and 'intelligence' synonymously is to create even greater confusion. Let us rather state the following in number 3 below

3

- Skill enhancement via mediatory interventions which seek to elicit hitherto inaccessible potential may lead to increased predictive validity of IQ score increases. As IQ is not synonymous with intelligence (for the moment let this be a given) it is highly likely that just such a scenario can plausibly manifest. However, skill enhancement via mediatory interventions which seek to elicit hitherto inaccessible potential may not necessarily lead to increased intelligence. And herein lies the crux. The EPA is stated as struggling to effect far transfer as assessed via increased predictive validity of IQ scores but this issue is itself an unresolved minefield. Recall one explanation of the Flynn effect: pure exposure to test items and technology results in an increase in scores. This does not reflect on intelligence as phenotypical response and this argument is buffered from evidence in heritability studies. So perhaps the EPA's authors should change their criterion of transfer attainment and seek to assess for more biologically attuned measures of intelligence increase (if this is possible). The counter argument could then perhaps be that the exercise would be a pointless effort given the unchanging nature of biologically determined intelligence; why bother at all? Who is to tell? But changing the criterion might well be a start. Another argument in favour of utilising the current criteria is that both items in the EPA and items traditionally utilised in intelligence assessment batteries are similar in nature - having as their predicates the foundations of inductive and analogical reasoning

4

The authors also ask whether IQ is a mediating factor in improvements. Once again we are confronted with tautologous arguments which seem to forever spiral away from a solution. If IQ is a mediating factor (and in *this* instance it is assumed that IQ = intelligence) then the arguments can continue as follows:

- The larger the IQ the greater the probability of improvement
- Individuals traditionally sought are those presenting with low IQ levels which, if invoking the argument above, concludes with the statement tantamount to advocating a fatalistic stance regarding low IQ individuals. It is known that such individuals perform well on subsequent dynamically assessed for tasks. So then, IQ is not necessarily a mediating factor, because if it was then it would most likely have played a role in not affecting change scores (as opposed to the increase in change scores evidenced)
- Hence IQ is a mediating factor as it increases the likelihood of increased performance from already high functioning individuals
AND IQ is NOT a mediating factor as improvement can be elicited regardless of IQ result

5

The question of transfer is also related to improvement in functioning as it is assumed that changes elicited are inherently acquired. If this is the case, then change should be permanent and hence transferable across domains. It is possible that items assessed for on both near and far transfer are not themselves good measures of such skill transference and it could also be that domain-specific skill enhancement is simply not transferable across larger domains. See section 5.2.3 on the targeting of domain specificity within dynamic assessment. Linking the concept of transference to the highlights above, the authors state the following: "it is important to ask to what extent strategies trained by means of test items (which assesses a certain cognitive construct) can generalise to the test construct" (2000, p.313). Cognitive task and cognitive construct as well as inferred intellectual construct being assessed for are three separate entities which require unique description. The fact that cognitive task skill acquisition does not exhibit far transfer might be telling of the dissimilar nature of the construct been assessed for across domains and not the assumed similarity! So perhaps our assumptions as to construct stability are incorrect. It could, as mentioned, also indicate the weakness of training for skills over time. When assessment batteries from dynamic perspectives issue statements regarding the degree of transferability and the nature of the construct assessed for, one needs to consider that not only is dynamic assessment saddled with its own unique burdens but is also saddled with greater issues within intelligence research which it unfortunately inherited from "above" or "below" as the case may be

The nature of the task, construct and other related constructs come into play when assessing for transfer. The experimental results from the EPA indicate that transference between tasks and construct are two different concepts altogether, which is hardly surprising given the difference already discussed regarding the nature of the hypothetical construct and the substantive construct. In this example, one could construe the test items as assessing for a substantive construct (whether or not it represents the same hypothetical construct is open to debate) and the test construct as evidencing the theoretical construct. The following is cited as it pertains directly to this very important epistemological and ontological question “[the] EPA training could not be transferred at the level of the construct (even to the same construct) when that construct was assessed by items different from those used in the training program (Fernandez-Ballesteros & Calero, 2000, p.314). A few aspects of note arise here:

1. “even to the same construct” - this is an assumption the authors are making, because it cannot be proven that the construct is ever the same as we have no way of even fully substantiating the hypothetical construct in the first place. Different items supposedly assessing for the same hypothetical construct confound the nature of the defined construct by assessing for other possible latent traits. Strictly speaking and if one wants to be pedantic about the issue, in its most technical form, the argument can state that by merely changing the item’s wording one effects the nature of the construct. We assume, based on measurement dictates (in some instances this can also be said for IRT), that items testing for a certain construct assess for other latent traits over and above that for which it was originally developed. The further one espouses the ideal of far transfer the less defined the construct is going to become. The less fine the construct the less likely one can measure because measurement necessitates definition at varying levels. The fuzzier the definition, the fuzzier the results and the more open to interpretation become the findings
2. far transfer is possible to obtain but is virtually impossible to measure but near transfer is easier to obtain and is easier to measure but becomes less useful if one wishes to extrapolate to a wider area
3. the EPA seeks both near and far transfer but is unable to muster enough evidential support in this regard. This is not a fault of the EPA nor is it the express duty of the EPA to solve this conundrum. This is an issue pervading the broader field of intelligence assessment. The EPA as delineated merely attempts to address this issue and it does so remarkably well given the nature of the near and far transfer tasks it assesses for

The authors state the following regarding a certain EPA study “an important finding of this study remains unexplained: how and why a learning potential test based on a non-verbal task can predict improvements in a verbal intelligence score after a long-term cognitive training program [referring here to the FIE programme]” (2000, p.317). The explanation they offer as to why this may occur is that verbal measures are in fact used when describing the tasks to be completed; in other words, during training, the tasks are verbalised. Another reason could be an underlying feature of the tests: *g*. This is a tentative offer but one which is looked upon favourably by this author. The counter argument could be that regardless of *g*, verbal and non-verbal scores should not be so closely related. But recall the predicates of the tasks - inductive reasoning. Inductive reasoning loads highly on *g*, although not all sub-components of the matrices load as highly on *g* as others. It is being speculated that some sub-components do not assess for inference at all. As mentioned in chapter 2, wherever you go, you are sure to find *g*. One is assessing for the predictive relationship here though not the relation between the two tests. These muddled issues make measurement even more arduous. Measurement is a debated method of assessing for an elusive quality, and added to this difficult endeavour is construct fuzziness and transfer dilemmas. There is yet a path to travel before these issues are sorted but they first need to be acknowledged and confronted - an aspiration of this thesis.

5.2.10 Application of cognitive functions scale (ACFS)

The most outstanding feature of the cognitive functions scale (ACFS) is its concern with curriculum (Lidz, 2000b). Unlike more generic models of dynamic assessment seeking to assess and mediate for deficient functions, the ACFS is located on a level which can be said to be pragmatic in nature. The overwhelming need in modern society for average and above average school performance dictates that children achieve certain levels of educational attainment at certain stages. To argue against this very large system in society is almost pointless, so in a manner it can be concluded that it is better to join the system in order to beat it! Individuals placed at a distinct disadvantage in terms of intellectual, cognitive and educational functioning nevertheless need to fulfil the barest mandate instituted by reigning educational philosophy and as such, skill within the academic domain cannot be disregarded. The alignment between education and assessment is so close that it can be said to be one and the same thing. Education is all about assessment and assessment is (usually) all about education in its varied forms. As the ACFS is curriculum-based it stands to reason that the cognitive functions assessed for will somehow be criterion referenced. Six subscales assess young preschoolers’ mastery and responsiveness to tasks and interventions which illustrate the dynamic nature of the tool and the tasks are reflective of what is required of preschoolers in the United States. It is interesting to note that cognitive skill requirement can be encased in six subscales and links back to the earlier discussion in chapter 3 regarding parsimony within science. To what extent is information lost through delineation of skill at a level seemingly subsuming general cognitive functioning? Or is very little information lost due to efficient scaling down of overlapping hypothetical skill? This is not a question lodged at the ACFS in particular as this can be asked of many test batteries, but is highlighted here as it comes to the fore. To accommodate for behavioural aspects during assessment, judgements regarding interaction with materials making up the battery are translated into ratings which remain consistent across scales. This standardisation of rating allows for

comparisons to be made on an equitable basis which will of necessity highlight the differences on an intra-individual basis. Task-specific behaviour can indicate a multitude of cognitive deficits not picked up on other scales.

The role of teachers within assessment is hardly an issue worth highlighting but is done so here in order to differentiate the differences between curriculum-based testing which is typically the domain of the teacher and dynamic assessment which is typically the domain of the school or educational psychologist. Knowledge of curricula, educational assessment, psychometrics and psychology is warranted on the basis of fair, valid and relevant assessment. Importing notions of learning potential assessment directly into the classroom is quite urgent when the crisis of educational backlog is studied in South Africa. The drawback in this argument for the use of dynamically assessed curricula is that one will by definition be classifying many learners as learning deficient but the ultimate goal is "to remain close to, yet move beyond and below, the content demands of a specific curriculum" (Lidz, 2000b, p.408). Many developing countries are unable to afford even the of minimum of resources when it comes to education so the case for dynamically assessing masses of young children really does become problematic. Nevertheless, Rome was not built in a day. An intervention is placed between a standardised pre and posttest which itself is partially standardised allowing for more accurate assessment but also facilitating lee-way in an effort to entreat more dynamic assessment components. This is yet another instance of a battery's need to remain marketable in a climate demanding psychometric robustness but attempting to remain true to central tenets of dynamic modes of assessment. Similar items are utilised for both the pre and posttests but intervention materials are unique. The rationale underlying the ACFS can be tentatively sketched as follows:

Logical argumentation for the use of dynamic assessment within the curriculum

Premise 1: school tasks need to be mastered in order to function in modern society

Premise 2: failure to do so results in underutilised resources (potential yet to be tapped) and wasted resources (those already ploughed into education)

Premise 3: dynamically assessing for cognitive deficits has evidenced a moderate track record of success

Premise 4: cognitive strategies undergird school-related tasks

Premise 5: if these strategies are identified then it is possible to remediate for them in the school setting

Premise 6: cognitive processes and principles can also be highlighted

Conclusion: by bridging curriculum and dynamic assessment it is possible to assess and remediate those strategies within the classroom ensuring greater likelihood of better educational attainment at school

Logical argumentation for the potential difficulties to be encountered when utilising dynamic assessment within the curriculum

Premise 1: mass education invariably loses students along the way. Not only does "intelligence" aid in the future success of students but adaptability too becomes a key ingredient in success (the two are likely linked)

Premise 2: cognitive strategies underlying school-based tasks may not necessarily be completely known as manifest constructs mirroring the hypothetical constructs

Premise 3: the practical utility of dynamic assessment still precludes the assessment of those most deserving of the technique which results in a paradox

Premise 4: remediating cognitive skill specific to the curriculum may not necessarily be generalisable to the broader cognitive skill realm (this is not the fault of the assessment technique but of educational skill requirement)

Premise 5: if these skill-enhancing tools are not generalisable beyond the school context, the ecological validity of such endeavours becomes questionable

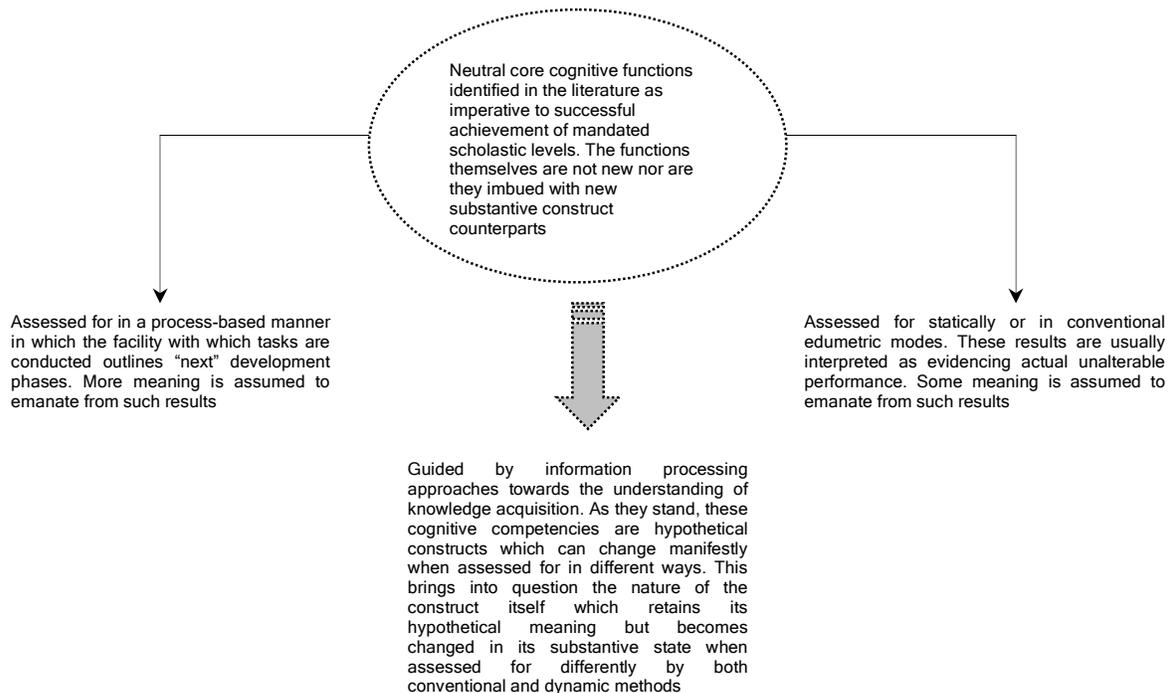
Conclusion: the nature of societal concern for school-specific constructs results in many individuals falling behind, not all of whom can be assisted. Dynamically assessing for such specific skill can aid in better educational attainment at school but the generalisability beyond the curriculum-context remains unconvincing. This is not, however, a fault that can be directed at this manner of assessment

5.2.10.1 A(i) Ontology

The ACFS is one of the more explicit models discussed in this chapter and what is meant by this is the firm contention upheld throughout the discussion as to its model status as opposed to its theory (or atheoretical) status. Lidz (2000b) maintains that the model is premised upon the universal attainment of specific cognitive functions which in the case of school-related tasks ensures progression in the developing child. Reference is made to the link between cognitive functions and intelligence but the ensuing argument frames the discussion on how the two are not synonymous, at least not as it pertains to intelligence. Ontology refers to what can be known and in the case of the ACFS battery processes can be known and potential attainment given the right circumstances can also be known even though there is a limit to which such processes can be assessed and accounted for.

Capacity can be scored and so can strategy, the latter being the more indicative of potential than the former. The conception of the individual is also more explicitly accounted for in the discussion around the battery and although evidencing no direct relation to the neurological underpinning of functioning, Lidz (2000b) acknowledges, through the use of neuropsychological test batteries, that human cognitive and behavioural functioning cannot be disassociated.

The individual is conceived of as a socio-cultural being with physiological underpinnings leading to behavioural (cognitive) outcomes. Yet the undertone is one of acknowledgement for human freedom of action where initial choices are guided by the appreciation of limitations and potentialities both of which undergo change from re-evaluation. Lidz (2000b) is not only mindful of conventional assessment techniques but embraces their use sensing their power in delivering noteworthy results on overall functioning. The coalescence of conventional and dynamic measures affirm the necessity for both techniques to work in tandem in order to conserve what can be useful from both sides of the assessment continuum (assuming that it can be analogised as such). Adequate functioning of core cognitive functions are necessitated by school-based curricula which lead to the eventual attainment of requisite levels of functioning. School-based tasks are well known for their links to conventional IQ assessment batteries and so it can be concluded that cognitive functions underlie some items on IQ tests. However, what is emphasised in the ACFS discussion is not the achievement of adequate functioning but the processes involved in obtaining these levels. It is maintained that not only can process-based assessment tap unexplored areas of actual and “next” development but that it can aid in the facilitation of improved achievement in static-based assessments. Emphasis is also placed on strategy use which is more indicative of how and why certain solutions are chosen over other solutions which in turn leads to more fruitful information pertaining to intellectual functioning. Below is a representation of what process-based assessment of the same cognitive functions can offer educationalists in addition to and as opposed to conventional assessments of these same functions. Note that the cognitive functions themselves are adopted from the information processing framework within assessment.

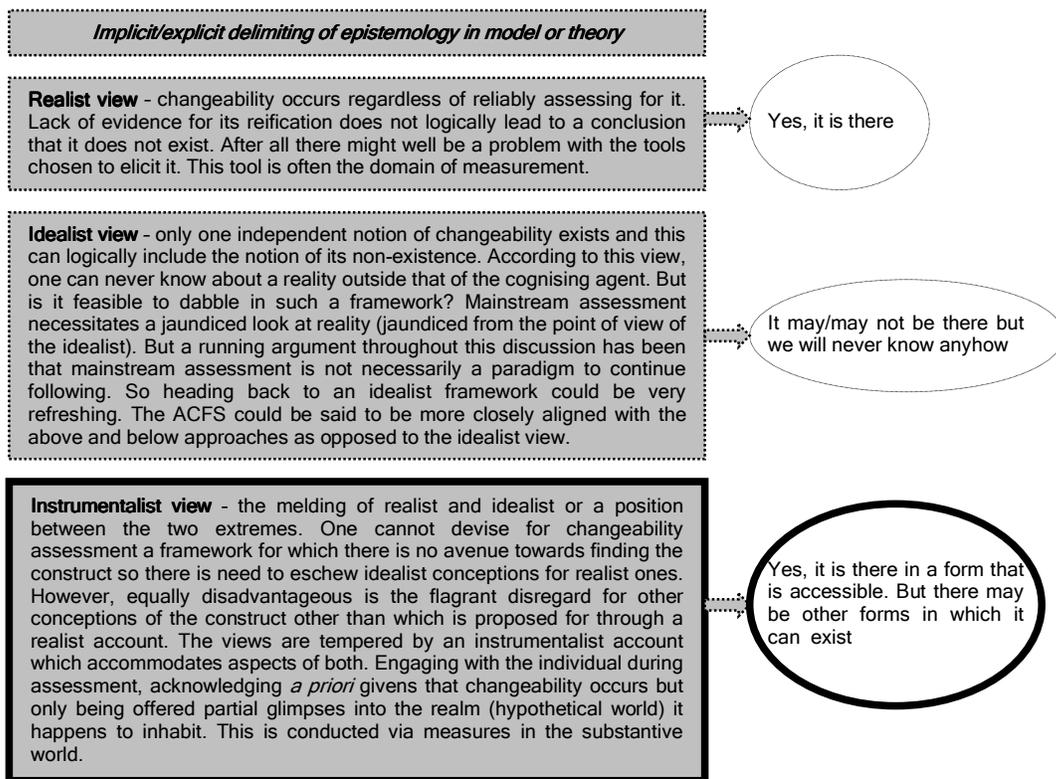


5.2.10.2 A(ii) Philosophy

If one has to position the ACFS in terms of its basic epistemology and how it can elicit what it does, a safe bet may be to describe it as instrumentalist in approach. Taking as ideal the notion of the existence of changeability and reifying results through measures on scales via realism, a hybridised approach towards assessment can be considered as instrumental in essence. In other words, changeability occurs but eliciting this ‘truth’ or given (*a priori* meta-stratum governing premise) is valuable only in so far as it is functional. Functional or utility value of an assessment is the level at which measurement is placed as it serves the agenda of utilitarian frameworks. This issue has been discussed at length throughout the prior chapters but is clearly evident in the ACFS’s premises and operationalisation of hypothetical terms. Madsen’s conceptualisation of epistemological theses and the evidentiary relations expressed in the ACFS can be viewed below. The research methods employed in extracting data typify traditional measurement and quantification and through these methods as well as through the data language used the ACFS is scored according to mainstream tenets. The ACFS does however negotiate a way around pure

quantitative concerns by assessing for metacognitive behaviour and change in a manner which hovers between quantitative and qualitative assessment.

Certain tasks lean heavily on Vygotskian notions regarding engagement of the child with the social environment and the use of symbols suffused with meaning-making sentiments which the child employs in order to make sense of the task. Curriculum-based dynamic assessment as exemplified through the ACFS is a unique modern-day test battery and the idea fits in conceptually well with Vygotsky's approach towards assessment and education. Dynamic assessment is historically linked with educational assessment (Binet, Vygotsky and Feuerstein's grappling with how best to deal with backlogged children in educational environments). It is then perhaps odd that an approach such as the ACFS has taken quite so long to make its appearance in this domain even though there are other instances of curriculum-based dynamic intervention studies. A possible reason as to why this may be so, is that dynamic assessment is often encountered in situations where intelligence is measured. Sooner or later the link becomes entrenched and the move away from intelligence testing towards mediatory interventions aimed at changing cognitive functions becomes more difficult. Lidz (2000b) as much as admits this when she refers to the battery as not being an intelligence test. This focuses the argument once again on the need for dynamic assessment to move away from the realm of pure intelligence-related assessment and into another more profitable realm where it can be practised unabated by concerns now plaguing the intelligence field. That it can be deployed in intelligence settings is not questioned but to place it there exclusively is perhaps not the most feasible step to take.



5.2.10.3 B(i) Hypothetical terms

Both this stratum and the accompanying two below form part of the hypothetical stratum which is where theory comes to the fore. Theory, as now known, is the best educated guess serving at any particular time until a "better" theory comes along to replace it. What is deemed "better" is contentious and open to debate and this is best illustrated with the example of what constitutes a parsimonious theory - an issue yet to be resolved. Multitudinous explanations exist as to what exactly intelligence is and how it can be assessed. One avenue to pursue would be via cognitive tasks which utilise cognitive structures which overlap quite considerably with intelligence constructs. In keeping with Madsen's notion of hypothetical terms providing coherence to a system of observables for which there is as yet no explanation, the ACFS navigates through cognitive tasks in attempting to assess for change at this level. Ploughing any deeper into a difficult construct results in a move further away from the original idea of what is meant by intelligence. In moving "down" a level, the constructs become concretised and yield measures suitable to current mainstream assessment practices. One task within the ACFS, the short term visual memory task is employed to ascertain the current level of memory functioning in the child but the assessment also ensues at the level of metacognitive awareness for the need to choose certain strategies over others.

There is possibly more to metacognitive awareness of memory skill than what is typically assessed for here, but understanding that two levels can be simultaneously assessed provides not only novel information in terms of mainstream assessment but also concretises a concept seemingly more difficult to quantify. Metacognition was highlighted in chapter two as playing an ever-increasing role in assessment literature and is well on the way to becoming a mainstay within intelligence assessment as a whole, seeing as 'meta' governs cognition first and foremost. Metacognitive awareness can be assessed as surely as cognitive awareness, so the retort to this might well be that there is hardly much difference philosophically speaking in how these are measured. An argument can be made for the suitability of metacognition as indicator of learning potential where cognition is used as indicator of learning and herein lies the hypothetical terminology redolent of theoretical constructs yet to be substantiated via empirical results. Most hypothetical terminology used in the ACFS reflects mentalistic (H_m) and organismic (H_o) terms as originally highlighted by Madsen and reflects mental states as well as mental processes of which there are numerous examples. These include terms such as grouping (H_m); communication (H_m or H_o); retrieve (H_m); detect (H_m); planning (H_m); self regulation (H_o); persistence (H_o); frustration tolerance (H_o); motivation (H_o); flexibility (H_o); interactivity (H_o) and responsivity (H_o). The latter half is the behavioural repertoire assessed for in the qualitative rating scale whilst the former reflects the actual tasks themselves which is quite interesting. This could indicate that task-orientated terms are more consistent with mentalistic hypothetical terms and behavioural oriented terms are more in keeping with organismic terms. These terms can be construed as theoretically laden but become manifest via operationalisation and subsequent measurement and link metacognition to potential. The ACFS is one among many models that employ hypothetical terminology as above but it is a particular example of how strictly Madsen's framework can be used to distinguish between varying types and levels of hypothetical terms.

5.2.10.4 B(ii) Scientific hypotheses

The ACFS is predicated upon process-based assessment and as such emphasises qualitative behavioural ratings which are not normative but are in fact scored in order to offer a semblance of comparison to pretest performance as well as to offer information about how cognitive functions change over the duration of intervention and assessment. Being linked to curricula itself necessitates a re-look at scoring methodology as aid as a diagnostic tool which is the primary function of dynamic assessment. HQ determination was conceived of by Madsen as an index of assessing testability of theory or model. His assumption was made after the fact; in other words original HQ determination resulted from prior assessments of what most theories encompassed, namely testability. Along with testability comes quantification. Process-based assessment is not limited to qualitative investigations but quantifies behavioural patterns thus offering the opportunity for HQ determination. The number of scores produced from the ACFS is numerous, due to their construct repetition (a score is assigned to the same behavioural dimension on both pre and posttest as well as across tasks). As with all the models assessed for in this section, the decision to accord certain hypothetical constructs theoretical or empirical status is a decision based on interpretative readings of these models and as such could differ from other readings. HQ results are based on the initial attempts to secure for the formula the requisite constructs and their associated meanings with the numerical index as resultant score. Theoretical (hypothetical) scientific hypotheses are deduced from "basic cognitive processes" as well as "learning strategies" (Lidz, 2000b, p.411) and reflect hypotheses which appear more concretised than some hypotheses illustrated in other models. The H-H hypotheses therefore tie in closely with their associated H-S and H-R hypotheses as can be seen below.

Theoretical (hypothetical) scientific hypotheses

- H (cognitive processes) - H (specific curriculum objectives)

Empirical (substantive) scientific hypotheses

- H (basic cognitive processes/learning strategies) - S (classification)
- H (basic cognitive processes/learning strategies) - S (perspective taking)
- H (basic cognitive processes/learning strategies) - S (short term auditory memory)
- H (basic cognitive processes/learning strategies) - S (short term visual memory)
- H (basic cognitive processes/learning strategies) - S (verbal planning)
- H (basic cognitive processes/learning strategies) - S (sequential pattern completion)
- H (metacognitive behaviour) - R (self regulation)
- H (metacognitive behaviour) - R (persistence)
- H (metacognitive behaviour) - R (frustration tolerance)
- H (metacognitive behaviour) - R (motivation)
- H (metacognitive behaviour) - R (flexibility)
- H (metacognitive behaviour) - R (interactivity)
- H (metacognitive behaviour) - R (responsivity)

$$\Sigma (H-S) = 6; \Sigma (H-R) = 7; \Sigma (H-H) = 1$$

Hence HQ for the ACFS model = $1/(6+7) = 0.07$. The ACFS is tightly interwoven into the curriculum it seeks to tackle and due to the static nature of curricula, the very low HQ result is not wholly unexpected. Behavioural ratings are included in the HQ system but as Lidz (2000b) states, if needs be, the interventions need not follow so closely the guidelines offered but can be attuned to the individual in context.

5.2.10.5 B(iii) Hypothesis system

This stratum operates as a type of mediating link between governing meta concern and data-based lower strata where the wheels of science turn. The HQ ratio as evidenced above immediately primes one to consider for the ACFS a nomological hypothesis system as the system employed can be repeated with various individuals. However, its concern with idiographic explanation is emphasised by the manner in which the individual is targeted specially in terms of cognitive functioning. This statement does not comment on the explanatory system as such as it pertains to deductive and inductive reasoning but rather how it conceives of its subject matter and subjects. Seeking to gain entrance into the mainstream arena yet tempering this with concern for the uniqueness of the individual the ACFS aligns with curricula presently deployed in mainstream early education. This is the framework in which the battery is housed. The model is premised on cognitive functioning as opposed to intelligence predicates as the former underlies much of what makes intelligence what it is. Lidz (2000b) hypothesises that in combining two unique approaches towards assessment (process and curriculum-based) the nature of what the child can do will surface over and above the tasks that the child can or cannot already accomplish.

What can be done (proximal) and what is already correctly or incorrectly done is assessed through knowledge of strategies to employ and not only the chosen strategies that are employed (metacognitive skill). Awareness and assessment of metacognition is an *a priori* given (i.e. the reigning meta framework) and the measurement of the construct is the data concern. Metacognition is, of course, only one feature assessed for in the ACFS intervention. Lidz (2000b) does state that assessment is hypothesis driven which can indicate that already-formed deductive explanations are at hand before the assessment becomes dynamic in nature; typically towards the end of an assessment programme. Being hypothesis-driven the model tacitly acknowledges that certain premises are held which are resolved upon further investigation or assessment. If one were to lodge the type of explanatory system on a two dimensional continuum (which the author is fearful of doing given the critique of dichotomies), the ACFS could find itself positioned at the half-way mark as it shares concerns at both ends.

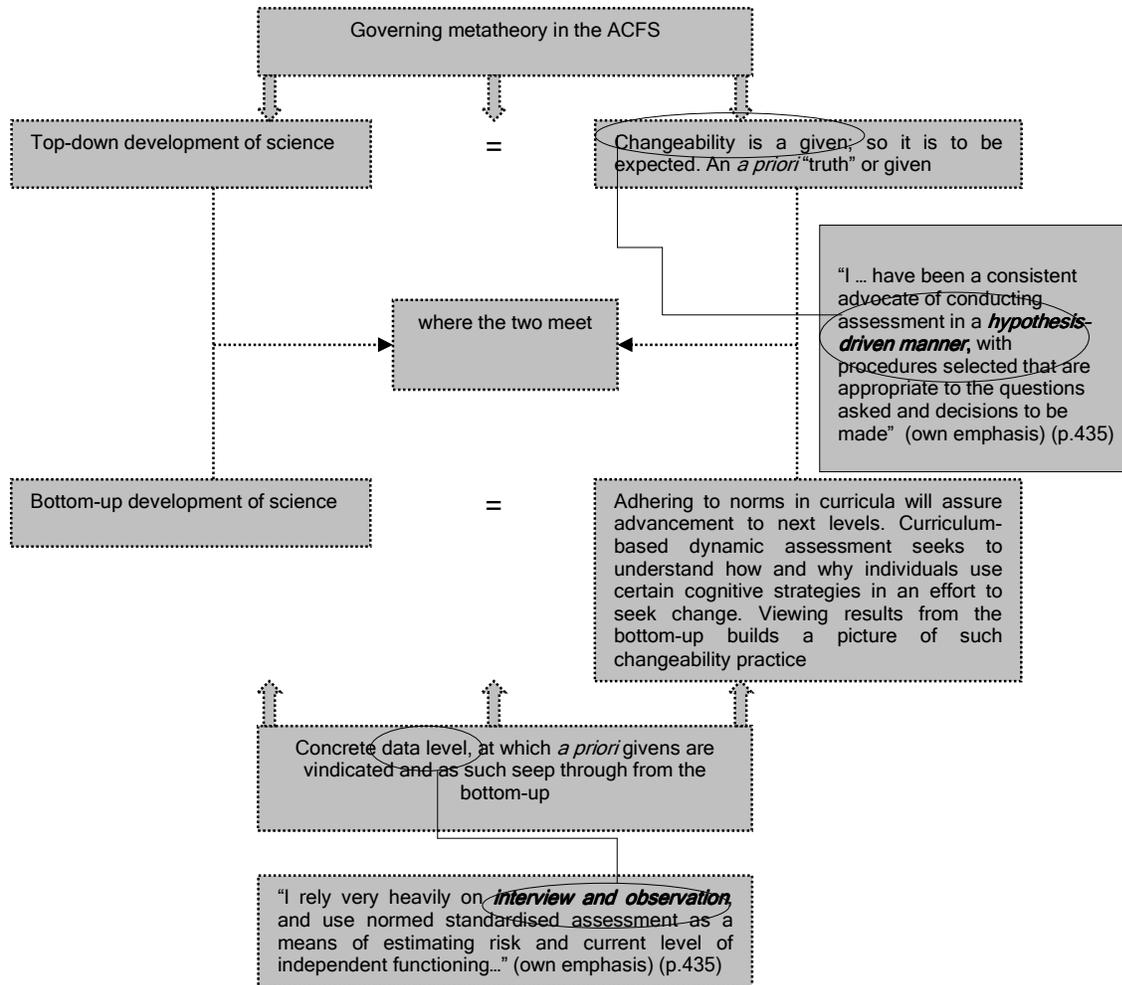
5.2.10.6 C(i) Abstract data

Lidz's (2000b) delineation of abstract data theses is representative of one of the more grounded attempts within a model attempting to link abstract data within the hypotheses system and utilises no hypothetical terminology. This is, strictly speaking, what Madsen had in mind when he conceived of this level's characteristics where empirical relations are enmeshed with data theses. Lidz (2000b) refers to Madsenian abstract data theses as processes. These processes become the H-R hypotheses in the HQ determination above. The differentiated nature of abstract data and hypothetical terms is evidenced by the fact that no hypothetical terminology is utilised within the data theses. In ascertaining the level of cognitive functioning, sub-skills are necessary for successful completion of items. These subskills are translated into data theses which means they employ abstract terminology but do not employ abstract hypothetical terms themselves; the processes in turn are considered as hypothetical terms. Each process is accompanied by three to six subskills (data theses). For instance, objects need to be grouped according to abstract features of the system during the classification task and data theses include detection of features, response to directions, grouping along lines of classifications and across types of objects. Perspective-taking assesses, among other features, behavioural reactions to other people (typically the assessor) as seen from this view point and also assesses verbal and non-verbal behaviour. The difference between an abstract data thesis such as "perspective-taking" and an hypothetical term such as "inductive thinking" is the concretisation of the former. A retort to this statement could be that inductive reasoning is reality concretised via measures seeking to assess inductive thought. However, the former is easier to concretise via Lidz's (2000b) criteria which are set out below each assessed process. The ACFS behavioural rating scale features are less concretised due to their more qualitative features, yet according to empirical data gathered on the ACFS, inter-rater reliability appears high enough to warrant adequacy of behaviour rating. The behavioural scale criteria nevertheless form part of the HQ determination and cannot be judged as data theses. In accommodating for data theses which link back to their hypothetical term counterpart, Lidz's (2000b) model encompasses testable features alongside qualitative features, making it amenable to measurement and meaning-making, although it is more skewed towards measurement at this stage. The data theses are provided in both functional and correlational manners. Functional relations can be seen in all the process-based tasks and correlational relations are illustrated in the behavioural rating scales where metacognitive aspects are rated across tasks allowing more general statements to be made concerning certain metacognitive skills.

5.2.10.7 C(ii) Concrete data

Curriculum-based assessment is by nature data driven as students need to fulfil certain prerequisites in order to be considered eligible for assessment at higher levels of functioning. Having said this in no way precludes behavioural interventions and dynamic assessments which seek to alter cognitive functions but these too are scored according to conventional dictates. The

meta-level concern in the ACFS is identifiable on the basis that it seeks to remediate for cognitive backlogs in the school environment and does not assess for intelligence. Garnering evidence to attest for changeability in a zone other than which is tested for in school bespeaks of the ACFS's concern for static-based assessment reality and it can successfully meld with proximal development endeavours. The scoring structure allows information to be gathered which can make manifest the zone of next development in a manner not seen in typical school settings. Madsen was of the opinion that science proceeded on the basis of top-down development but the ACFS illustrates that dual concerns of both top-down and bottom-up development of theory and model meet midway as shown below.



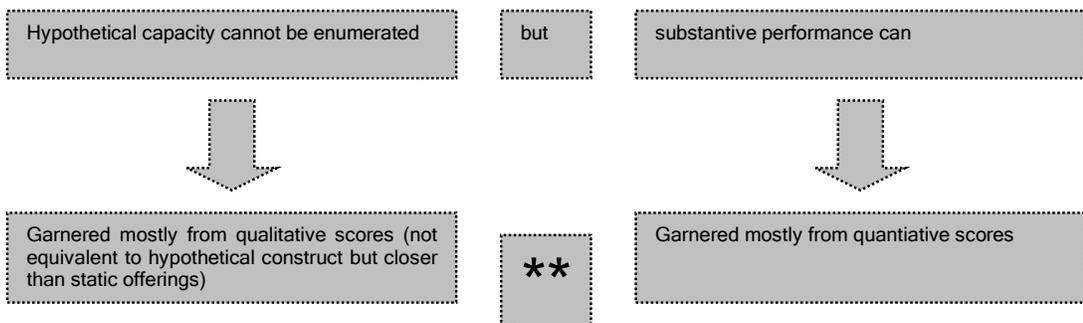
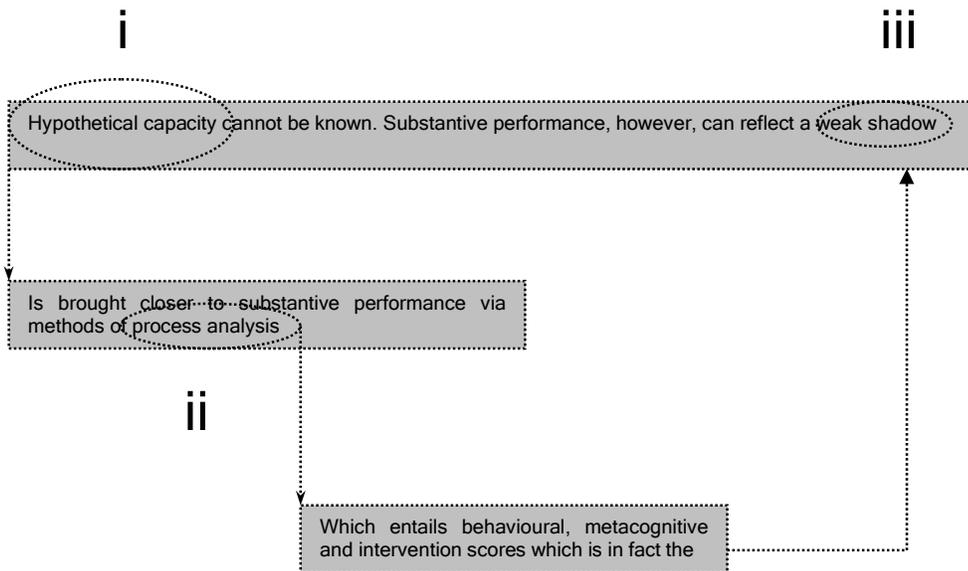
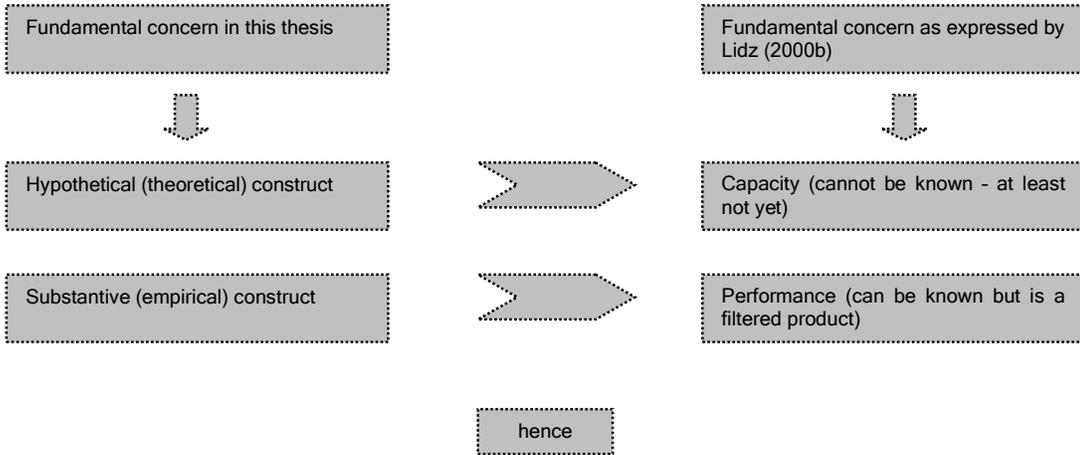
5.2.10.8 C(iii) Prime considerations

The ACFS accounts for both quantitative and qualitative assessment and yields both numeric and qualitative scores. The scores in and of themselves are not as meaningful when considered out of context as they reflect interaction with the task and not the result of interaction and herein lies a main philosophical focus. The tasks vary and due to the stability of criteria across tasks the resultant scores offer a base of comparison highlighting difficulties encountered with each task as well as the nature of the difficulty. The emphasis as always within dynamic assessment is to understand what mistakes are made, how they are arrived at and why they occur hence the normative test results. Below can be seen the tabulated rendering of the ACFS scoring.

Quantitative scoring	Qualitative scoring
<ul style="list-style-type: none"> • Use of raw scores and percent mastery scores on both pretest and posttest items across six tasks • Decision as to what to measure based on “rational task analysis” and observations • Rational task analysis is based on the requisite skill for each task in order to correctly solve it • Error analysis • Cross-task comparison • Approach towards task • Reaction to assessor intervention • More diagnostic approach allows for deviation from semi-scripted intervention • Scores for both task completion and ability to respond successfully to intervention if required • Testing the limits can also be utilised as method if so chosen • Ceiling level is inherent but ACFS it not a battery to be utilised by high functioning performers • Scoring <ul style="list-style-type: none"> ○ Score for accomplishment of each task ○ Pretest score (6 scores) ○ Posttest score (6 scores) ○ Gain score (near transfer) score calculated in the traditional manner of pretest subtracted from posttest (6 scores) 	<ul style="list-style-type: none"> • Initially based on observations of children interacting with tasks and agreed upon after inter-rater analyses • Score apportioned to utilisation of metacognition • Behavioural rating scores are given along six to seven dimensions and ranked according to the occurrence of each dimension. Each subscale is thus rated twice (during the pretest and after intervention) on these dimensions which results in 36 (pretest) and 42 (after intervention) respectively. This totals 78 scores • Summary scores for all pretests and posttests, total scores for each behaviour across each task and total scores for the behaviour rating scale • Note that each behavioural dimension is scored not only per task but can be successfully compared across tasks so that use of flexibility for instance can be monitored not only on one occasion and not only on one task but for all tasks. What does this indicate about behavioural functions? Are reactions task-specific or general in nature? The answers to this can yield a wealth of clues about domain-specific and domain-general cognitive tasks and the role this plays within transfer tasks. If individuals present themselves in a very flexible manner on sequential pattern completion but not for short-term visual memory, what does this reflect? Is it reflective of task more so than of the behaviour? The between and within task analysis fits in well with the underlying rationale of the ACFS, notably, “projections about the child’s ability to profit from instruction [are] to be made from evidence that is grounded in experience rather than inference” (Lidz, 2000b, p.436)

Lidz (2000b) is adamant that static assessments, which are normed and standardised, should be used within a setting necessitating diagnostic results in addition to the complementary findings resultant from dynamic assessment batteries which emphasise process and not product. The additional information cannot be “harmful” and only beneficial towards mediatory interventions. In sum, the ACFS is a battery which functions within a process-based model and although does not give to dynamic assessment novel constructs, seeks to assess for conventional constructs in novel ways (assessment of common behaviours within and across tasks). The reason behind the lack of meaning-making novel constructs could be Lidz’s (2000b) assertion that true abilities can never be known as one works with applications of cognitive functioning and not the nascent ability. This is critically important in this model’s discussion when it is assessed from the meta-framework employed. The ongoing discussion in this thesis regarding the definition of words and meanings highlights the controversy surrounding the hypothetical and substantive construct.

Although Lidz (2000b) does not refer to the hypothetical and substantive construct dilemma it is evident in her thoughts when the difference between performance and capacity is pointed out. This is illustrated below. For instance, it is not capacity for memory per se which is sought but the use of various strategies that are sought. What is the point in demonstrating poor performance when poor performance is already known? This nullifies the reason for assessment in the first place. At-risk individuals are at risk already, their status is known. The point is to aid in their return to normal status which can only be effected via interventions which in the ACFS is undertaken at the cognitive level which is not necessarily reflective of intelligence. Once again, the domain is critiqued in its unclear definitional status surrounding intelligence. The assessment of cognitive tasks is indicative of certain aspects within intellectual functioning but cannot be necessarily equated with the concept of intelligence (this being partially due to the unfixed nature of what intelligence refers to). Nevertheless, a case can and has been made in other instances for the use of numerics in determining levels of attainment for certain cognitive tasks, if only to place results on a normative scale. Lidz (2000b) evidences early yet impressive results attesting to the ACFS’s construct validity and the facility of the battery to differentiate between children with and without developmental disabilities as well the high correlation found between the behavioural rating scale and task competence. The semi-structured scale has evidenced good inter-rater reliability suggesting that the guidelines along which to establish metacognitive scores are equally understood across assessors.



** See table above

5.2.11 Analogical reasoning learning test (ARLT)

This dynamic assessment is aimed at a particular designated sample, namely, moderately mentally retarded individuals (Schlatter & Büchel, 2000). Dynamic assessment often comes into its own when differentiation is warranted on the basis of learning disability, retardation, poor cultural mediation and the like. Individuals whose poor performance is based on these predicates are often misclassified. Some evidence poor performance due to variables other than those accounted for by traditional assessment which in turn usually shows floor effects; i.e. all respondents seem to perform at very much the same level, namely low levels thus not allowing for sufficient discrimination between such individuals. Latent traits are not assessed for in detail in such a manner as to allow for finer discrimination. Note that the premise of this approach and test battery is firmly lodged in the understanding that poor performers do not necessarily harbour great intellectual potential but it does state that poor performers' functioning has been homogenised to such an extent that the situation is treated as hopeless. This is not necessarily the case. As language poses special problems with this target population, it is anticipated that interventions will be cognisant of this hurdle. Thinking as tool is emphasised in Vygotskian theorising as it serves as link between thought and world (language). Concepts are dealt with in the world through language and private speech after which the transition is made to the inner world of thoughts. Attention problems are another aspect of behaviour affecting this group and as has been evidenced in other models and approaches, dynamic assessment can alter the behaviour-based responses via varied techniques allowing for the truer picture of intellectual functioning to burgeon. Short-term memory is also a variable within commonly understood variables representative of typical mainstream assessment. As part of the cognitive approach towards the understanding of intelligence, cognitive tasks assessing for such underlying constructs come to the fore and include inductive reasoning task performance.

The fact that short-term memory functioning is poor, does not only reflect poor memory storage and retrieval but points to what lies behind the construct (lack of various strategies designed to promote memory; there are a number of such variables). If, for instance, short-term memory (which one must recall is not the same as working memory, although the authors do not make this distinction) is a main propellant of intellectual functioning on many cognitive tasks (i.e. short-term memory does account for many underlying task performance variance) then it can be said that any ability "lying beneath" this process will be unable to surface adequately, if at all. The counter-argument to this could be that short-term memory is by nature transfixed and intimately bound up to ability and the two cannot be separated. The aim then, via dynamically assessing for this, is to endeavour to look beyond short-term memory and plough beneath to other underlying constructs. By paving a path towards these constructs one need not conceive of short-term memory as sole conduit of intellectual manifestation. Flexibility is another variable or construct noted in this population for whom it poses a particular problem. There is an underlying information processing feel to the discussion on the use of cognitive functions by the authors. Transfer, both near and far are crucial indicators that strategies have indeed been mastered and can be adapted to other areas of similar strategy employment. Notably, analogical reasoning is once again cited as main cognitive function involved in the ability to successfully transfer from one mode to another. As with other approaches, this assessment is premised upon cognitive education theory and information processing and incidentally, the authors categorise dynamic assessment along with these two approaches which indicates that they have tacitly implied that dynamic assessment can be possibly construed as construct innovator in the sense of bringing to the assessment a novel construct not assessed for in mainstream assessment. This is a debateable statement, but according to the division described in section 5.2.5.2. in which a dynamic assessment approach is discussed, this statement can be defended. This is another instance of where dynamic assessment finds itself immersed in what can be considered "minor" technicalities but it is argued that these very technicalities are what has resulted in dynamic assessment's quandary in the first place (see chapter 1). A citation is warranted:

"Our research is based on a variety of empirical and theoretical studies in domains such as dynamic assessment, cognitive education, information processing, and others" (Schlatter & Büchel, 2000, p.159).

Hence, it is not entirely incorrect to assume that "cognitive education" as domain and "dynamic assessment" as domain are treated as equivalent categorizations. They are lumped together into one seemingly greater domain of construct meaning-making. The authors do not single out dynamic assessment as *method or approach* but prefer to construe it as a *bona fide* model consisting of its own constructs. What might these constructs be? There seems to be confusion as to this differentiation in the literature. One can also view the situation thus: is cognitive education a framework housing models and theories or is it a method or approach towards the study of intelligence?

5.2.11.1 A(i) Ontology

Human beings, especially those designated as mentally retarded are offered renewed hope at more accurate assessment and categorisation (if such a terms can still be used). Short-term memory, as implicated in this model, is at the core of why processing of information cannot proceed along the lines more revealing of their actual underlying levels of functioning. Poor short-term memory is also invoked as being caused by mental retardation and not itself causing mental retardation. In many instances, such retardation is organically derived. The conception of the human is more explicitly set out in this model than in some others discussed in this chapter. Cognisance is taken of the chromosomal abnormality resulting in retardation and thus a specific distal feature can be used as explanation in the poor functioning of these individuals. There is no skirting around the issues of hope for extant constructs, there is a definitive reason behind most cases of this sort. Humans are seen as not being assessed in terms of intelligence but in terms of learning and analogical reasoning is cited as playing a central role in *learning*. The motivation behind the assessment can already be seen to be slightly altered in comparison to conventional assessments. The rationale behind this model of approach is the nature of assistance that can be afforded as part of the assessment. This is accomplished in three ways; sequenced hints, scaffolded instruction and mediation including metacognitive awareness strategies as espoused by Feuerstein. The authors have sourced dynamic assessment ideas, notions and meanings from a number of noted luminaries within the field including Budoff, which is not surprising due to his pioneering efforts in seeking finer differentiation among mentally retarded individuals and those not suffering the effects of retardation but being misclassified as such. In essence, the ontological premise is that intelligence is knowable when assessed for in a mediatory environment.

5.2.11.2 A(ii) Philosophy

One can make a case for the appeal of the model as it pertains to rationalism in terms of the vindication of learning capacity as empirical construct which it receives via measurement of its hypothetical counterparts. This is carried out in a process involving scoring over a number of phases for the assessment of potential. Idealist interpretations of learning potential as construct can rightly be said to reflect dynamic assessment's concern with qualitative explanations of functioning which can ultimately never be known. Moving over into the realm of operationalisation where such ethereal concerns become physical reality is what this model seeks to accomplish. It is concerned with both the individual and methodology as the functioning relates back to a standardised method of assessing for this. Idiographic and nomothetic concerns are evident in the plea to better discriminate between mentally retarded individuals who possess the capacity for further development and those who do not. The model is entrenched in the reality of the situation where status as retarded does not necessarily indicate that potential has gone missing but attests to some cases where such potential is resident but which has not been identified as such. The data language is ensconced in scoring of responses and a summary of numerical results builds the case for the attribution of resident potential. The data language is reflected in the HQ score given to this model which is not a surprising result. It can be seen across models in this section that those evidencing higher HQ scores are those with concomitant higher instances of data filled jargon. Does this mean that data cannot be in the form of numericised attribute? This is a perennial issue surfacing within the social science, psychology and more specifically within dynamic assessment.

5.2.11.3 B(i) Hypothetical terms

Both mentalistic and organismic hypothetical terminology is utilised within the model. Learning capacity, retardation, competencies and language as tool for instance are highlighted. Organismic terms are easier to operationalise due to their manifest reality particularly in this population and behavioural features of low-level functioning include metacognitive characteristics which are usually not developed to the same extent as they are in normal populations. Typical manifestations include lack of impulse control, external locus of control (well documented for this population), motivational factors, poor attention and memory span and less as opposed to more self-control in general. Process terminology is utilised during the three phases of the model as the assessment progresses. Dynamic and directive variables are the measurable equivalent of the terms stated above and are thus translated into numerical indices.

5.2.11.4 B(ii) Scientific hypotheses

This model goes to great lengths in accommodating as many aspects of dynamic assessment as possible given the constraints within which it operates. It seeks to not only define and account for various hypothetical constructs but to operationalise them as well through conventional means of numerical assignation. This is not a criticism as almost every model assumes that representational measurement is applicable. This, as is now known, cannot be assumed as is evident in the discussion on measurement within the social sciences in chapter 4. The model is both a qualitative and quantitative effort at building an accurate picture of individual functioning and appears to have covered as many areas of "measurement" as possible. This model, in comparison to some other models discussed in this section, utilises far more existential hypotheses, which, within the Madsenian framework, proffers the existence of hypothetical constructs which are later operationalised. The model also utilises functional hypotheses which serve as links between sets of hypothetical constructs which is evidenced by transfer items. Learning potential as existential hypothesis is defined or made manifest through maintenance and transfer scores; the functional



hypothesis linking the new found cognitive tools to the ability to understand and retain such tools. The following hypothetical terms can be identified within this model:

Theoretical (hypothetical) scientific hypotheses

- H (intelligence; indication of retardation) - H (Language)
- H (intelligence; indication of retardation) - H (Attention problems)
- H (intelligence; indication of retardation) - H (Short-term memory)
- H (intelligence; indication of retardation) - H (Inductive reasoning)
- H (intelligence; indication of retardation) - H (Flexibility)
- H (intelligence; indication of retardation) - H (Analogical reasoning)

Empirical (substantive) scientific hypotheses

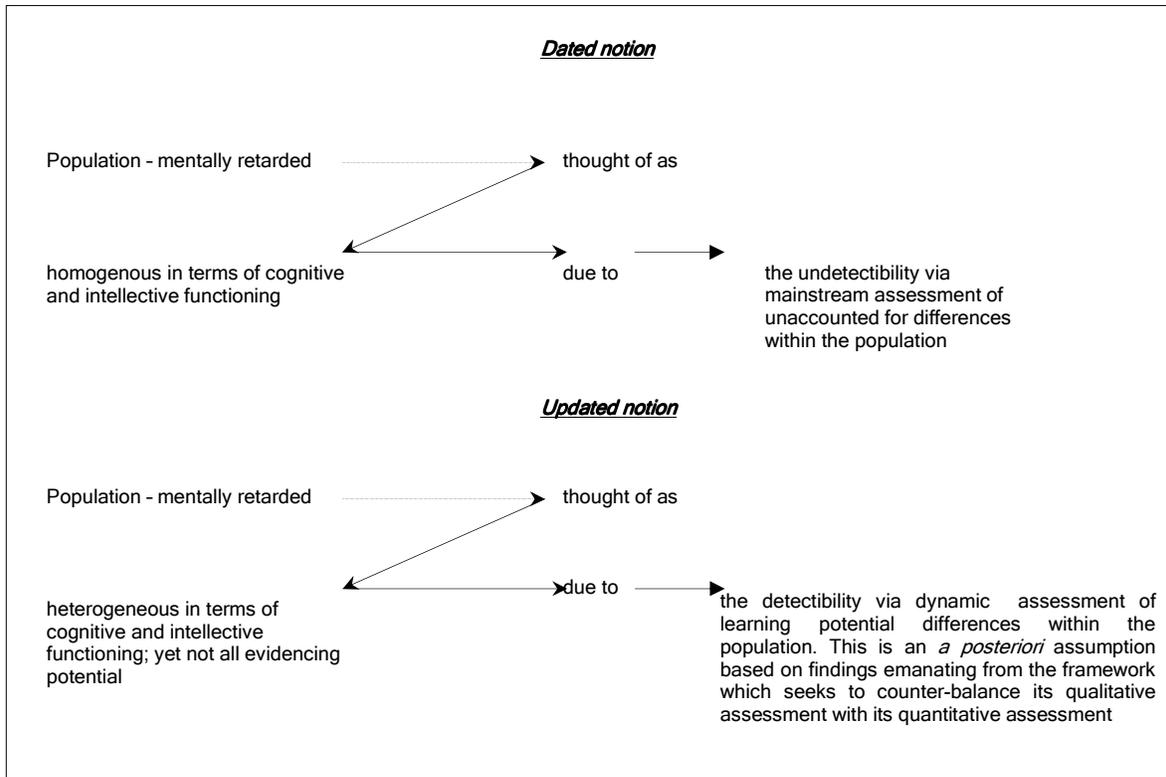
- a. H (Learning) - S(Structured error-related hints)
- b. H (Learning) - S(Scaffolding)
- c. H (Learning) - S(Compensation)
- d. H (Learning) - S(Mediation)
- e. H (Learning) - S(Metacognitive control)
- f. H (Learning) - R(scores utilised for express purposes of aiding in structured hints during learning phase)
- g. H (Learning) -R(potential via end-status category)

$$\Sigma (H-S) = 5; \Sigma (H-R) = 2; \Sigma (H-H) = 6$$

Hence HQ for the ARLT = $6/(5+2) = 0.85$ which is to be expected. The ARLT model is testable in comparison to other mediatory models mainly due to its high loading on empirical construct validation via the scoring processes involved. The link between empirical testability and quantification is brought to light in this model and illustrates how easy it is to link testability to numericised concepts. Is it true to state though that such practices and thoughts are warranted? The discussion of parsimony, theory testability, verifiability and falsifiability bears relevance to this issue as discussed in chapter 3. The pressure to conform to nomological modes of conceiving and practising a science is felt in such instances. The need to prove its scientific status forces a compromise between legitimate dynamic models and conventional assessment models.

5.2.11.5 B(iii) Hypothesis system

If capacity is not assessed for during mainstream testing, and there is yet evidence that the population (mental retardates) does not present homogeneously and if dynamic assessment is able to elicit unaccounted for variance, then it can be stated that capacity can be sourced. This is one manner in which the model can be expressed and is done so within a deductive explanatory system. Hypotheses are couched within this system and serve to link the logical arguments via hypothetical terms and hypotheses as stated above. The degree of abstraction in the model is limited to verifiable theses and is thus not complicated. The model is more or less a true reflection of what the philosophical hypotheses imply. In this instance, the model is more reflective of a framework according to which its logic ensues. A possible schematic rendering of the model can be posited as follows:



5.2.11.6 C(i) Abstract data

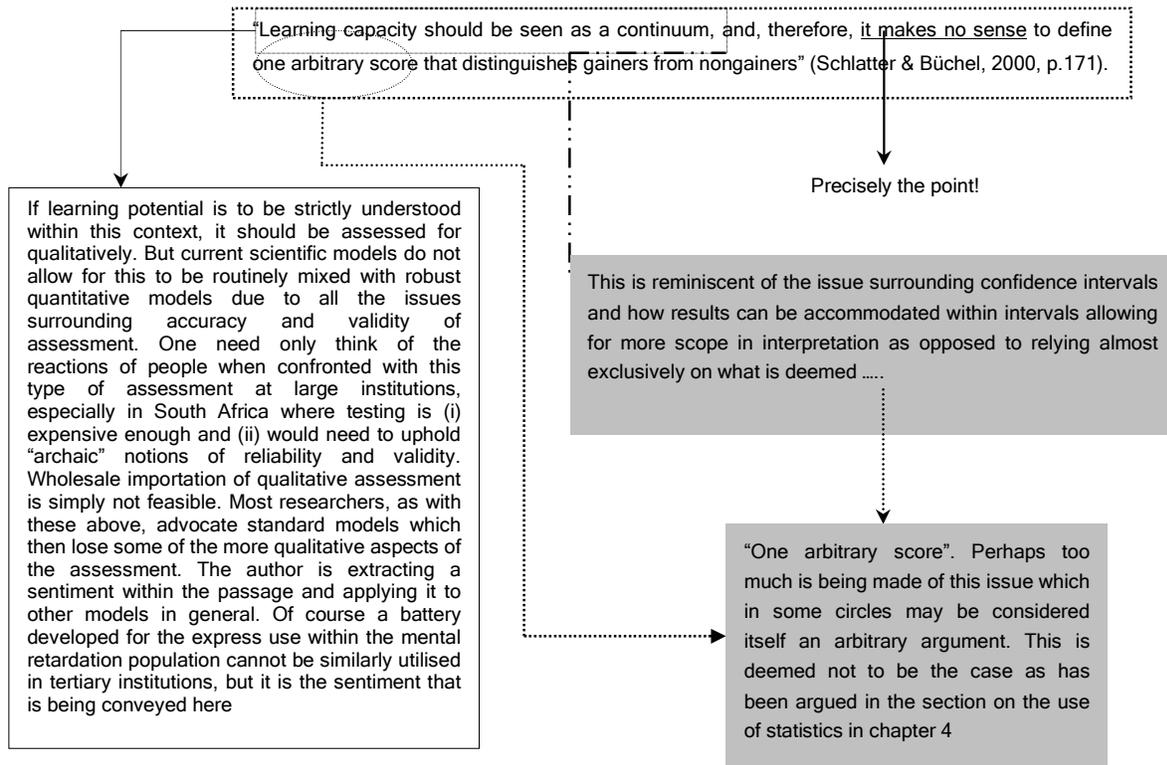
The data stratum houses functional relations between variables more so than correlational variables. The model highlights both observable and non-observable behavioural characteristics especially behavioural traits such as metacognitive aspects. Most of the data stratum variables are accommodated in the concrete data stratum.

5.2.11.7 C(ii) Concrete data

The level of concrete empirical data from which conclusions are drawn and proceed from bottom-up. *A posteriori* assumptions are likewise verified (differing levels of potential are resident in some more so than in others) or nullified (that no potential exists). The system of scoring yields the data level which is in abundance within this approach.

5.2.11.8 C(iii) Prime considerations

Of note in this model and what is particularly striking is the concern for the construct, learning potential, which is construed as a continuum. There is no traditional notion of what can be considered as numerically classifiable “learning potential evident” and “learning potential not evident”. Criteria are nevertheless available which allow for distinctions to be made between those classified as gainers and nongainers which, if assessed for traditionally, would not be highlighted as possessing latent potential and would invariably disappear if not acknowledged as extant. The ARLT has undergone reliability and validity trials in terms of its applicability and usability and has thus employed the usual compliment of statistical techniques for endeavouring to find favourable results. A pertinent issue is discussed below.



A big “however....”

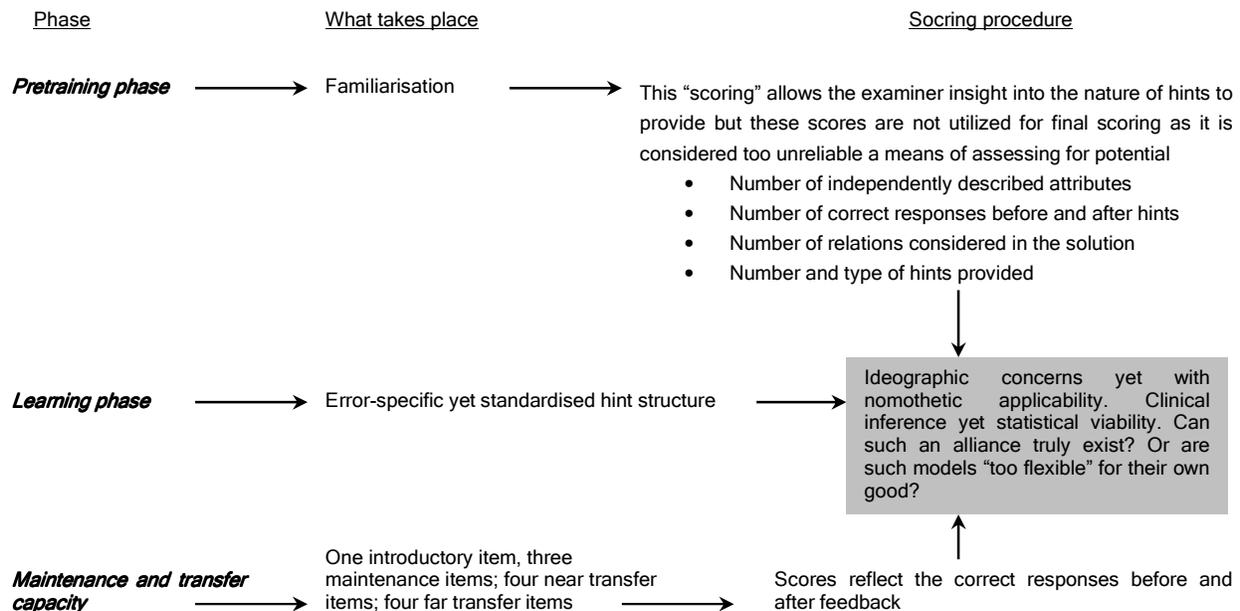
Recall the controversy surrounding statistical and clinical interpretations of data and the manner in which judgements are made based on either of these two approaches. See chapter 3 for Meehl’s arguments in this regard which has pertinence here. On the one hand the following cogent arguments are given in support of qualitative, one-on-one assessments of individuals (the clinical approach):

- i. It offers deeper insight into the nature of the problem and allows for individualised and in some instances tailor-made assessment strategies
- ii. Provides information that would have otherwise have been missed by statistical derivation over a number of cases and added to this is the central concern for the individual and not the spectrum of individuals. Hence, it is logical not to invoke statistical decision-making

On the other hand we have arguments supportive of statistical rendering even in such dynamic assessment set-ups:

- i. In nearly every case to date (see relevant literature – in chapter 3 with heavy emphasis on Meehl) statistical decision-making has proven to be more robust, reliable and accurate when compared to human judgement. This truly is factual and cannot be denied on the evidence alone
- ii. This leaves decision-making as a whole in questionable status, robs human judgement of its often at times insightful arguments and in general does not bode well for dynamic assessment as leveraged on individual qualitative assessment. Yet, this does not mean that such assessment cannot carry on unabated, it merely implies that broad generalisations (any generalisation at all in fact) cannot be made. So, carry on in an uninterrupted manner but keep qualitative assessment to the realm of individual assessment and keep quantitative assessment to its own realm. What does this mean for arguments against NHST? Clearly, the arguments in favour of Meehl’s approach cannot be invoked as supportive of NHST - it is not, because it is not the same thing. NHST and inferences based on statistical renderings of findings are two completely different areas of concern; they, merely look superficially similar. But they are not.

The system employed in the ARLT is similar in concept to that employed in the ADAFI as discussed in section 5.2.1. Structured hints are built into a research design with a designated pretraining phase, a learning phase which includes standardised error-specific hints and lastly the maintenance and transfer stage. The third phase assess for maintenance via similar items assessed for during the learning phase; near transfer is assessed for via items administered at the first complexity level followed by items at the second level of complexity thus assessing for far transfer. Depending on the facility with which these items in the learning phase are completed, the hint structure follows different paths. Hints serve to guide the learner and summarise responses and as analogical reasoning tasks lend themselves to multiple considerations within any one item (i.e. in a typical format, one needs to hold in memory one relation whilst applying one or more other relations) scoring becomes more systematised in approach. For instance, applying one relative rule within the same analogy results in a score less than the total possible but does not result in no score at all. Learners' potential is determined by the scores received on the third phase. This has allowed for initial differences in task approach and application to be evidenced as well as indicating rate of learning taking place. As the third phase is separated by a week from the first two phases, transfer and maintenance can more readily assess for learning that has taken place in an ingrained manner. The qualitative analysis of results occurs during the first two phases and the point of administering the two phases is to induce change via learning techniques and tools that are transferred to the individual. The Vygotskian notion of assisted developed is clearly evident here. Of note is the avoidance of utilising scores obtained during the learning phase as it is acknowledged that these scores are not good indicators of future outcomes. In conclusion, the scoring procedure can be viewed as follows leads one to ask uncomfortable questions which nevertheless need to be addressed:



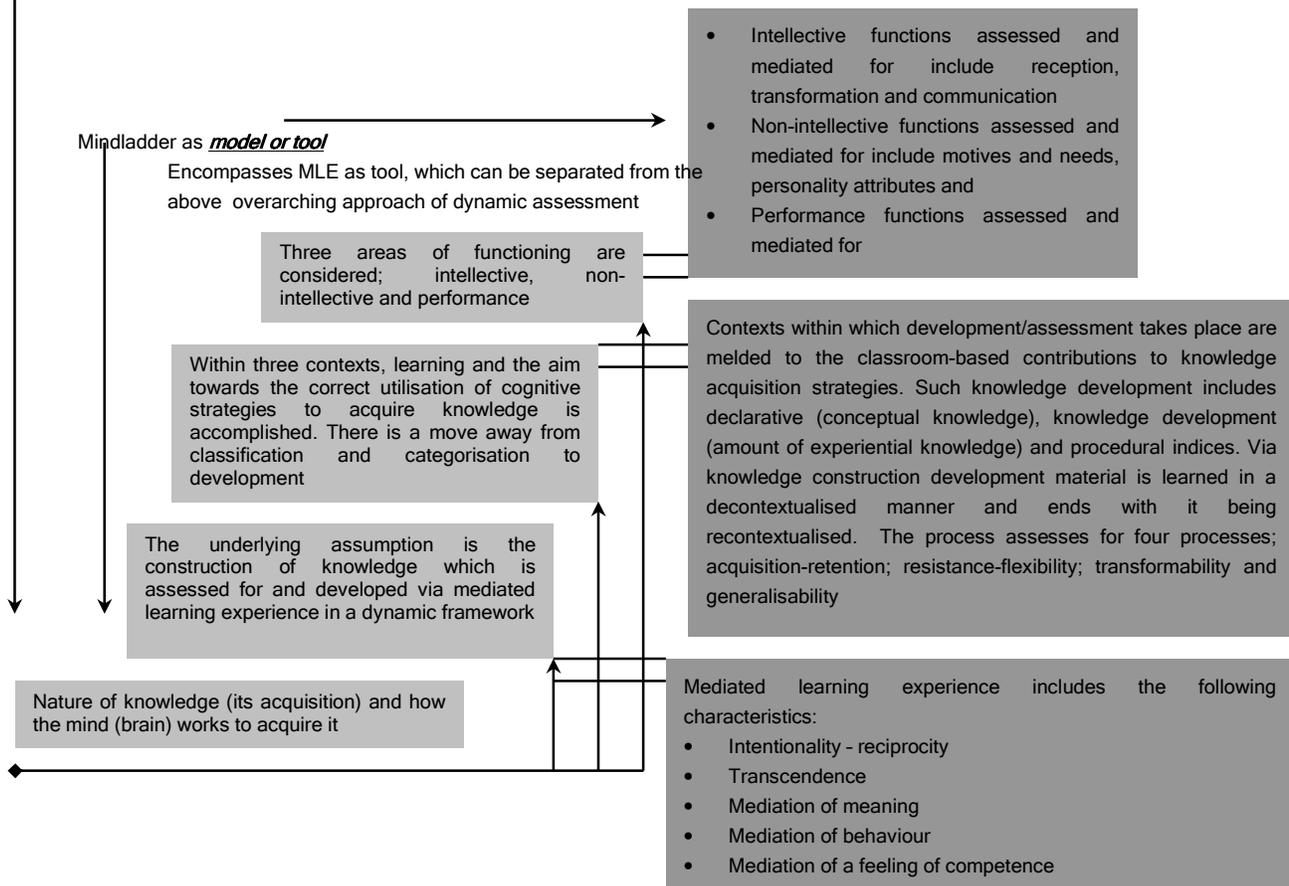
5.2.12 The Mindladder model: using dynamic assessment to help students learn to assemble and use knowledge

In comparison to most of the models positioned within the Madsenian framework, the Mindladder model (Jensen, 2000) can be said to avail itself of more attributes in the framework than most others due to the set-up and layout of its functioning and the clear formulation of the variables both hypothetical and empirical which fit comfortably within the HQ formulation. The model is explicitly conceptualised as a model; in other words, as in the case with some approaches/models/theories where the terrain of terminology is not always sharply delineated, this model presents a cogent argument for the use of certain hypothetical terms, empirical markers, theory-grounded concerns and cognitive terms within a dynamic assessment frame of its own. As discussed in chapter 3, the role of metaphor within psychology has often aided in elucidating often complicated processes involved in the description of psychological behaviour and functioning. Due to social science's slippage into terrain marked by inconsistency, ill-defined conceptualisation and lack of evidential yield, analogising often secures a firmer footing for tentative models deployed for the very function of making the aforementioned process streamlined, understandable and workable. There are a number of salient features of the mindladder model which make it particularly appealing as useable within intelligence assessment. It utilises computer assisted technology (briefly attended to in chapter 2) with which it shares a number of procedural similarities. It can be customised to any of three domains of functioning, namely, the home environment, the school-classroom environment and the larger school administration environment. The model is premised on constructivist notions of knowledge understanding which inherently relates back to the environment and cultural contexts in which learning takes place. This has somewhat of a relativist tinge to it as basis from which to build a model and from what has transpired in chapter 2 in terms of a relativist



argument of knowledge gathering, a few issues surface regarding the theory behind the model. Vygotskian, Feuersteinian and Piagetian views of knowledge gathering and sense-making all espouse the context validity of information processing and how such information is taken in (assimilated) and applied (accommodated) and as discussed, the emphasis placed on context was not as pronounced within the Piagetian framework as was the case within the former two. Knowledge acquisition is dependent on cognitive developmental stages and as such contextualised items form part of assessment and development. In this manner curricular questions are emphasised as it makes sense to redeploy the learner into the context in which he/she originally functions. Hence, curriculum concerns and assessment are parallel concerns. Recall the co-construction of knowledge via mediation from peers or teachers within Vygostkian thinking; the role played by proximal and distal features within cultural settings espoused by Feuerstein and the origins of cognitive stage-like processing of information and cognition from Piaget. These theoretical concerns, the emphasis of constructivist learning, the role played by dynamic assessment of such knowledge acquisition and construction processes; the role of assessment as just another step in the process of the development and assistance in utilising knowledge structures more effectively make the model an all-encompassing one.

The model is applicable to both special and normal learner populations across a wide age spectrum and can be administered either via electronic means or traditional pencil-and-paper means. Learning, the assistance needed to more effectively implement learning strategies and assessment present a continuum along which progress is monitored. The facility for far transfer is also included which is often lacking in most such assessments, however this needs to be initiated by parent or teacher. The model is implementable in both traditional and non-traditional environments. The analogy of the mind as a traversable ladder indicates that movement can proceed both up and down on what can be considered a cognitive structure and cognitive strategy implementation. The incorrect utilisation of certain strategies warrants a downward movement in order to accommodate the correct manner of cognitively dealing with certain strategies. Traditional parametric scoring is used throughout in order to keep track of progress and performance and in this way acquiesces to conventional demands placed on verifiable and objective tests. One wonders how often such traditional concerns are implemented to satisfy the status quo and in fact how fitting they are to the models conceived of originally. The differences highlighted by Jensen (2000) between his model and that offered in general from the static model will not be given here as it has already been highlighted in chapter 2. Constructivist learning is mediated with a view towards knowledge construction. The nature and subsequent theory of knowledge acquisition via overarching context is assessed for in a participatory “partnership” in a dynamically assessed for scenario. “Assessment intends to play a key role as integral part of deliberate efforts to help students learn how to assemble and use knowledge” (Jensen, 2000, p.188). The dynamic assessment feature plays the role of approach more so than model. The *model* is exemplified through Mindladder; the *approach* is exemplified through dynamic assessment and herein lies the crux in terms of the arguments debated in section 5.2.5 (dynamic assessment as offering constructs - new or derived concepts or dynamic assessment as offering method - approach). Constructs are *a priori* considerations which are operationalised and empirically manifested through the model which is constructed specifically to aid in the development of skill. Dynamic assessment functions in the latter capacity as offering method through which this model is realised. A snapshot of the model can be offered as follows:

Dynamic assessment as *approach or technique*

5.2.12.1 A(i) Ontology

The model's ontological status is more explicitly stated in comparison to other models surveyed in this chapter. Knowledge, it is stated, is constructed and is dependent on the pervading cultural context (distal) and local environmental circumstances (proximal). The Feuersteinian conceptualisation of mediated environment impinging on the transmission of knowledge acquisition is particularly evident in the model's emphasis on three independent contexts in which the model and programme can function. Development of strategies in a co-participatory set-up highlights progress and prescribes for the mediatory attempts necessitated by results offer a framework which assumes that mind (Jensen's use)¹³ can undo and redo learned effective strategies in terms of processing information and hence aiding in the acquisition of knowledge. People are not to be classified or categorised but aided via diagnostic techniques which nevertheless offer objective and widely applicable measures. Information is not seen as being transmitted but as being constructed and this can be aided via tools or functions afforded the learner in a dynamic manner. Mind is imbedded in culture and wider context and forms strategies based on mediation and learning but can perform optimally given the right conditions and guidance. Because knowledge is constructed it is assumed that it can be deconstructed which is where the notion of climbing up and down a ladder enters the metaphor for intellectual functioning. Rarely is it stated that cognitive processing tasks need to be unlearned so that new learning can take place for it is often likewise assumed that no learning has taken place in the first place. This is not necessarily the situation in all cases. Recall the applicability of this model to normal and special learners. Normal learners will indeed have learned strategies not necessarily befitting the task which will need to be relearned. The ontological concern can be summed up in this citation by Jensen (2000, p.188) "the Mindladder assessment model is designed to be able to function as an *integral part of the actual development* of students' learning ability, knowledge acquisition and problem solving skills" (own emphasis). The model assumes change is inherent, that human functioning is modifiable and that scores do in fact represent meaning beyond its numericised assignation. This last point is, however, contested. No definitive argument is offered in support of this contention. Knowledge cumulates in a context and the model cannot function without the "other" present in this context.

¹³ As is by now obvious, the author herself prefers the term "brain" in this regard but will not interfere with the original work.

5.2.12.2 A(ii) Philosophy

The model is espoused as having wide applicability yet being readily focused on the individual within context. Both nomothetic and idiographic concerns play forth as each developmental programme is conducted on a one-on-one basis. The data language is strewn with H-S and H-R terminology which immediately evidences its empirical and testable status. It is anticipated that this model will present as highly testable hence conforming to scientific dictates of nomological verifiability and falsifiability. The scoring of over 75 knowledge construction functions in the model is buoyed by the impact of mediated learning experience strategies for more effective use of cognitive strategies in the solving of problems. The assessor mediates via feedback the learning experience by “scheduling and framing the appearance of stimuli, by selecting them according to purposes and goals, by grouping them according to attribute and by imbuing them with meaning ...” (Jensen, 2000, p.196). Qualitative ecological assessment also forms part of the total assessment and relies on parental and teacher investment in terms of information provided regarding home and school environment.

5.2.12.3 B(i) Hypothetical terms

In an attempt to provide coherence to interlaced variables, hypothetical terms offer threads which weave the web of coalescing hypotheses and empirical data. The Mindladder model provides a dispersion of hypothetical terms across mentalistic, organismic and constructive terminology. Higher order thinking skills, learning ability, metacognitive and self-regulatory skills and knowledge refer to mentalistic functioning whilst aspects characteristic of the learner’s experiential reality are embedded within the mediated learning experience infrastructure. The main constructive term used in the model is of course the analogy of mind as ladder emphasising process as opposed to structure. Process and function are deeply rooted in most models but are most often not explicitly so stated. Dynamic variables make their presence in the more qualitative aspect of the programme. This is evident in the non-intellective functions in the theory of mediated constructivism such as motives and needs as well as personality attributes. These attributes are listed and assigned codes within the model. Among such attributes are the needs for feelings of competence, the need for achievement, the need for mastery, the need for novelty, the need for aspirations and so on. Within the ambit of personality attributes the following is considered: desire for approval, curiosity, frustration tolerance, optimism, self-efficacy and lastly performance attributes which include rapidity and precision, attention and persistence, habit formation, perception of need for effort, awareness of own improvement, enthusiasm and insight. The intellective functions provide more insight as to directive variables and include reception, transformation and communication variables. All these variables will be housed under the H-R variable delineation.

5.2.12.4 B(ii) Scientific hypotheses

Existential hypotheses become manifest through their functional counterparts. Causal patterns are what is hoped for in terms of the workability of underlying theses, which in this case as with other models include the hoped for elicitation of potential via concerted development progression.

Theoretical (hypothetical) scientific hypotheses

- H (learning ability) - H (plasticity)
- H (knowledge acquisition) - H (potential)
- H (problem solving) - H (structural changes evidenced via baseline, acquisition and retention)

Empirical (substantive) scientific hypotheses

- H (knowledge development) - S (mental representation)
- H (knowledge development) - S (memory retrieval)
- H (knowledge development) - S (hypothetical and inferential thinking)
- H (knowledge development) - S (goal seeking and setting)
- H (knowledge development) - S (planning)
- H (knowledge development) - S (searching for cause and effect))
- H (knowledge construction) - S (intentionality-reciprocity)
- H (knowledge construction) - S (transcendence)
- H (knowledge construction) - S (mediation of meaning)
- H (knowledge construction) - S (mediated regulation of behaviour)
- H (knowledge construction) - S (mediation of feeling of competence)
- H (knowledge construction functions) - R (of which there are 75, only a few will be delineated below)
 - H (knowledge construction functions) - R (intellective functions)
 - H (reception) - R (closure)
 - H (reception) - R (attention)
 - H (reception) - R (temporal orientation)



- H (reception) - R (traces, signs and symbols)
- H (reception) - R (explanatory behaviour)
- H (transformation) - R (stream of consciousness)
- H (transformation) - R (pre-analysis)
- H (transformation) - R (mental representation)
- H (transformation) - R (planning)
- H (transformation) - R (cause-effect relationships)
- H (communication) - R (trial and error)
- H (communication) - R (projection of virtual relationships)
- H (communication) - R (verbal tools)
- H (communication) - R (attention to outcomes)
- H (communication) - R (self-regulation)
- H (knowledge construction functions) - R (non-intellective functions)
 - H (motives and needs) - R (need of feelings of competence)
 - H (motives and needs) - R (need for achievement)
 - H (motives and needs) - R (need for mastery)
 - H (motives and needs) - R (need for novelty)
 - H (motives and needs) - R (need for aspirations)
 - H (personality attributes) - R (desire for approval)
 - H (personality attributes) - R (curiosity)
 - H (personality attributes) - R (frustration tolerance)
 - H (personality attributes) - R (optimism)
 - H (personality attributes) - R (self-efficacy)
- H (knowledge construction functions) - R (performance functions)
 - H (performance functions) - R (rapidity and precision)
 - H (performance functions) - R (attention and persistence)
 - H (performance functions) - R (habit formation)
 - H (performance functions) - R (perception of need for effort)
 - H (performance functions) - R (awareness of own improvements)
 - H (performance functions) - R (enthusiasm)
 - H (performance functions) - R (insight)
- H (knowledge development) - R (acquisition-retention)
- H (knowledge development) - R (resistance-flexibility)
- H (knowledge development) - R (transformability)
- H (knowledge development) - R (generalisability)

$$\Sigma (H-S) = 11; \Sigma (H-R) = 75; \Sigma (H-H) = 3$$

Hence HQ for the Mindladder model = $3/(75+11) = 0.03$ which is incredibly low in comparison to the other models surveyed and hence highly testable. What makes this all the more interesting, is that Jensen has conceptualised this as a model and has not proffered it as a theory even though theory underpins the model. This is a very testable model and it seems almost warranted to refer to it as a theory. It is due to the fact that each construct delineated has been operationalised that allows for its testability even though it is based on a very qualitative and untestable theory of mediated learning experience - traditionally the cornerstone of qualitative dynamic assessment interventions. This makes for an interesting analysis within the present Madsenian framework. Clearly, the feel of the model does not come across as one of extension in both areas of accommodating for static and dynamic properties.

5.2.12.5 B(iii) Hypothesis system

The Mindladder model is an instance of a 'model-within-model' approach, as a profile of learners' performance is compiled during the process which is displayed on a monitor. A depiction of what the model might look like is given above in the introductory section. Knowledge is assumed constructed and aided by prevailing contextual circumstances which implies its subsequent deconstruction in order for the re-learning of knowledge acquisition strategies. The model is housed within a very detailed deductive explanatory system primarily due to its detailed operationalisation. This does however raise the issue surrounding the veracity of deductive arguments based purely on operationalised constructs. Merely "empiricising" variables does not mean that a deductive system is at work; it merely reflects its nomological status as verifiable or falsifiable model. The structure is so finely worked out that it would prove a task to debate against its deductive claim but a claim can nevertheless be lodged.

5.2.12.6 C(i) Abstract data

There are myriad H-S and H-R variables identified within this model although they are not stated as such. Functional relations highlight the relations within each of the intellectual, non-intellectual and performance functions but as is standard knowledge, there are relations between these levels as well. This is seen with improved performance on items requiring cognitive operations due to improved (via mediatory facilitation) non-cognitive or non-intellectual behaviour. These can be tentatively stated as correlations between variables.

5.2.12.7 C(ii) Concrete data

The concrete data stratum is evidenced through the performance of these abovementioned three functions. The mediated learning experience also adds to the facilitation of the improved knowledge acquisition skills and includes mediatory tools such as modelling, focusing, anticipation, scaffolding, fading, error recognition and reflection in gaining greater self control and mastery.

5.2.12.8 C(iii) Prime considerations

Once again in capitulation to mainstream assessment requirements, the Mindladder model avails of quantitative constructs which is faithfully compared to conventional methodology. In this regard it is worth citing the following from Jensen (2000):

“Scores acquire meaning from parametric analyses of quantitative information comparing post-mediation performance to pre-mediation performance and from qualitative information obtained in quasi-controlled experiments” (p.189) whilst concurrently maintaining that “we can measure either the properties of stability or the properties of modifiability but never both at the same time” (p.203). In as unsophisticated a comment as that, it becomes very clear that the two approaches, static and dynamic, are at odds. However, Jensen has failed to mention change-based IRT models which could accommodate for this disparate schism.

Parametric analyses of data within the programme allow for significant change results to be reflected. Data are treated as conventionally treated within static assessment. A counter-argument could be that once change has been effected, the need to “prove” this becomes almost impossible to resist. The use of conventional statistics within the change process itself and its use after change has been effected illustrates two separate issues. The use of statistical techniques during construct meaning-making is perhaps questionable (in fact, very questionable) but is perhaps not as deviant after the construct has been identified. Although this too is open to debate.

5.2.13 The cognitive modifiability battery (CMB)

The cognitive modifiability battery (CMB) is one of a number of dynamic assessment batteries for young children developed by Tzuriel and his colleagues. The niche carved out by the availability of a pool of young children affords dynamic assessment methodology the opportunity to engage with learners from as early a developmental age as possible, particularly those at greater risk of developmental delays caused by any number of factors. Childhood assessment is framed from the outset by a number a factors that focus mediation and assessment in a particular manner such as offered below (Tzuriel, 2001):

- The test materials are more concrete in nature
- They are attractive to children and appear game-like
- Children’s typically shorter attention spans are considered
- The assessor’s tone during assessment plays a particularly large role as it not only facilitates the assessment process but forms part of the mediational strategy itself and is referred to as “rhythmic intonation”

The test battery’s purpose is two-fold: a tool for diagnosis and intervention and it need not necessarily be utilised for both purposes in one setting. Moreover it can be functionally utilised in either a measurement/research or clinical/educational mode within the diagnostic process. In this manner, the CMB appears to cover as wide an area as possible within the dynamic assessment array in similar fashion to other models and batteries which aim to accommodate for as many factors as possible. The scientific imperative of remaining streamlined (parsimonious attribute) is sometimes eschewed in favour of greater diversity of methodology. As mentioned in this chapter, this can play against the need for robust science practice but given the nature and context in which dynamic assessment finds itself, movements away from a strict nomological model is not necessarily to be construed as negative.

The appeal and allure of dynamic assessment is particularly strong when dealing with children as the techniques are intuitively adaptive to this population group evidencing typically less developed verbal acuity and cognitive abilities in general. In a manner, assessment is more complex as extrapolations and inferences are called upon for measurements of constructs which

are not necessarily as concretised as evidenced in older learners. The premise behind the design of the CMB is to identify, assign and intervene in order to aid for future academic studies; such studies being typically identifiable with static-based assessments. Mediatory intervention takes on a number of characteristics depending on the nature of the individual and assessment item. Highlighted aspects within the CMB include the nature of the constructs assessed for and the degree to which such constructs are related along a number of dimensions. The CMB does not engage in novel construct meaning-making but rather, uses existing constructs in a manner befitting the elicitation of modification. Leaning heavily on Vygotskian and Feuersteinian notions of dynamic assessment and indicating the differences between the approaches, the CMB sub-tests are premised upon related cognitive constructs. These relations are attested for on the basis of factor analytic results. The nature of scoring also influences statistical results such as correlations between residual posttest scores. Partial-credit and “all-or-none” scoring provide different diagnostic value and is a scoring technique that has been in use for many years and was in fact one of the earliest methods of scoring in South African dynamic assessment research. The defining feature of the CMB is its reliance on standardised statistical techniques in identifying constructs and relations between such constructs and doing so in tandem within a development framework. In other words, natural age-graded development provides the reference for changes induced or produced before and after teaching or mediation.

5.2.13.1 A(i) Ontology

The most difficult sub-strata to elucidate upon in this chapter is the ontological and philosophical, not because they are by their nature difficult to tease apart but because most dynamic assessment batteries and models do not explicitly state their underlying ontology and governing philosophy. The CMB is one particular such instance where it becomes difficult to infer because not much is given from which to extrapolate. The basic premise behind the CMB is similar to basic ontologies underlying much of dynamic assessment: that elicitation of potential is not only possible but measurable in both qualitative and quantitative manners befitting standardised modes of currently accepted psychometric practices. In this particular instance, the identification of specific deficient functions comes to the fore in both the clinical and intervention set-ups. Regarding Madsen’s conception of man and the related psycho-physical theory it can be said that children are very amenable to change and more so within the cognitive realm. This latter implicit assumption however is open to debate as it is knowledge or “fact” that is necessary for the school system to function. Most school related tasks are cognitive by nature and by identifying deficits and encouraging modification through engagement such cognitive skills can show improvement. This is of necessity a belief which needs to be believed in if events are to turn out for the better. So in essence, there is an *a priori* assumption that manifest change can result within the cognitive domain and the need to induce for such change and subsequently measure it becomes the next logical step to take. This is as much as can be said regarding the ontological base of the CMB’s construction. As Madsen has stated, each theory or model has some layered ontology whether it is explicitly or implicitly stated which can be at the very least inferred. There are instances though where inferences can only go so far and no further.

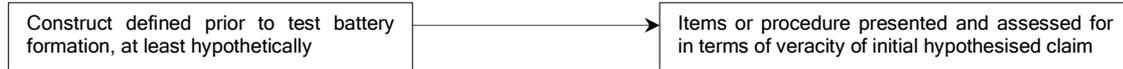
5.2.13.2 A(ii) Philosophy

The manner in which the CMB isolates factors from exploratory analyses evidences the surety with which results from such techniques are taken to be true. The CMB is of course only one of innumerable instances where reality (potential) is assumed to exist and via methods from various instruments can be plausibly extracted, at least with a veneer of scientific credibility attached to it. Epistemologically, it can be quite tricky to ascertain what philosophy the CMB adheres to but one can make a case for its realist and instrumentalist philosophies. Madsen’s assortment of such epistemologies dictate that models and theories capture either implicitly or explicitly these underlying ideas. Potential and the identification of deficits is understood to be a reality regardless of the hitherto inability of science to get at the core of this issue. By assuming that such deficits can be remedied is an assumption about the reality of their existence as well as an assumption about the facility of getting access to them. However it can be argued too that an instrumentalist account of this reality overshadows the former by emphasising that only a certain section or view of the truth can be “got at”. In other words, the only truth available is pragmatic truth, truth that is useful in understanding the hypothetical construct. What lies beyond such pragmatic truth cannot be known. Here we are constrained by the techniques used to extrapolate from such pragmatic truth (exploratory factor analysis for instance). Can it be said that the CMB’s truth value is only as good as the method used to extract the hypothetical construct and make it manifest? In this instance, one has to work backwards from the technique used to locate the construct and understand how this technique is positioned philosophically in the first place. The CMB discussion does not avail of finer philosophical renderings of what can be considered epistemological ponderings.

5.2.13.3 B(i) Hypothetical terms

The CMB sub-tests in addition to the methodology utilised to come by the information required is very grounded in approach and the description of the entire test seems to solidify this view. Very few hypothetical terms pervade the CMB discussion and those that emanate from factor analyses are offered as statistical hypothetical instances of underlying loadings. As there is no definitive explanation surrounding the substantive construct of “learning potential” (chapter 2’s exposition into the realm of neurological isomorphism touched on this issue) the only tangible aspect to which one can grab onto in terms of making manifest such a construct is the test battery or sub-component used. The CMB exposition by its silence on the topic of construct-

making highlights this lack of hypothetical terminology but more specifically the lack of a networked system of supportive hypothetical terminology. The usual case within dynamic assessment usually presents as follows:



but in the instance of the CMB the case is presented as follows:



This is not a critique as such but an observation, although it allows for better placement of the CMB when compared to other dynamic assessment batteries of models on a number of criteria. Due to the nature of the above discussed matter, the CMB presents as rather sparse when hypothetical terms and hypothesis system is considered.

5.2.13.4 B(ii) Scientific hypotheses

The CMB would seem to lend itself to straightforward hypothetical quotient analysis especially in the format of a diagnostic tool. The rationale behind its existence

“has been motivated by the need to decide about [young children’s] educational assignment... early identification of specific deficient functions ... that limit the child’s cognitive functioning is of crucial importance for facilitating development of learning strategies and to enhance the child’s preparation for school learning and academic success” (Tzuriel, 2000c, p.375).

Static conceptions infuse the rationale behind the need to assess and mediate for skill development and enhancement in a dynamic manner. The inescapable conventions force the CMB to operate within traditional measurement systems where identification, classification and assignment are important functions of assessment so as to secure for the individual a place within the conventional educational set-up. There is nothing intrinsically wrong about this motivation because the justification behind this is that reality as is, dictates that individuals succeed as best they can within such reality; the reality being school learning and academic success. This ultimately leads to tertiary success and eventual optimal functioning within a world (not just certain societies) which empahsies these particular attributes. One cannot run away from these blunt facts. The need to assess according to these dictates is not going to go away. However, the path leading to such assessment and eventual outcome can indeed be tailored to certain dynamic approaches. One regime only will never prove viable as sole alternative for all time. History has shown this. The continual change and flux brought about by modification and change within conventional systems is what makes history. The need to assess for both within static and dynamic modes is emphasised. Neither has a future in the offing without requisite acknowledgement of what the other has to offer. Once again, the reader’s attention is called to the plight of assessment in the future and the place carved out for dynamic assessment. The two can co-exist but issues need to be solved or the two must continue within different realms of meaning-making (or paradigms).

The CMB assesses “cognitive processes and mental skills” (Tzuriel, 2000c, p.375) which is somewhat vague as a general statement and does not really say anything much of value as most psychological tests assess for cognitive processes and mental skills, hence the rather vague allocation of H-H terms below. But what is striking is the CMB’s application within a developmental framework of child development. Recall the discussion in chapter 4 where mathematical and statistical IRT change-based models accommodate both numerical and cognitive development aspects; two aspects of change which are at once closely aligned but epistemologically very divergent in nature. “Meaningful cognitive changes” are said to have been brought about after use of the CMB as an intervention tool but the support underlying this contention is supplied by statistical significance which is known to be valid as a statistical outcome only (see discussion on NHST within the social sciences in chapter 4). Statistical significance and cognitive developmental change as significant cannot be easily equated as the two types of change emanate from two different ontological and epistemological realms. The reliance on statistical renderings of meaningful change within the CMB is problematic for this reason and is perhaps the most obvious aspect to critique but this must be understood within the context described above in the grey text box. There is an understanding of the need to engage in



conventional modes of assessment as argued but to rely exclusively on such premises can also be problematic. The solution? As advised on numerous occasions throughout; separate or unite but somewhere changes will need to be effected. For purposes of HQ definition the following can be tentatively be offered following Madsen's dictates for identifiable hypotheses systems:

Theoretical (hypothetical) scientific hypotheses

- H (cognitive area; "mental skills") - H (factor loading 1)
- H (cognitive area; "mental skills") - H (factor loading 2)
- H (cognitive area; "mental skills") - H (factor loading 3)

Empirical (substantive) scientific hypotheses

- H (cognitive area; "mental skills") - S (seriation)
- H (cognitive area; "mental skills") - S (reproduction of patterns)
- H (cognitive area; "mental skills") - S (analogies)
- H (cognitive area; "mental skills") - S (sequences level 1)
- H (cognitive area; "mental skills") - S (sequences level 2)
- H (cognitive area; "mental skills") - S (memory)
- H (Learning potential) -R(type and amount of mediation)
- H (Learning potential) -R(level of task difficulty)
- H (Learning potential) -R(deficient cognitive functions manifested)
- H (Learning potential) -R(analysis of non-intellective behaviour - need for mastery)
- H (Learning potential) -R(analysis of non-intellective behaviour - tolerance for frustration)
- H (Learning potential) -R(analysis of non-intellective behaviour - anxiety)
- H (Learning potential) -R(analysis of non-intellective behaviour - resistance to mediation)
- H (Learning potential) -R(analysis of non-intellective behaviour - self-confidence in response)
- H (Learning potential) -R(analysis of non-intellective behaviour - locus of control)
- H (Learning potential) -R(analysis of non-intellective behaviour - level of independence)
- H (Learning potential) -R(analysis of non-intellective behaviour - degree of vividness and vitality)

$$\Sigma (H-S) = 6; \Sigma (H-R) = 11; \Sigma (H-H) = 3$$

Hence HQ for the CMB = $3/(6+11) = 0.17$. The CMB as it presents as a diagnostic tool is highly testable given the number of H-R and H-S variables it seeks to measure although the nature and number of hypothetical constructs is kept to a minimum with construct clarification relying on exploratory factor analysis as well as smallest space analysis. The outcome of this exercise could well be different when the CMB is considered in its role as intervention tool where it would most likely evidence a higher HQ due to less reliance on numerical score in its qualitative effort to mediate and enhance skill in a programme of longer duration.

5.2.13.5 B(iii) Hypothesis system

Measurement rationale behind the CMB considers both the individual clinical overview and measurement of intervention efficacy of the testee and is more in keeping with an idiographic depiction of behaviour. The CMB presents as a technique more so than as a model. By utilising sub-tests in a manner seeking to shed light on behaviour and as a means towards improving functioning and by relying on statistical results evidencing factors assumed to underlie these sub-tests, the course of assessment and intervention more typically resembles dynamic assessment than do the actual sub-test constructs. The governing hypothesis in the CMB is one which is familiar to most if not all dynamic assessment efforts: to identify as early as possible the specific deficit functions in the child's cognitive functioning. These deficits are then remediated. There does not appear to be an integrated network of other hypotheses in this assessment battery.

5.2.13.6 C(i) Abstract data

Both functional and correlational data relations are indicated and utilised by the CMB. Madsen's data stratum contains experimental trials and observations and is one of the most obvious examples of the models discussed in this chapter where experimental data describe and statistically validate construct delineation from empirical techniques such as factor analysis and other correlational studies. Although the data do not themselves contain hypothetical terms, they evidence hypothetical relations between them which is precisely what characterises the abstract data stratum. Correlational results are not necessarily causal in nature and correlational matrices forms the foundation for factor analysis as well as weighted smallest space analysis so it can be tentatively stated that factor loadings account for variance between various scores and that a significant amount of variance is accounted for by a common factor. Of course the retort is that variance accounted for has nothing to say of causal relations

between sub-component loadings other than the fact that variance is accounted for in explaining the relation between the various sub-components. Nevertheless, the rationale behind the CMB's constructs is built up from abstract data these interfacing between the concrete data results and higher order construct determination.

5.2.13.7 C(ii) Concrete data

The CMB revolves around its concrete data findings and due to its reliance on exploratory statistical techniques is very dependent on this level or stratum. Once constructs are identified on the basis of factor loadings (numerical) and clustering along multiple dimensions (visual) investigations into theoretical realms becomes possible. However, as with science in general, bottom-up processing is the route followed by the nomological-deductive system. Theory is informed from data in the practical world with the assumption being that such data is useful and scientifically accountable. The CMB offers qualitative data as well but the HQ system above was based on the quantitative data gathering technique. Nevertheless, qualitative data garnered from non-cognitive factors "codetermine modifiability" within intervention (Tzuriel, 2000c, p.404) and is equally valuable as concrete data. Transfer scores are indicators of possible change even though change as assessed for by pretest-posttest scoring remains problematic, however residuals of postteaching scores are used to assess for such change. Concrete data in any of its varied forms will always function as foundation for any assessment or intervention. The degree to which one relies on such data is determined by the tool used as well as the rationale behind the need for such "scores" in the first place. Utilising data to draw statistical conclusions warrants a research design specifically tailored to seeking such numerical data whereas other designs emphasise process over numerical data but nevertheless presents concrete data of its own.

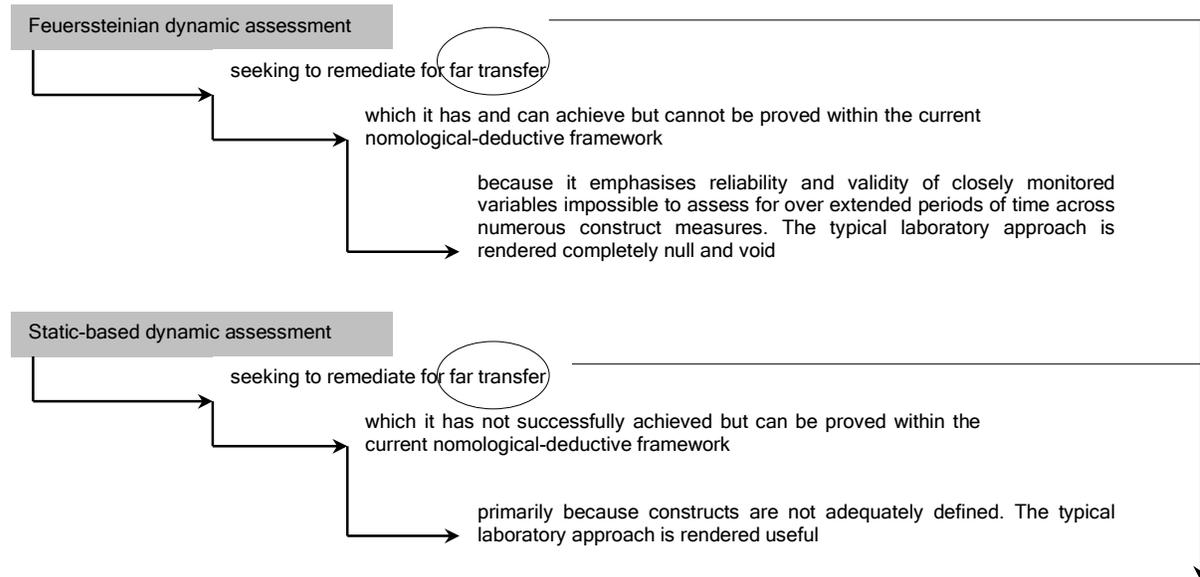
5.2.13.8 C(iii) Prime considerations

The models and test batteries investigated in this chapter are not typically assessed in terms of their reliability and validity criteria as the interest in this thesis is the ontological and epistemological background to the models. However, in some instances the model and battery are so closely tied in to their statistical background that it becomes difficult to always separate the eventual test outcome from its development. More quantitatively driven models capitalise on the offerings of statistical techniques in order to define constructs whereas wholly qualitative models find recourse to face validity criteria when detailing the nature of their constructs. However, Tzuriel (2000c) stresses that "crucial factors in the assessment and intervention processes are non-intellective factors" (p.378). Nevertheless, numerical scores are assigned to correct answers on dimensions within sub-tests as well as scores to correct entire solutions regardless of the correctness of dimension results. This type of scoring scenario is typical of what is found when mathematical items are scored; the incorrect subsequent steps might still be scored even though the path chosen is incorrect. In this way, the learner is not penalised to the fullest extent but receives some credit for having correctly proceeded along an incorrect path towards a solution. Once again, the gain score is calculated in a simplistic manner by subtracting the pretest score from the posttest score. It is an empirical study aimed at deducing the nature of the constructs underlying the CMB for which exploratory factor analysis employing orthogonal rotation is utilised. Exploratory factor analysis is used when the nature of the dimensions assessed for are not altogether known and the degree of overlap between the sub-test components is also largely unknown. Tzuriel (2000c) also employs weighted smallest space analysis which in essence confirms the factor structure found in the exploratory factor analysis and illustrates the factor clusters visually along three dimensions.

The question that arises when the sub-components of the CMB battery are analysed in this manner is the rationale behind the choice of the components if they are not completely understood prior to implementation. The usual case with established test batteries is to employ confirmatory factor analysis. *A priori* factor loadings are usually established once assessment commences but new batteries would first need to establish such factor loadings. Principle components factor analysis minimises the number of dimensions without concomitant loss of information and rotating the factors orthogonally assumes that factors are not correlated with one another hence their rotation at right angles which in turn minimises the number of variables loading highly on any one factor. In this manner, *a priori* decisions are being made about the nature of the data before such conclusions can be made. The sub-components in the CMB yield three distinct factors with noticeable eigenvalues attesting to this distinction. The weighted smallest space analysis (WSSA) visually illustrates virtually the same scenario and clusters factors along three dimensions, two of which relate to the degree of abstraction of construct and task-specific strategies of the measured construct. One wonders about the nature of far and near transfer tasks if learners are assessed and remediated for these seemingly task-specific tasks regardless of the degree of abstraction involved.

Far transfer within dynamic assessment often presents as a mirage; a goal seemingly close in view but unobtainable in most instances. What is far transfer? Is it the extent that learned skill is implemented on similar tasks eliciting similar construct manifestation? Does dynamic assessment not strive for far transfer of understanding of concept? Concrete skill is one thing but understanding of underlying concepts underpinning varied skills is entirely another. Feuersteinian dynamic assessment strives for the latter hence its low reliability and validity indices. Batteries offering skill mediation such as the CMB favour Feuersteinian modes in general but acquiesce to current dominant forces within traditional psychometry. This is nothing new, but highlights the role played by skill and concept training for far transfer. Dynamic assessment interventions possess a poor track record of far transfer maintenance over ever-extended periods of time. Poor results buffer support for traditionalists arguing against efficacy

of long-term programmes humbly attempting to bring about change. If dynamic assessment is to procure for itself more favourable support regarding its methodology it would do well to focus on near and far transfer issues which currently plague its programme. Unfortunately this scenario presents what could be construed as a dichotomy (which, as already discussed throughout this study, is reflective more of how we ascribe to nature our own biased conceptualisations of how we think the world works as opposed to the actual rendering of reality as it presents in the real world).



Far transfer: the degree or extent to which what has been learned can be transported to novel situations assessing for similar constructs. *If manifest skill is mediated for then superficial far transfer may take place but will tend to fade over time as the construct moves away from the original construct mediated for. This can at least be measured and tested because it is very domain specific. Structural change does not necessarily occur*

Far transfer: the degree or extent to which what has been learned can be transported to novel situations assessing for similar constructs. *If underlying general competency is mediated for and if this mediation incorporates substrates closer to heritable intelligence then far transfer becomes more probable but concomitantly vague and hence not amenable to measurement because it is very domain general. Structural change occurs*

Cognitive modifiability results in transference and the ultimate goal is to achieve far transfer. CMB correlations evidence results to the contrary illustrating non-significant relations between modifiability scores among sub-components of the battery. In other words, cognitive modifiability is weakly related among the various domains.

As typifies conventional battery reliability and validity indices and the quest towards providing such statistics, the CMB has yielded statistically significant results attesting to its predictive facility. However this chapter does not focus on the statistical results from battery implementation but is mentioned here due to Tzuril's (2000c) obvious concern for the model's conventional veracity. The need to prove this is understandable but not necessary when it comes to the assessment of the model in terms of its placement philosophically.

5.3 Summary of HQ results

Dynamic assessment theories and models are assessed in this chapter in a very specific manner utilising a very specific format. The conclusions reached, by way of the hypothesis quotient results, is a focused effort at quantifying what at times can be said to be non-quantifiable. The debate ensuing within psychological assessment as to the nature of the construct being assessed for bears heavily on the nature of the quantification of such constructs. This thesis may well have adopted and adapted another framework according to which testability of models and theories could be assessed. The conclusions are thus couched in the following two concerns;

1. that theories and models should be quantified in an effort to bring some means of common metric to the field (even though the technique can be disputed, it nevertheless allows for a unanimous means of assessment of these theories and models)

2. the chosen Madsenian framework resuscitates a useful yet dated method of deploying just such an analysis. Another method or framework could have been utilised thus resulting in its own common metric of testability. The issue here is not the nature of the chosen method but that a common denominator allows for equal treatment of all models. A counter-argument may well be lodged against the use of Madsen's framework, namely, the emphasis on a testability criterion as premise. Chapter 3 has argued for and against the use of such a premise within psychology-as-science. This issue is far from resolution and it is not the intention of this thesis to find a final solution to this aspect, this remains to be fleshed out and solved (if a solution is even in sight). The emphasis in chapter 5 was to seek a novel means of comparing theories and models in a manner befitting psychological theories. Madsen's framework bequeathed to this study just such an opportunity. Another framework could very well have emphasised constructionist approaches towards the understanding of intelligence and potential but acknowledging the real world in which assessment takes place forces one to affirm the nature of quantification as paramount in current assessment methodology. Whether this is 'healthy' is a matter addressed in chapter 3.

Disagreement about the framework deployed is not of concern when studying the summary of HQ results as it has been stated that alternative measures could likewise have been utilised. The following testability summary is presented and follows the sequence of models discussed above. The table following this presents the models' range of testability from highly testable to least testable as evidenced from the HQ scores.

<i>Section</i>	<i>HQ score</i>		<i>Section</i>	<i>HQ score</i>	
	<i>HQ variation between 0 (highly testable) and 2 (not very testable)¹⁴</i>				
	<i>For the sake of expediency, a score of 1 is considered as partially testable</i>				
Section 5.2.1	LLT = 0.5 and ACIL = 0.1	Highly to partially testable	Section 5.2.8	LPAD = 0.06	Highly testable
Section 5.2.2	LIR = 0.08	Highly testable	Section 5.2.9	EPA = 0.13	Highly testable
Section 5.2.3	Dynomath = 0.4	Highly to partially testable	Section 5.2.10	ACFS = 0.07	Highly testable
Section 5.2.4	Carlson and Wiedl model = 0.375	Highly to partially testable	Section 5.2.11	ARLT = 0.85	Highly to partially testable
Section 5.2.5	Analogical reasoning = 1.88	Not very testable	Section 5.2.12	Mindladder = 0.03	Highly testable
Section 5.2.6	LEM = 0.22	Highly testable	Section 5.2.13	CMB = 0.17	Highly testable
Section 5.2.7	S-CPT = 0.16	Highly testable			

<i>Range of testability: from highly testable to least testable as evidenced from HQ scores</i>		
0.03	Mindladder	highly testable
0.06	LPAD	highly testable
0.07	ACFS	highly testable
0.08	LIR	highly testable
0.10	ACIL	highly testable
0.13	EPA	highly testable
0.16	S-CPT	highly testable
0.17	CMB	highly testable
0.22	LEM	highly testable
0.37	Carlson and Wiedl model	highly to partially testable
0.40	Dynomath	highly to partially testable
0.50	LLT	highly to partially testable
0.85	ARLT	highly to partially testable
1.88	Analogical reasoning	not very testable
<i>Average HQ</i>		
0.35 which results in the average model reflecting its status as one of highly to partially testable model		

In flagrant disregard for qualitative aspects within the theories and models and accounting only for measurable properties as evidenced from the framework it is evidenced that, of those theories and models assessed, Jensen's (2000) Mindladder model yields high testability and thus is most amenable to nomological-deductive discussions pertaining to its utility as theory and

¹⁴ An HQ score of 0 has been obtained for behaviourist theories of motivation which is obviously anticipated due to their stance on the non-existence of non-observables. Freud's theories on motivation score a 2 which is indicative of the difficulty in testing for this theory. There could well be higher HQ scores but Freud's score has been used as an instance of highest HQ score possible.

model within dynamic assessment in psychological measurement. However, eight models are close on this model's heels in this regard evidencing a difference of only 0.19 between the Mindladder model and Hessel's (2000) LEM. This result is, at most, superficially indicative of testability differences between the two models. The figure of 0.19 is meaningless as a numeric as Madsen does not provide a means of sensibly making meaning of such a difference; it is an exercise purely engaged in for the sake of curiosity. As one moves down the range of the testability table it becomes apparent that four models are less testable than the models falling in the former range. These models are nevertheless testable within a quantified framework. The least testable model within this framework is Karpov and Gindis's (2000) analogical reasoning model. The question expected from adversarialists would be the degree of unsuitability of their model within the current framework of mainstream assessment. Judging the model within a framework other than the one chosen may result in a conclusion different in nature to the one reached. Recall that less testable means that the model is less amenable to strict replicability within the nomological-deductive model of how science is currently practiced. A move away from this *modus operandi* can polarise the discipline in such a way that Karpov and Gindis's (2000) model becomes the model most sought-after within another less nomologically aligned methodology of science practice. The mean HQ result of 0.35 reveals an interesting trend within dynamic assessment (as exemplified by the Madsenian framework) and that is as a unique and historically richly informed manner of potential assessment emphasising process above product it nevertheless seeks to satisfy current mainstream methodological approaches towards robustness, predictability, reliability and validity of quantified numerical results. It seems as if this sub-discipline is talking at cross-purposes; advocating its strengths as undeniable in the modern world of assessment but revealing a close affinity to old school premises. This leads to the following questions being asked of this method of potential assessment:

- can dynamic assessment break the mould?
- does dynamic assessment break the mould?
- should dynamic assessment break the mould?

Each answer needs to be clearly identified as to its basic philosophical tenets. This chapter has not answered these three questions but has led to them being asked. This is the first step in a process that has still to be revealed.

5.4 Conclusion

Chapter 5 should not be construed as *the* culmination of preceding chapters as the entire thesis has endeavoured to understand dynamic assessment's predicament from varying perspectives. What the chapter does achieve, is a manner of bringing to a focal point the possible reasons for dynamic assessment's chosen path within psychological assessment. It does culminate in the deployment of the framework which itself was informed from discussions in previous chapters. The previous four chapters can stand alone as a closed summation of the status of the field of dynamic assessment as well as psychological assessment per se. In concretising the models and theories, the framework brings to life various issues which plague the field and gives meaning to these issues which were only abstractly referred to in the first four chapters. By identifying specificities within the chosen models, evidence of sorts is presented to support certain claims that were tabled in the previous chapters. The chapter opened with a retort to main contentions upheld in the study and sought to put to rest certain counter arguments which can legitimately be put forward in such a study.

Working from the test battery and inferring assumptions prevalent in the model/theory leading to the implicit and explicit meta-framework ensconced within the theory allowed the framework to be implemented on various practical test batteries. The small number of batteries thus assessed serves to inform future possible batteries but the main reason behind such analysis, apart from determining an HQ score, is to highlight a unique issue inherent within each model. Each model yielded a novel contentious aspect the roots of which can be traced back to grounding philosophical and methodological issues pervading the discipline of psychology as a whole. The startling finding, which at the outset of the study was an educated guess at best, was that many of the problems affecting dynamic assessment models were problems emanating from meta-levels within psychology as an entire enterprise. Many problematic issues have been inherited from above so to speak and are not the exclusive domain of dynamic assessment. Attempting to get psychology's house in order will result in many smaller rooms being cleared of remaining debris and confusion. Some issues, of course, are solely within the ambit of assessment and other issues are even more constrained within the specific domain of potential assessment.

The most often cited highlighted concern revolved around core constructs and how in fact such constructs can be reified (if at all), the nature of construct delineation and the manner in which change is accounted for in constructs which are not similarly altered. These issues were highlighted in sections 5.2.1, 5.2.2, 5.2.3, 5.2.5, 5.2.9 and 5.2.11. Another issue, that of altering the construct and seeking to measure it as changed construct is addressed in sections 5.2.2, 5.2.7 and 5.2.12. The nature of change as well as ways of seeking to account for such change (rate of change for instance) was highlighted in sections 5.2.2, 5.2.6 and 5.2.12. The nature of methods seeking to alter change such as various mediatory interventions (clinical vs. structured methods) and the resultant near/far transfer dilemma was highlighted in sections 5.2.4, 5.2.6, 5.2.7, 5.2.9, 5.2.12 and 5.2.13. *A priori* concerns surrounding change resulting in such *a priori* concerns as givens was highlighted in section 5.2.13 with the veracity of dynamically altering static-based concepts highlighted in section 5.2.7. An issue discussed at length in chapter 3 and highlighted



as a concern in section 5.2.8. is the argument for and against the use of statistical and clinical appraisals within dynamic assessment interventionist strategies. Cognitive assessment and the role that cognition plays within curriculum-based assessment was highlighted in section 5.2.10.

The determination of HQ results attested to various models' testability only within the nomological-deductive framework which, it must be emphasised, is not a model adhered to nor accepted as standard science model by all practitioners alike. There is no "winner" of the HQ contest. This is not the point of the exercise, for most of the results can quite happily hang in the air. It is the interpretation of each score *within this governing framework* that is important. Utilising another framework can radically alter any conclusions resulting from such a comparative study. It may be useful then to devise a new meta-theoretical framework with novel concerns and priorities and study the results emanating from such a newly devised framework. There is no right or wrong answer, there is no correct model nor is there finality in any sense of the word. The framework allows for *a* conclusion to be reached, one conclusion among many possible conclusions; it is a framework-specific result only. This does not mean, however, that this framework deployed and its varied conclusions are not useful, because they are, seeing as any supported framework will ineluctably lead to novel highlighted concerns and conclusions. Doubtless there are concerns which were not highlighted due to the specific nuances of the framework utilised as one framework cannot possibly be attuned to every possible outcome of analyses. This is an inherent limitation within the study and is acknowledged at the outset. A firm conclusion, if this is at all possible, can only state the average model as analysed within the Madsenian framework results in a highly testable to partially testable model in the conventional sense of assessment practices. In fifty years time, such a conclusion could well be meaningless, as by then, it is possible that assessment will have moved on to different ways of working. Dynamic assessment springs from process-based accounts of change where the individual is emphasised as paramount concern but its practical outpourings tell a different story. Methods and philosophies are not communicating as effectively as they should. The governing system within which dynamic assessment works is partially to blame for this state of affairs. Whence dynamic assessment? But perhaps more importantly, whence psychology?

Chapter 6 Conclusion

6.1 Introduction: the value of the study in terms of its coverage

This brief chapter serves in the capacity as summarising recombining; a necessary pull to the knot in order to secure the package in one coherent and complete whole.¹ The study opened with an introduction decrying with fervour the thesis of lapsed concern for the fundamental issues striking at the very heart of the matter of assessment within the broader scope in which psychology itself is lodged as competitor science. In support for the rationale for the study, an overview of the crises within the field of dynamic assessment was depicted and an in-depth look at why such a tentative and exploratory study is warranted was argued. Although this study is “contextless” in terms of its arguments being applicable in the broadest sense, the country of origin, South Africa, from which the author was writing cannot be said to have played no role at all and it was with this in mind that further spurred the need for this exploration to be undertaken. In essence, the method of dynamic assessment holds much potential of its own and if it is to be worthy of utilisation within educational contexts hoping to aid in the identification of potential of previously disadvantaged students, then a case for it can be made on the basis of a renewed look at its own core of method and philosophy. South Africa is sorely in need of some means of “attuned assessment”. Attuned to this country’s specificities in as broad a sense as possible.

In an initial and precursory attempt to aid dynamic assessment’s future, which at times does not bode well, the idea of a revamp was tackled from the point of view of meta-theory even though dynamic assessment consists of varied models and not always varied theories. In order to put in place an infrastructure from which to view and assess the plight of this manner of assessment the building blocks had to be layered first. This took the study to the next chapter which concerned itself with epistemological and ontological issues of philosophical import.

6.2 Fundamentals

This exploratory theoretical undertaking covered issues ranging from epistemological and ontological concerns surrounding personal philosophical affiliations and how they may or may not have impinged on the study at hand and viewed the following issues in turn:

- The socio-geo-political and educational contexts in which the study is located and how this has influenced the area under investigation as well as how this has influenced the researcher’s orientations
- Personal experiences within psychological assessment and various implicit assumptions made manifest in the study which were subsequently informally viewed in terms of the psycho-epistemological profile as delineated by Joseph Royce and included the following dimensions:
 - Rationalistic ; Intuitive (metaphorism); Empirical; Realistic; Idealistic and Authoritarian
- The study’s main focus and area of concern in terms of the underpinnings of what was discussed, was the fundamental core of epistemology and ontology, which had important (perhaps imperative) ties with the following linked issues:
 - The mind-body “problem”; consciousness; G - dominated vs. multiple intelligence (MI) - dominated leanings towards the understanding of intelligence; emergence (irreducibility) vs. reductionism; realist vs. relativist approaches towards research; nature/nurture and static - dynamic assessment of intelligence and potential. These issues are considered pivotal within the assessment realm precisely due to their often implicit status permeating studies dealing with intelligence and assessment and in particular dynamic assessment. The main intention with this section was to make consistent the author’s qualms about certain issues which in turn would be necessitated in the subsequent discussions. Plainly, if the author was to critique implicit assumptions in others, her own assumptions would by the very same token need to included

The case for unappealing dichotomies was made in terms of the aforementioned dimensions of concern to the area of psychological assessment. The conviction for the need for an assessment calling itself dynamic is made alongside the at-times seemingly opposed notions of the very basic outlook towards human beings that the author posited. Although the core of dynamic assessment philosophy emanates from an implicit and explicit notion of change and innate potential recognition and one which is shared by the author, one cannot turn a blind eye to issues with which the method of assessment grapples and one also cannot stay quiet on certain very prominent yet often implicitly stated issues in terms of how the functioning human being is

¹ As such no new information will be referred to.



viewed from various angles. These angles were highlighted above. To mention one aspect in this concluding chapter: the conclusions drawn in chapter 2 illustrated the view of the biological human being within the socio-cultural context and not the other way round. Taking cognisance of the important role of context is secondary to the notion of biological innate predispositions; a view perhaps at odds with the notions considered as paramount within dynamic assessment. It is this and other reasons why the field finds itself in the situation it currently does. This is of course a thesis only; but one which is propounded and defended throughout.

Dynamic assessment within intelligence, its fundamental concerns, its historical narrative, Soviet influences in the future course of this method and lastly current trends were highlighted in the next major section of chapter 2. The methodological concerns within this approach were discussed as they pertained to their entrenchment within the larger domain of intelligence assessment. Constant reminders about the method's placement within psychology served to underlie the importance attributed to the method's awareness of the biological basis of functioning. All its notable predecessors are acknowledged supporters of the need to understand physiological underpinnings of human behaviour and how therapeutic interventions take place within such a context. Stalwarts in the domain received a critical overview in terms of their major contributions to the method and how their histories coloured their future work, being contextualised in both time and place. Points of emphases included in this discussion looked at the suitability of construct definition within both intelligence and dynamic assessment arenas. This discussion also investigated the disparate and overlapping concerns in dynamic assessment as 'helping discipline' and dynamic assessment as progressive science each viewed from their own vantage point and specific mandated platform of expertise. The reason for having chosen to briefly highlight the climate within which Vygotsky worked was due to the particular nature of ideology practised within the Soviet Union at the time of his theorising. Nowhere it is more evidenced that prevailing climate dictates science to a large degree and this aspect features in chapter 3 during the discussion of the deployment of Madsen's framework which is premised upon reigning meta concerns. Sociological approaches towards the study of human beings included Marxist views of the nature of existence and humankind's place within the larger system. Added to this ideology was a central preoccupation of all members of society as equally endowed intellectually which of course had a resultant effect on the construal of *g*-based research. This forms a link to intelligence research in the early twentieth century in the Soviet Union. To further contextualise Vygotsky's work, an even briefer overview of Soviet psychology was given so as to allow for a more refined take on the history of this discipline in a country considered as alien to the West for much of its history.

Intelligence assessment has become very much a franchise enterprise with a past and future which seems to have forever existed and will seemingly continue to exist in the future. Retracing the historical roots of intelligence and understanding the climate in which current researchers work is sobering for studies such as this one which are attempting to spur the field to re-look its reason for existence. Not only is the assessment field riddled with construct definition fuzziness but it has also reached a stalemate of sorts in its regard for the individual as static being. Which is where dynamic assessment enters the picture. Displaying what at times seem to be contrary notions of the understanding of the governing forces in human development, both intelligence and dynamic assessment as coalescent whole have much to offer the larger assessment machine. Many progressive avenues are being pursued in intelligence research particularly in the domain of physiological basis of intelligence. Quantitative trait loci is a case in point but is perhaps still far too distant a realm with which dynamic assessment can align. Although this can be debated. Intelligence research and its preoccupation in many instances with *g*-based research surfaces in many dynamic assessment batteries. The argument here is that if dynamic assessment is to follow in the wake of intelligence research then it should guard against the as yet unresolved issues and pursue a novel avenue towards assessment in terms of intelligence fundamentals. The chapter ended with an overview on cognition and its place within intelligence assessment as it is cognitive functions which are most often assessed for in intelligence and dynamic assessment batteries as well as the increasingly important role accorded non-intellective factors in assessment which play a crucial role in many dynamic assessment models. Once the fundamentals were firmly in place the next issue of ensconced framework was discussed in chapter 3.

6.3 Theory, meta-theory, science, the social sciences, psychology and assessment

Chapter 3 emphasised the importance attributed to the discussion of theory, meta-theory, theory appraisal, science, the social sciences and psychological assessment in this exploratory study. The necessary preliminaries were discussed and viewed in terms of psychology's placement within the greater realm of science; an issue which is prominent within the philosophy of the social sciences and one which impinges heavily on the placement of assessment within a designated area entitled "science". This subject has yet a long road to travel before it reaches anywhere near a limit of exhaustion. The at times moot area concerning the consilience of both the natural and social sciences was discussed and, as with almost every subject within academia, certain viewpoints are considered worth many volumes' discussion whilst others will view these same issues as a waste of time. A stance is taken and defended throughout; which is, if the social sciences and in particular psychological intelligence assessment is to continue in a progression of greater knowledge acquisition then it will have to consider a greater propensity to merge with the methodology and ideas engrossed within the natural sciences.

The notion of science in general was discussed along with its preferred tried-and-tested methods which were not always successful and when taken as benchmark, leave the social sciences pitifully behind. This discussion was almost always leveraged on a philosophy of science vantage point. Explanatory mechanisms within science were discussed next followed by an exposition on scientific theories, which, throughout the discussions were tied up to certain key issues highlighted in chapter 2. The social sciences were then overviewed but was done so with the constant hovering of natural science philosophy overhead. Explanatory mechanisms within the social sciences were the next to receive attention followed by the coverage of psychology as discipline within science in general and its explanatory mechanisms and theory development. Theory appraisal is perhaps one of the most contentious issues within the philosophy of science and much hinges around the need to develop for science and the social sciences a workable framework from which to base and compare various theories in varied domains. However the natural sciences might proceed with debates on this aspect, one can be assured that the debates in the social sciences camp are more heated, voluminous and fractured. The over-reliance of psychology on natural science models of how science should and does progress was discussed with the intention of initiating for psychology some alternative path which to forge for itself. There is no such theory appraisal framework set in stone but only guidelines which, it is hoped, will aid in the navigation towards an end goal for social science theory and in particular psychological assessment theory. A considerable number of issues are far from resolution yet many disciplines are expected to make headway in as efficient a manner as possible. This mandate is one with which the disciplines are struggling to cope, assessment being one such subsidiary area of research.

Once various elements such as concepts, definitions and propositions within theory development and theory appraisal were discussed, attention was turned towards the development and deployment of a meta-theoretical framework from which to contextualise the study. The chosen framework was based heavily on the ideas of the late K.B. Madsen, a Danish meta-psychologist, whose framework, although dated in approach towards the study of overarching meta-theory development, was deemed a suitable choice for utilisation. Meta-theory is not a developed subject area within psychology and in some quarters it is considered an enterprise lacking in stable foundations and certainty in criteria. Moreover, detractors claim that deployment of just such a framework is far too early an attempt to bring to the field of psychology a limiting or delimiting framework; psychological theory simply is not as evolved as it should be for such an endeavour to proceed. The author has taken an alternative view and whilst acknowledging inherent limitations in any framework utilised to bring the field some semblance of coherence, decided to proceed along the lines of comparative theory/model analysis of dynamic assessment within intelligence assessment. Although psychology literature was available which dealt with issues concerning meta-theory, no single coherent framework was located or known to the author at the time of writing other than the framework developed by Madsen and lest this seem to be the sole reason for making use of the framework, a claim to the contrary can be made as there are many aspects discussed within Madsen's framework that are of importance and that linked up elsewhere with various epistemological and ontological dynamic assessment issues. The choice of Madsen only came after the initial discussions in chapter 2. The framework was timely and applicable yet had to undergo revision and attenuation if it was to serve a more apt purpose within this study.

Madsen's framework was discussed in terms of how he viewed psychology during the 1950's, specifically motivational psychology and the numerous theories then pervading the field. Entrenched in the language of stimulus-response, Madsen's framework might appear dated but his strategic emphasis on core philosophical issues is pertinent to any discussion concerning theory comparison and evaluation. His unique² method of calculating the empirical nature of theories and thus turning towards accountability within theories was particularly incisive even in its simplicity. In order to attenuate his framework, however, further deliberation of psychology's prime considerations were necessary for an all-inclusive meta-theory framework.

6.4 Prime considerations: quantification as imperative

At what might at first glance have appeared to be a rather disparate area of concern within this study, following so closely after chapter 3. Chapter 4 focused on prime considerations within assessment as posited within this thesis. A major thesis within this study was the lack of acknowledgement of the role that the philosophy of mathematics and more importantly the foundations of mathematics play within assessment. In order to understand the almost obsessive need to measure, analyse and in turn reify constructs within psychology, an evaluation of what it means to count, measure, manipulate and statistically analyse constructs was necessitated as paramount. In order to understand psychological measurement, merely viewing measurement theory is not enough because this view is already a view predicated on a number of implicit assumptions and it is these implicit assumptions which form the backbone of measurement. If one does not question the fundamentals preceding measurement per se, one is likely to forever traverse in a circle. An analogy might be of some service here: in the attempt to alleviate poverty by dumping stocks of food at food stations and kitchens in poverty-struck areas is one teaching the necessary and requisite skills to people in order to aid them in their own betterment? In order to alleviate such poverty the first course of action would be to get people to help themselves as opposed to donating yet more food that they could be harvesting. The root then of measurement issues

² As far as this author is currently aware, Madsen's HQ calculation is the only one of its kind. The author could very well be in error.



within psychology will not be found with the techniques (many of which are sound) but with the fundamental underlying philosophies and principles governing what it means to measure in the first place and to decide whether the area of assessment truly is amenable to such measurement in the first place. The omnipresent quantitative imperative concern within psychological assessment served as introduction to the sub-sections to follow. In order for psychology to remain on a supposedly equal footing with the natural sciences, the need to quantify and subsequently measure psychological attributes has been and still is deemed of the utmost importance. Yet, the foundational concerns within quantified measures are debateable. The prevailing logic of the need to quantify then measure and lastly to conclude scientifically utilising this method as base has resulted in the erroneous use of measurement especially in areas such as dynamic assessment which could quite easily avoid the trap that has already befallen mainstream assessment. The notion of a valid construct which could be measured was flawed in two ways: there was no construct that was proved valid and secondly quantification of the construct has not helped in the measurement process. This state of affairs is compounded by the illogical silence on the part of the research community which, through its silence, practically endorses the view.

In an attempt to evaluate the predicates of measurement, mathematical philosophy and history were briefly sketched and it can clearly be seen that the very pure objective science of mathematics is itself merely another discipline bogged down in its own theoretical and philosophical issues. Far from a clear-cut understanding of the enterprise of mathematics, one is confronted with predicates in need of discussion and although the enterprise carries on in an almost unconcerned fashion (the need to establish its philosophical foundations is more of an issue to this study than it would perhaps be to applied mathematicians; this much is conceded) one has to question its roots too if clarity is to be sought concerning the just-as-vast enterprise of psychological measurement. Attention was next turned to statistics within psychology and assessment. Psychology's misplaced allegiance to null hypothesis significance testing is telling of its absurd rationalisation of working within a framework which is not even a correct one, at least historically. The miscegenation of Fisherian and Neyman-Pearson statistics and the ignorance of Bayesian statistics has had profound and detrimental effects within psychology and especially within assessment.

The remaining area of concern was the measurement arena. Measurement presupposes a quantified structure which assumes an additive structure, axiomatic representation and analysis which translates directly from object to number. Unfortunately, much of psychology's measurement history has been a concerted effort at inadequate substantiation in terms of assuming such a relation to exist in the first place. Initial efforts called for a controlling process of error detection, reliability considerations and inferred true scores from a reality that did not yield such true scores. Techniques attempting to control for as many extraneous factors as possible resulted in a theory of mental tests that was virtually unrivalled for many decades in term of noticeable competition. Modern attempts to derive for psychological measurement techniques more in keeping with axioms of representation and measurement have begun to permeate the larger core of work within the measurement arena but still have a way to go before it can be considered the main contender in the field of mental test theory.

6.5 The exploration of dynamic assessment within a meta-theoretical framework

Dynamic assessments are still not translatable into models or theories that are detailed and defined, which is hardly surprising as the intelligence field (under which dynamic assessment is situated) is still saddled with basic premise issues and confusion about definitions of fundamental concepts. From what grand theory does dynamic assessment emanate? Dynamic assessment models are often based on tenuous sentiments as opposed to solid theories of cognition and if a closer look was given to theory or sets of theories within cognitive psychology particular issues could be partially solved, as Deary (2001) states, "as psychology progressed through the twentieth century the attempts to relate the psychometric intelligence differences to cognitive elements were desultory until the rise of cognitive psychology in the 1970's" (p.129). Dynamic assessment has a base but it is very loose and is premised more so on initially intuitive ideas and concepts from which it has grown, without first having worked out the smaller problems which have now become contentious issues (such as change scores and how to assess and interpret them). A possible preface to this study may well have contextualised the author's predilection for favouring dynamic assessment's rationale and could possibly have been titled "A chain of proofs to the contrary"³ - dynamic assessment as evidencing hidden potential". An ironic title indeed given the origin of the cited quote. However, preferring to delve into deeper issues beyond that which is usually evidenced in favour of dynamic assessment's efficacy, the study highlighted areas of concern within each model or application analysed within the Madsenian framework. Dynamic assessment is by nature educationally-bound, in other words, this field more often than not finds applicability within the broader field of education and is firmly lodged within the sub-field of educational assessment (although by no means exclusively so). When Aldridge, Kuby and Strevy (1992) lamented the lack of meta-theory within education, similar sentiments were found to be echoed within the field of dynamic assessment years later. The authors acknowledged the differences between psychology and education as separate disciplines but the need for a metatheory in education as opposed to reigning psychological meta-theories was highlighted. The broad field of application

³ "A chain of proofs to the contrary" was taken from Francis Galton's 1869 text "Hereditary Genius" (extract in Sahakian, 1981, p.224) in which Galton made clear his understanding of the overpowering role and influence of heredity over-and-above environmental influences regarding intelligence and that 'natural equality' was a misnomer in every sense.



within which dynamic assessment finds itself would make for easier development if “a unifying whole or world view [were] proposed first in order to move from molar to molecular ideas in describing [dynamic assessment]” (p.686). It would be more feasible if dynamic assessment practitioners were to agree on certain core assumptions and issues within the assessment arena.

Working from test batteries and extrapolating back to what this author considers each models' premises and philosophical underpinnings, each model-analysis proceeded along exactly the same lines in order for a base measure of sorts to manifest. Thirteen such batteries were explored in this manner and each was quantitatively defined through its HQ score; a partial determiner of testability and falsifiability. A highlighted concern within each model was delineated and unpacked as it pertained to major discussions throughout the thesis and included issues such as construct definition, statistical and clinical decision-making, transfer issues, the nature of change and its construal as both novel and old construct, intelligence and dynamic assessment as two divergent entities not necessarily inhabiting the same level of description as well as *a priori* concerns within assessment. Concerns raised in chapters 2, 3 and 4 found fertile ground in each model, although it is ceded that such prior judgements were made at the outset of this study. Once highlighted, these chapters merely served to emphasise what can be considered to be perennial issues within dynamic assessment and intelligence. There was nothing surprising about each highlighted concern encountered within each model - it was expected. Although only very superficial, the HQ scores were tabulated from lowest to highest (most testable to least testable) and an average HQ score was calculated. The meta-model utilised to explore each model in turn is founded upon tenets of natural science dictates and practices. The HQ score as culmination or end-point of this meta-model makes sense within the ambit of its nomological-deductive framework. Utilising a quantified score in a study criticising the unthinking use of numbers can be considered ironic but the case for the utilisation of this has been discussed. Jensen's (2000) Mindladder model evidences the lowest HQ score of the models assessed in this study with Karpov and Gindis's (2000) model evidencing the highest HQ score. Is the goal of dynamic assessment as expressed through its varied models and approaches, to be testable, verifiable and reliable? The goals of CTT within psychological assessment would proffer this as good science within the discipline. And indeed, which model would not wish to have as its underpinning solid reliable bases from which to develop? The question posed within this study is the unquestioning pandering to CTT-like approaches within a sub-domain that does not necessarily have to rise to these tasks in the first place.

The tabulation of lowest to highest HQ score denotes nothing more than one framework's deployment of method of accounting for testability. Nothing more. It brings to light numerous other issues too without which the discipline would perhaps be the poorer but in no way can it be said to be a final answer on the question of dynamic assessment's ill-fit within intelligence assessment and the broader psychological assessment domain. Dynamic assessment as methodology as opposed to dynamic assessment as model (as manifested in batteries) are two approaches with two very different ideas about assessment. The one does not necessarily give to the discipline something the other does not but the differences between construct innovator and construct poacher is particularly significant to the debate surrounding dynamic assessment's place within intelligence assessment. Being a construct innovator pushes dynamic assessment into an area filled with novel meanings and novel methodologies (IRT change-based assessments) whereas dynamic assessment as construct poacher remains lodged within the traditional realm of intelligence assessment unable to escape a noose too constrictive for it to burgeon into a more fully developed domain. One solution is tantalisingly within reach: remain as construct poacher within the traditional realm of assessment but acquiesce to the tenets as specified for this domain and in so doing enrich an already fertile area of intelligence assessment or move away to newer grounds as construct innovator and lay the foundations for an entirely new domain which can develop alongside its traditional counterpart. Neither is wrong or correct as such; this is not a competition in which victory is sought over one or other domain; it is all about practising a science within a new science-framework or existing in an already carved out science framework. But as it currently stands the situation does not look promising.

6.6 Limitations of the study

All studies have their limitations and this is particularly the case with this study. Although strong contentions are made and firmly entrenched arguments are offered in defence of the numerous theses housed within this study, there are yet a number of limits within the conclusions. Firstly, as with most studies time and resources were of prime concern and no more will be said of these. Secondly, many arguments within the study, although cognisance was given to opposing views, were supported as if they presented themselves on a dichotomous continuum. This is not the case in most instances of philosophical debate and less often the case with psychology. Yet, in order to firmly establish at the outset certain key positions held within the study, such strategic placements had to be made in order for a consistent case to be maintained throughout. Thirdly, the main research source utilised within this study was the literature. Apart from the meta-analysis of dynamic assessment studies in South Africa and the content analysis of returned questionnaires, no statistical or practical methodology was used in support of any of the contentions made herein. Fourthly, philosophical issues are perhaps the most debated issues in any area of concern and are thus prone to attack in many guises. The thesis, of course, might be wholly incorrect from certain vantage points and no finality will ever come from these debates when conducted on paper. Fifthly, the statistical robustness of the meta-analytic results is tentative as it is based on the initial small number of studies included which is not a fault of this thesis but is an indication of the quantity and



quality of dynamic assessment research in South Africa. The low response rate to the questionnaire also does not bode well for those interested in keeping the area alive although the reasons behind the lack of responses were detailed in appendix 2.

The true litmus test, will undoubtedly be the future historical record. There is thus a looming danger of being hopelessly incorrect in any or all of the manifest assumptions argued for in this study. In this manner then, it should be regarded as precursory and tentative and in keeping with its title: a mere exploratory study into the at-times uncharted territory of dynamic assessment and intelligence. Given ten years, this study might well look completely different and early attempts might come across as quite tentative. The need to read was halted at a point in the study as there was always the tendency to obtain yet more sources. It is lamentable that there are numerous books and articles that could have been cited but a demarcating line had to be drawn and this line was drawn toward the end of 2005. The choice of models and theories assessed in chapter 5 was chosen due to their availability of signifiable constructs which allowed for the calculation of Madsen's HQ and due to their history within the field of assessment as existing models and applications with at least some research results to their name. This can be considered a limitation as *a priori* characteristics were conceived of beforehand in order to effect an HQ and newer models (since 2000) may well have been left out. Inclusion of 'dynamic assessment as method' would nullify the need for an HQ and thus would not allow for it to be placed on the HQ spectrum and this also resulted in the non-inclusion of dynamic assessment as clinical tool only (where the process of engagement is emphasised over-and-above demonstrable dependent and independent variables). This limits the nature of the dynamic assessment model or theory included in the study and in so doing the highlighted models were not necessarily representative of dynamic assessment per se.

As was evident from chapter 5 and the analysis taking place within one specific meta-framework, certain themes continually arose across models making manifest the perceived notion at the outset of the study that certain issues were in need of attention at least at an exploratory level. The mandate of taking assessment back to its measurement roots resulted in comments initially lodged at dynamic assessment in particular but was found to be applicable at a level presiding over this small area of concern within psychology. The issues did not stop there and progressed beyond the 'narrow' confines of psychology as scientific discipline and found its way to the pinnacle of the hierarchy: the governing social sciences. Comments, criticisms and thoughts were thus not limited to dynamic assessment as such. Placing dynamic assessment as this study saw fit results in a singular view of anticipating what the next step will be in its development as method, model and theory. As some comments have shown in both the literature and in the returned questionnaires, the feasibility of situating the study within a meta-theory is itself open to debate. But this latter assertion is precisely the point within a progressive discipline's evolution. Due to the context of the study, the varying emphases throughout on measurement, philosophy, science, meta-theory, meta-analysis, content analysis, item response theory (as it pertained to change scores) and analysis of models in a very specified fashion results in limited inferences that can be drawn if at all. The thesis has endeavoured to stay close to its initial intent: to explore a meta-theoretical framework for dynamic assessment and intelligence.

References¹

- Abbott, E. A. (1992). **Flatland: a romance of many dimensions**. Toronto: Dover
- Abelson, R.P. (1997a). On the surprising longevity of flogged horses: why there is a case for the significance test. **Psychological science**, **8**(1), 12-17.
- Abelson, R.P. (1997b). A retrospective on the significance test ban of 1999. (If there were no significance tests, they would be invented). In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.117-141). New Jersey: Lawrence Erlbaum.
- Abercrombie, N., Hill, S. & Turner, B.S. (2000). **The Penguin dictionary of sociology**. London: Penguin.
- Abrams, D. & Hogg, M.A. (2004). Metatheory: lessons from social identity research. **Personality and social psychology review**, **8**(2), 98-106.
- Abulkhanova-Slavskaya, K.A. & Brushlinsky, A.V. (1996). Theory in Soviet psychology: major dynamics. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.187-206). London: Greenwood Press.
- Achinstein, P.(2001). Observation and theory. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.325-334). Massachusetts: Blackwell.
- Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. **Journal of experimental psychology**, **117**(3), 288-318.
- Ackerman, P.L. (1994). Intelligence, attention, and learning: maximal and typical performance. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4: theories of intelligence** (pp.1-27). New Jersey: Ablex.
- Ackerman, P.L. (2005). Ability determinants of individual differences in skilled performance. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.142-159). Cambridge: Cambridge University Press.
- Ackerman, P.L. & Beier, M.E. (2005). Knowledge and intelligence. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.125-139). London: Sage
- Adams, R.J., Wilson, M. & Wang,W-C. (1997). The multidimensional random coefficients multinomial logit model. **Applied psychological measurement**, **21**(1), 1-23.
- Agassi, J. (1981). **Science and society: studies in the sociology of science**. Dordrecht: D. Reidel.
- Ageyev, V.S. (2003). Vygotsky in the mirror of cultural interpretations. In A. Kozulin, B. Gindis, V.S.Ageyev & S.M. Miller (Eds.), **Vygotsky's educational theory in cultural context** (pp.432-449). Cambridge: Cambridge University Press.
- Aldridge, J., Kuby, P.& Strevy, D. (1992). Developing a metatheory of education. **Psychological reports**, **70**, 683-687.
- Alfassi, M. (2002). Communities of learners and thinkers: the effects of fostering writing competence in a metacognitive environment. **Journal of cognitive education and psychology**, **2**(2), 124-140.
- Al-Khalili, J. (2003). **Quantum: a guide for the perplexed**. London: Weidenfeld & Nicolson.
- Allal, L. & Ducrey, G.P. (2000). Assessment of - or in - the zone of proximal development. **Learning and instruction**, **10**(2), 137-152.
- Altarriba, J. (Ed.), (1993). **Cognition and culture: a cross-cultural approach to cognitive psychology**. London: North Holland.
- Altarriba, J. (1993). The influence of culture on cognitive processes. In J. Altarriba (Ed.), **Cognition and culture: a cross-cultural approach to cognitive psychology** (pp.379-384). London: North Holland.
- Anastasi, A. (1988). **Psychological testing 6th edition**. New York: MacMillan.
- Anastasi, A. (1998). Intelligence as a quality of behaviour. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.19-21). New Jersey: Ablex.
- Anderson, J.A., Silverstein, J.W., Ritz, S.A. & Jones, R.S. (1977). Distinctive features, categorical perception, and probability learning: some applications of a neural model. **Psychological review**, **84**(5), 413-445.
- Anderson, M. (1992). **Intelligence and development: a cognitive theory**. Oxford: Blackwell.
- Anderson, M. (Ed). (1999). **The development of intelligence**. Sussex: Psychology Press.
- Anderson, M. (1999a). Project development - the shape of things to come. In M. Anderson (Ed.), **The development of intelligence** (pp.3-15). Sussex: Psychology Press.
- Anderson, M. (1999b). Project development - taking stock. In M. Anderson (Ed.), **The development of intelligence** (pp.311-332). Sussex: Psychology Press.
- Anderson, M. (2005a). Marrying intelligence and cognition: a developmental view. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.268-287). Cambridge: Cambridge University Press.
- Anderson, M. (2005b). Cortex forum on the concept of general intelligence in neuropsychology. **Cortex**, **41**(2), 99-100.
- Anderson, M. (2006). What we need is better theory, not more data. **Behavioral and brain sciences**, **29**, 125-126.

¹ Studies cited in Appendix 1 section A and section B are included in this reference list.

- Anderson, O.R. (1997). A neurocognitive perspective on current learning theory and science instructional strategies. **Science education**, **81**(1), 67-89.
- Andrén, G. (1981). Reliability and content analysis. In K.E. Rosengren (Ed.), **Advances in content analysis** (pp.43-67). Beverly Hills: Sage.
- Andrews, S.G. (1996). **A small-scale investigation of the group administration of Feuerstein's learning potential assessment device**. Unpublished M.Ed. dissertation, University of Natal, Pietermaritzburg.
- Andrey Kolmogorov. (2006). Retrieved March 6 2006 from <http://en.wikipedia.org/wiki/Kolmogorov>.
- Anrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? In E.V. Smith & R.M Smith (Eds.), **Introduction to Rasch measurement: theories, models and applications** (pp.143-166). Minnesota: JAM.
- Angus, J. (1985). The Luria model of information processing. **Australian journal of educational technology**, **1**(1), 59-67.
- Applebaum, A. (2003). **Gulag: a history**. London: Penguin.
- Archer, M. & Green, L. (1996). Classification of learning difficulties. In P.Engelbrecht, S. Kriegler & M. Booysen (Eds.), **Perspectives on learning difficulties: international concerns and South African realities** (pp.123-129). Pretoria: Van Schaik.
- Archer, R.P.(1996). MMPI-Rorschach interrelationships: proposed criteria for evaluating explanatory models. **Journal of personality assessment**, **67**(3), 504-515.
- Arievitch, I.M. & Stetsenko, A. (2000). The quality of cultural tools and cognitive development: Gal'perin's perspective and its implications. **Human development**, **43**, 69-92.
- Arija, V., Esparó, G., Fernández-Ballart, J., Murphy, M.M., Biarnés, E. & Canals, J. (2006). Nutritional status and performance in test of verbal and non-verbal intelligence in 6 year old children. **Intelligence**, **34**(2), 141-149.
- Arredondo, P., Rosen, D.C., Rice, T., Perez, P. & Tovar-Gamero, Z.G. (2005). Multicultural counselling: a 10-year content analysis of the Journal of counselling and development. **Journal of counselling and development**, **83**, 155-161.
- Arthur, W., Bennett, W. & Huffcutt, A. (1994). Choice of software and programs in meta-analysis research: does it make a difference? **Educational and psychological measurement**, **54**(3), 776-787.
- Asbury, K., Wachs, T.D. & Plomin, R. (2005). Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. **Intelligence**, **33**(6), 643-661.
- Ashman, A.F. (1992). Process-based instruction: integrating assessment and instruction. In H.C. Haywood & D. Tzuril (Eds.), **Interactive assessment** (pp.375-396). New York: Springer-Verlag.
- Ashman, A.F. & Conway, R.N.F. (1997). **An introduction to cognitive education: theory and applications**. London: Routledge.
- Asimov, I. (1987). **Asimov's new guide to science**. Middlesex: Penguin.
- Asimov, I. (1993). **Understanding physics vols. 1-3**. New York: Barnes and Noble.
- Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: a proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), **The psychology of learning and motivation: advances in research and theory (Vol 2)** (pp.90-195). New York: Academic Press.
- Augoustinos, M. & Walker, I. (1995). **Social cognition: an integrated introduction**. London: Sage.
- Austin, S. & Prilleltensky, I. (2001). Contemporary debates in critical psychology: dialectics and syntheses. **Australian psychologist**, **36**(1), 75-80.
- Avtonomova, N. (1995). On some contemporary tendencies in Russian-Soviet psychology. In I. Lubek, R. Van Hezewijk., G. Pheterson & C.W. Tolman (Eds.), **Trends and issues in theoretical psychology** (pp.95-101). New York: Springer.
- Ayer, A.J. (1983). The *a priori*. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.315-328). Cambridge: Cambridge University Press.
- Babad, E.Y. (1977). Pygmalion in reverse. **Journal of special education**, **11**(1), 81-90.
- Babad, E.Y. & Budoff, M. (1974). Sensitivity and validity of learning-potential measurement in three levels of ability. **Journal of educational psychology**, **66**(3), 439-447.
- Bacon, F. (1959). Preface to "Novum Organum". In G. Schwartz & P.W. Bishop (Eds.). **Moments of discovery: the origins of science. Vol 1**. (pp.33-36). New York: Basic books.
- Bagley, C. & Verma, G.K. (Eds.), (1986). **Personality, cognition and values**. London: Macmillan.
- Bailer-Jones, D.M. (2002). Models, metaphors and analogies. In P.Machamer & M. Silberstein (Eds.), **The Blackwell guide to the philosophy of science** (pp.108-127). Massachusetts: Blackwell.
- Baker, B.D. & Richards, C.E. (1999). A comparison of conventional linear regression methods and neural networks for forecasting educational spending. **Economics of education review**, **18**, 405-415.
- Baker, E.L. (2001). Testing and assessment: a progress report. **Educational assessment**, **7**(1), 1-12.
- Baker, F.B. (2001). The basics of item response theory. ERIC: Eric Clearinghouse on assessment and evaluation (available online from <http://ericae.net/irt>).
- Baker, W.J., Hyland., M.E., Van Rappard, H & Staats, A.W. (Eds.), Current issues in theoretical psychology: selected/edited proceedings of the founding conference of the international society for theoretical psychology held in Plymouth, U.K., 30 August - 2 September, 1985. Amsterdam: North-Holland.
- Ball, P.(2005). **Critical mass, how one things leads to another: being an enquiry into the interplay of chance and necessity in the way that human culture, customs, institutions, cooperation and conflict arise**. London: Arrow.



- Baltes, P.B. (1998). Notes on the concept of intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.23-27). New Jersey: Ablex.
- Balzer, W. (1992). The structuralist view of measurement: an extension of received measurement theories. In C.W. Savage & P.Ehrlich (Eds.), **Philosophical and foundational issues in measurement theory** (pp.93-117). New Jersey: Lawrence Erlbaum.
- Bangert-Drowns, R.L. (1992). Review of developments in meta-analytic method. In A.E. Kazdin (Ed.), **Methodological issues and strategies in clinical research** (pp.439-467). Washington: American Psychological Association.
- Baral, B.D. & Das, J.P. (2004). Intelligence: what is indigenous to India and what is shared? In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.270-301). **Cambridge: Cambridge University Press.**
- Barber, N. (2005). Educational and ecological correlates of IQ: a cross-national investigation. **Intelligence, 33**, 273-284.
- Bardis, P.D. (1985). Jensen, Spearman's *g*, and Ghazali's dates: a commentary on interracial peace. **Behavioral and brain sciences, 8**(2), 219-220.
- Barendregt, M. (2003). Interlevel relations and the prospects for an evolutionary psychology. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.396-404). Ontario: Captus University Publications.
- Barker, P.(2001). Kuhn, incommensurability, and cognitive science. **Perspectives on science, 9**(4), 433-463.
- Baron, J. (1985). Reliability and *g*. **Behavioral and brain sciences, 8**(2), 220-221.
- Baron, J. (1998). Capacities, dispositions, and rational thinking. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.29-33). New Jersey: Ablex.
- Baron, R.A. & Byrne, D. (1997). **Social psychology**. Boston: Allyn and Bacon.
- Barlow, R.E. (1992). Introduction to De Finetti (1937) "Foresight: its logical laws, its subjective sources". In S. Kotz & N.L. Johnson (Eds.), **Breakthroughs in statistics volume 1: foundations and basic theory** (pp.127-133). New York: Springer-Verlag.
- Barrett, P.T. (1998, June). **Science, Fundamental Measurement, and Psychometric Testing**. Paper presented at the BPS 3rd Test User Conference (Keynote Speaker), Brighton, United Kingdom. Published in Conference Proceedings, British Psychological Society, pp 11-20.
- Barrett, P.T. (2000, September). **Quantitative Science and Individual Differences: candidate models for intelligence and personality**. Paper presented at the ECP10 Conference, Kraków, Poland.
- Barrett, P.T. (2001, December). **The role of a concatenation unit**. Paper presented at the BPS Maths-Stats-Computing Meeting, London, United Kingdom.
- Barrett, P.T. (2002, April). **Measurement cannot occur in a theoretical vacuum**. Paper presented at the AERA Annual educational measurement conference, Rasch measurement SIG, New Orleans, United States of America.
- Barrett, P.T. (2003). Beyond psychometrics: measurement, non-quantitative structure, and applied numerics. **Journal of managerial psychology, 18**(5), 421-439.
- Barrett, P.T. (2005). What if there were no psychometrics? Constructs, complexity, and measurement. **Journal of personality assessment, 85**(2), 134-140.
- Barrow, J.D. (1992). **Theories of everything: the quest for the ultimate explanation**. London: Vintage.
- Barrow, J.D. & Tipler, F.J. (1996). **The anthropic cosmological principle**. Oxford: Oxford University Press.
- Barrow, J.D. & Webb, J.K. (2005). Inconstant constants: do the inner workings of nature change with time? **Scientific American, 292**(6), 56-63.
- Barton, R.A. (1997). Neural constructivism: how mammals make modules. **Behavioral and brain sciences, 20**(4), 556-557.
- Bates, E.A. & Elman, J.L. (1994). Connectionism and the study of change. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.623-642). Oxford: Blackwell.
- Bateson, M.C. (1995). On the naturalness of things. J. Brockman & K. Matson (Eds.). **How things are: a science tool-kit for the mind** (pp.9-16). London: Weidenfeld and Nicolson.
- Bauer, R.A. (1959). **The new man in Soviet psychology**. Cambridge: Harvard.
- Beals, D. (2000). Book review: Vygotsky and cognitive science. **Applied psycholinguistics, 21**, 425-427.
- Beatty, J. (1995). **Principles of behavioural neuroscience**. Madison: Brown and Benchmark.
- Beaujean, A. A. (2005). Heritability of cognitive abilities as measured by mental chronometric tasks: a meta-analysis. **Intelligence, 33**, 187-201.
- Bechtel, W. (1988). **Philosophy of science: an overview for cognitive science**. New Jersey: Lawrence Erlbaum.
- Bechtel, W., Abrahamsen, A. & Graham, G. (1998). The life of cognitive science. In W. Bechtel & G. Graham, (Eds.). **A companion to cognitive science** (pp.1-104). Massachusetts: Blackwell.
- Bechtel, W. & Graham, G. (1998). (Eds.). **A companion to cognitive science**. Massachusetts: Blackwell.
- Bechtel, W., Mandik, P. & Mundale, J. (2001). Philosophy meets the neurosciences. In W. Bechtel, P.Mandik, J. Mundale & R.S. Stufflebeam, (Eds.). **Philosophy and the neurosciences: a reader** (pp.4-22). Massachusetts: Blackwell.
- Bechtel, W., Mandik, P., Mundale, J. & Stufflebeam, R.S. (2001). (Eds.). **Philosophy and the neurosciences: a reader**. Massachusetts: Blackwell.
- Becker, B.J. (1994). Combining significance levels. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp215-230). New York: Russell Sage Foundation.

- Bedell, B., Van Eeden, R. & Van Staden, F. (1999). Culture as moderator variable in psychological test performance: issues and trends in South Africa. **Journal of Industrial psychology**, *25*(3), 1-7.
- Begg, C.B. (1994). Publication bias. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.399-409). New York: Russell Sage Foundation
- Bejar, I.I. (1984). Educational diagnostic assessment. **Journal of educational measurement**, *21*(2), 175-189.
- Benacerraf, P. (1983a). What numbers could not be. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.272-294). Cambridge: Cambridge University Press.
- Benacerraf, P. (1983b). Mathematical truth. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.403-420). Cambridge: Cambridge University Press.
- Benacerraf, P. & Putnam, H. (Eds.). (1983). **Philosophy of mathematics: selected readings**. Cambridge: Cambridge University Press.
- Benga, O. (2006). Heterogeneity in fluid cognition and some neural underpinnings. **Behavioral and brain sciences**, *29*, 126.
- Benjafield, J.G. (1993). **Cognition**. New Jersey: Prentice-Hall.
- Benjamin, L. & Lomofsky, L. (2002). The effects of observation of dynamic and static assessment on teacher's perception of learners with low academic results. **Journal of cognitive education and psychology**, *2*(2), 102-123.
- Ben-Shakhar, G. & Lieblich, A. (Eds.), (1995). **Studies in psychology in honour of Solomon Kugelmass**. Jerusalem: Magnes Press.
- Bereiter, C. (1962). Using tests to measure change. **Personnel and guidance journal**, *41*(1), 6-11.
- Bereiter, C. (1967). Some persisting dilemmas in the measurement of change. In C.W. Harris (Ed.), **Problems in measuring change** (pp.3-20). Wisconsin: University of Wisconsin Press.
- Berg, B.L. (2001). **Qualitative research methods for the social sciences**. Massachusetts: Allyn and Bacon.
- Berger, D.E., Pezdek, K. & Banks, W. (Eds.). (1987). **Applications of cognitive psychology: problem solving, education and computing**. New Jersey: Lawrence Erlbaum.
- Berger, J., Wagner, D.G. & Zelditch, M. (1989). Theory growth, social processes, and metatheory. In J.H. Turner (Ed.). **Theory building in sociology: assessing theoretical cumulation** (pp.19-42). Newbury Park: Sage.
- Berger, J. & Zelditch, M. (1993a). Orienting strategies and theory growth. In J. Berger & M. Zelditch (Eds.). **Theoretical research programs: studies in the growth of theory** (pp.3-19). California: Stanford University Press
- Berger, J. & Zelditch, M. (Eds.). (1993b) **Theoretical research programs: studies in the growth of theory**. California: Stanford University Press.
- Berka, K. (1992). Are there objective grounds for measurement procedures? In C.W. Savage & P.Ehrlich (Eds.), **Philosophical and foundational issues in measurement theory** (pp.181-194). New Jersey: Lawrence Erlbaum.
- Bernays, P. (1983). On Platonism in mathematics. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.258-271). Cambridge: Cambridge University Press.
- Bernoulli distribution. (2006). Retrieved March 6 2006 from http://en.wikipedia.org/wiki/Bernoulli_distribution.
- Bernstein, J. (1993). **Cranks, quarks, and the cosmos: Essays on science and scientists**. Oxford: Oxford University Press.
- Berry, J.W. (1996). A cultural ecology of cognition. In I.Dennis & P.Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.19-37). New Jersey: Lawrence Erlbaum.
- Berry, J.W. (1998). A cross-cultural view of intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.35-38). New Jersey: Ablex.
- Beth, E.W & Piaget, J. (1966). **Mathematical epistemology and psychology** (W. Mays, Trans.). Dordrecht: D. Reidel. (Original work published in 1961).
- Bethge, H-J., Carlson, J.S. & Wiedl, K.H. (1982). The effects of dynamic assessment procedures on Raven matrices performance, visual search behavior, test anxiety and test orientation. **Intelligence**, *6*, 89-97.
- Beveridge, M. (2000). Educational implementation and teaching: "school knowledge" and psychological theory. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.21-35). London: Routledge.
- Bickerton, D. (1997). Constructivism, nativism, and explanatory adequacy. **Behavioral and brain sciences**, *20*(4), 557-558.
- Bickle, J. (2003a). **Philosophy and neuroscience: a ruthlessly reductive account**. Dordrecht: Kluwer Academic.
- Bickle, J. (2003b). Philosophy of mind and the neurosciences. In S.P.Stich & T.A. Warfield (Eds.), **The Blackwell guide to philosophy of mind** (pp.322-351). Massachusetts: Blackwell.
- Bickley, R. (1977). Vygotsky's contributions to a dialectical materialist psychology. **Science and society**, *41*(2), 191-207.
- Bidell, T. (1988). Vygotsky, Piaget and the dialectic of development. **Human development**, *31*, 329-348.
- Bidell, T.R. & Fischer, K.W. (1997). Between nature and nurture: the role of human agency in the epigenesis of intelligence. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.193-242). Cambridge: Cambridge University Press.
- Biesheuvel, S. (1972). Adaptability: its measurement and determinants. In L.J. Cronbach & P.J.D. Drenth (Eds.), **Mental tests and cultural adaptation** (pp.47-62). The Hague: Mouton.
- Biggs, J.B. (1985). The role of metalearning in study processes. **British journal of educational psychology**, *55*, 185-212.
- Bigler, E.D. (1994). Neuroimaging and neuropsychological assessment, In C.R. Reynolds (Ed). **Cognitive assessment: a multidisciplinary perspective** (pp.1-34). New York: Plenum Press.



- Birenbaum, M. & Dochy, F.J.R.C. (Eds.), (1996). **Alternatives in assessment of achievements, learning processes and prior knowledge**. Boston: Kluwer.
- Birnbaum, R. & Deutsch, R. (1996). The use of dynamic assessment and its relationship to the Code of Practice. **Educational and child psychology**, **13**(3), 14-24.
- Bishop, D.I. & Solso, R.L. (1996). The institute of psychology revisited. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.124-134). London: Greenwood Press.
- Bjorklund, D.F. (1997). In search of a metatheory for cognitive development (or, Piaget is dead and I don't feel so good myself). **Child development**, **68**(1), 144-148.
- Black, J.E. & Greenough, W.T. (1997). How to build a brain: multiple memory systems have evolved and only some of them are constructivist. **Behavioral and brain sciences**, **20**(4), 558-559.
- Blackmore, S. (2000). **The meme machine**. Oxford: Oxford University Press.
- Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. **Behavioral and brain sciences**, **29**(2), 109-160.
- Blair, C., Gamson, D., Thorne, S. & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. **Intelligence**, **33**, 93-106.
- Blakemore, C. & Greenfield, S. (Eds.). (1989). **Mindwaves: thoughts on intelligence, identity and consciousness**. Oxford: Blackwell.
- Blinkhorn, S.F. (1997). Past imperfect, future conditional: fifty years of test theory. **British journal of mathematical and statistical psychology**, **50**, 175-185.
- Bly, B.M. & Rumelhart, D.E. (Eds.). (1999). **Cognitive science: handbook of perception and cognition**. London: Academic Press.
- Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: tracing the history of intelligence testing. **Journal of clinical and experimental neuropsychology**, **24**(3), 383-405.
- Boden, M.A. (1988). **Computer models of mind: computational approaches in theoretical psychology**. Cambridge: Cambridge University Press.
- Boekaerts, M. (1988). Arousal, telic dominance and learning behaviour. In R.M. Gupta & P.Coxhead (Eds.), **Cultural diversity and learning efficiency: recent developments in assessment** (pp.164-184). New York: St Martin's Press.
- Boeyens, J.C.A. (1989). **Learning potential: an empirical investigation**. Pretoria: Human Sciences Research Council.
- Bohm, D. (1997). **Causality and chance in modern physics**. London: Routledge.
- Bolhuis, J.J. (1997). Learning, development, and synaptic plasticity: the avian connection. **Behavioral and brain sciences**, **20**(4), 559-560.
- Bolig, E.E. & Day, J.D. (1993). Dynamic assessment and giftedness: the promise of assessing training responsiveness. **Roeper review**, **16**(2), 110-113.
- Bond, T.G. (2000). Measuring development: examples from Piaget's theory. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond** (pp.167-182). London: Routledge.
- Bondy, E. (1987). Thinking about thinking: encouraging children's use of metacognitive processes. **Journal for technical and vocational education in South Africa**, **113**, 7-10.
- Boniface, D.R. (1995). **Experiment design and statistical methods for behavioral and social research**. London: Chapman and Hall.
- Boolos, G. (1983). The iterative conception of set. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.486-502). Cambridge: Cambridge University Press.
- Borenstein, M. & Rothstein, H. (2005). **Comprehensive meta-analysis**. Retrieved April 13, 2005 from www.MetaAnalysis.com
- Borg, I. & Mohler, P. (Eds.), (1994). **Trends and perspectives in empirical social research**. Berlin: Walter De Gruyter.
- Boring, E.G. (1961). The beginning and growth of measurement in psychology. In H. Woolf (Ed.), **Quantification: a history of the meaning of measurement in the natural and social sciences** (pp.108-127). New York: Bobbs-Merrill.
- Borkowski, J.G. & Konarski, E.D. (1981). Educational implications of efforts to train intelligence. **Journal of special education**, **15**(2), 289-305.
- Borkowski, J.G. & Maxwell, S.E. (1985). Looking for Mr. good-*g*: general intelligence and processing speed. **Behavioral and brain sciences**, **8**(2), 221-222.
- Borland, J.H. & Wright, L. (1994). Identifying young, potentially gifted, economically disadvantaged students. **Gifted child quarterly**, **38**(4), 164-171.
- Borsboom, D. (2005). **Measuring the mind: conceptual issues in contemporary psychometrics**. Cambridge: Cambridge University Press.
- Borsboom, D. & Dolan, C.V. (2006). Why *g* is not an adaptation: a comment on Kanazawa (2004). **Psychological review**, **113**(2), 433-437.
- Borsboom, D. & Mellenbergh, G.J. (2002). True scores, latent variables, and constructs: a comment on Schmidt and Hunter. **Intelligence**, **30**, 505-514.
- Borsboom, D. & Mellenbergh, G.J. (2004). Why psychometrics is not pathological: a comment on Michell. **Theory and psychology**, **14**(1), 105-120.
- Borsboom, D., Mellenbergh, G.J. & Van Heerden, J. (2003). The theoretical status of latent variables. **Psychological review**, **110**(2), 203-219.



- Borsboom, D., Mellenbergh, G.J. & Van Heerden, J. (2004). The concept of validity. **Psychological review**, **111**(4), 1061-1071.
- Borsboom, D., Van Heerden, J & Mellenbergh, G.J. (2003). Validity and truth. In: H. Yanai, A. Okada, K. Shigemasa, Y. Kano, & J.J. Meulman (Eds.), **New developments in psychometrics. Proceedings of the International Meeting of the Psychometric Society 2001**(pp.321-328). Tokyo: Springer.
- Bouchard, T.J. (1997). IQ similarity in twins reared apart: findings and responses to critics. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.126-160). Cambridge: Cambridge University Press.
- Bowman, D.B., Markham, P.M. & Roberts, R.D. (2002). Expanding the frontier of human cognitive abilities: so much more than (plain) g! **Learning and individual differences**, **13**, 127-158.
- Boyer, C. & Merzbach, U.C. (1991). **A history of mathematics**. New York: John Wiley.
- Boyer, P.(1995). Ceteris Paribus (All else being equal). J. Brockman & K. Matson (Eds.). **How things are: a science tool-kit for the mind** (pp.169-175). London: Weidenfeld and Nicolson.
- Brainerd, C.J. (1993). Cognitive development is abrupt (but not stage like). **Monographs of the society for research in child development**, **58**(9), 170-191.
- Brand, C. (1985). Jensen's compromise with componentialism. **Behavioral and brain sciences**, **8**(2), 222-223.
- Brand, C., Egan, V. & Deary, I. (1994). Intelligence, personality, and society: constructivist versus essentialist possibilities. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4: theories of intelligence** (pp.29-42). New Jersey: Ablex.
- Brand, J.L. (1998). Theoretical psychology, yes; methodological egalitarianism, no! **American psychologist**, **53**(1), 66-67.
- Brandt, L.W. (1984). Logic and psycho-logic of science. In J.R. Royce & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 2** (pp.203-209). New York: Plenum Press.
- Bransford, J.D., Delclos, V.R., Vye, N.J., Burns, S. & Hasselbring, T.S. (1987). State of the art and future directions. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.479-496). New York: Guilford Press.
- Braybrooke, D. (1965). Introduction. In D. Braybrooke (Ed.). **Philosophical problems of the social sciences** (pp.1-18). New York: Macmillan.
- Braybrooke, D (Ed.). (1965). **Philosophical problems of the social sciences**. New York: Macmillan.
- Brennan, J.F. (1982). **History and systems of psychology**. New Jersey: Englewood Cliffs.
- Brinkley, A. (1993). **The unfinished nation: a concise history of the American people**. New York: McGraw-Hill.
- Broadbent, D.E. (1987). **Perception and communication**. Oxford: Oxford University Press.
- Brockman, J. & Matson, K. (Eds.). (1995). **How things are: a science tool-kit for the mind**. London: Weidenfeld and Nicolson.
- Brockmeier, J. (1996). Construction and interpretation: exploring a joint perspective on Piaget and Vygotsky. In A. Tryphon & J. Vonèche (Eds.), **Piaget-Vygotsky: the social genesis of thought** (pp.125-143). Sussex: Psychology Press.
- Brody, B.A. & Capaldi, N. (Eds.), (1968). **Science: men, methods, goals, a reader: methods of physical science**. New York: W.A. Benjamin.
- Brody, N. (1992). **Intelligence**. San Diego: Academic Press.
- Brody, N. (2004). History of theories and measurements of intelligence. In R.J. Sternberg (Ed.). **Handbook of intelligence** (pp.16-33). Cambridge: Cambridge University Press.
- Brody, N. (2005). To *g* or not to *g* - that is the question. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.489-502). London: Sage.
- Brogden, H.E. (1977). The Rasch model, the law of comparative judgment and additive conjoint measurement. **Psychometrika**, **42**(4), 631-634.
- Bronowski, J. (1974). **The ascent of man**. London: BBC.
- Brooks, R.A. (2003). **Robot: the future of flesh and machines**. London: Penguin.
- Brouwer, L.E.J. (1983a). Intuitionism and formalism. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.77-89). Cambridge: Cambridge University Press.
- Brouwer, L.E.J. (1983b). Consciousness, philosophy, and mathematics. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.90-96). Cambridge: Cambridge University Press.
- Brown, A.L. & Campione, J.C. (1998). Academic intelligence and learning potential. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.39-43). New Jersey: Ablex.
- Brown, A.L., Campione, J.C., Webber, L.S. & McGilly, K. (1993). Interactive learning environments: a new look at assessment and instruction. In B.R. Gifford & M.C. O'Connor (Eds.), **Changing assessments: alternative views of aptitude, achievement and instruction** (pp.121-211). Boston: Kluwer.
- Brown, A.L. & French, L.A. (1979). The zone of potential development: implications for intelligence testing in the year 2000. **Intelligence**, **3**, 255-273.
- Brown, B.L., Hendrix, S.B. & Williams, R.N. (2003). Deconstructing psychology's quantitative heritage: the "new look" of visual parables. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.78-89). Ontario: Captus University Publications.
- Brown, D.C. (1994). Subgroup norming: legitimate testing practice or reverse discrimination. **American psychologist**, **49**(11), 927-928.

- Brown, H.I. (2001). Role of judgement in science. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.194-202). Massachusetts: Blackwell.
- Brown, J.B. (2001a). Mathematics, role in science. In W.H. Newton-Smith (Ed.). **A companion to the philosophy of science** (pp.257-264). Massachusetts: Blackwell.
- Brown, J.B. (2001b). Social factors in science. In W.H. Newton-Smith (Ed.). **A companion to the philosophy of science** (pp.442-450). Massachusetts: Blackwell.
- Brown, J.B. (2001c). Who rules in science? An opinionated guide the wars. Massachusetts: Harvard University Press.
- Browne, M.W. & Du Toit, S.H.C. (1991). Models for learning data. In L.M. Collins & J.L. Horn (Eds.), **Best methods for the analysis of change: recent advances, unanswered questions, future directions** (pp.47-68). Washington: American psychological association.
- Brožek, J. (1966). Contemporary Soviet psychology. In N. O'Connor (Ed.), **Present-day Russian psychology: a symposium by seven authors** (pp.178-201). Oxford: Pergamon.
- Brožek, J. (1972a). History of Soviet psychology: some significant historical events in the development of Soviet Psychology. In J. Brožek & D.I. Slobin (Eds.), **Psychology in the USSR: An historical perspective** (pp.11-17). New York: IASP.
- Brožek, J. (1972b). Soviet psychology today. In J. Brožek & D.I. Slobin (Eds.), **Psychology in the USSR: An historical perspective** (pp.3-6). New York: IASP.
- Brožek, J. & Slobin, D.I. (1972). Noted figures in the history of Soviet Psychology: pictures and brief biographies. In J. Brožek & D.I. Slobin (Eds.), **Psychology in the USSR: An historical perspective** (pp.22-29). New York: IASP.
- Bruer, J.T. (1998). Education. In W. Bechtel & G. Graham, (Eds.). **A companion to cognitive science** (pp.681-690). Massachusetts: Blackwell.
- Bryant, A. (2003). Cognitive informatics, distributed representation and embodiment. **Brain and mind**, **4**, 215-228.
- Bryant, P. (2000). Piaget, mathematics and Vygotsky. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.131-144). London: Routledge.
- Bryder, T. (1981). Actor-reliability. In K.E. Rosengren (Ed.), **Advances in content analysis (pp.69-88)**. Beverly Hills: Sage.
- Buaer, R.A. (1959). **The new man in Soviet psychology**. Cambridge: Harvard University Press.
- Büchel, F.P. & Scharnhorst, U. (1993). The learning potential assessment device (LPAD): discussion of theoretical and methodological problems. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.13-18). Amsterdam: Swets & Zeitlinger.
- Budoff, M. (1987a). The validity of learning potential assessment. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.52-81). New York: Guilford Press.
- Budoff, M. (1987b). Measures for assessing learning potential. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.173-195). New York: Guilford Press.
- Budoff, M. & Friedman, M. (1964). "Learning potential" as an assessment approach to the adolescent mentally retarded. **Journal of consulting psychology**, **28**(5), 434-439.
- Budoff, M. & Hamilton, J.L. (1976). Optimizing test performance of moderately and severely mentally retarded adolescents and adults. **American journal of mental deficiency**, **81**(1), 49-57.
- Budoff, M., Meskin, J. & Harrison, R.H. (1971). Educational test of the learning-potential hypothesis. **American journal of mental deficiency**, **76**(2), 159-169.
- Bunge, M. (1985). From mindless neuroscience and brainless psychology to neuropsychology. In K.B. Madsen & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 3** (pp.115-133). New York: Plenum Press.
- Burden, R. (2000). Feuerstein's unique contribution to educational and school psychology. A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.45-54). Amsterdam: Pergamon.
- Burgess, J.P. (1992). Synthetic physics and nominalist realism. In C.W. Savage & P.Ehrlich (Eds.), **Philosophical and foundational issues in measurement theory** (pp.119-138). New Jersey: Lawrence Erlbaum.
- Burgess, R.V. (2000). Reuven Feuerstein: propelling change, promoting continuity. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.3-20). Amsterdam: Pergamon.
- Burns, M.S. (1996). Dynamic assessment: easier said than done. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.182-188). Ontario: Captus Press.
- Burns, M.S., Delclos, V.R., Vye, N.J & Sloan, K. (1996). Changes in cognitive strategies in dynamic assessment. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.96-102). Ontario: Captus Press.
- Burt, C. (1937). **The backward child**. London: University of London Press.
- Bushman, B.J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.3-14). New York: Russell Sage Foundation.
- Butterfield, E.C. (1998). Intelligent action, learning, and cognitive development might all be explained with the same theory. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.45-49). New Jersey: Ablex.
- Butterfield, E.C., Siladi, D. & Belmont, J.M. (1980). Validating theories of intelligence. **Advances in child development and behavior**, **15**, 95-162.

- Butterfield, E.C., Slocum, T.A. & Nelson, G.D. (1992). Cognitive and behavioral analyses of transfer of learning are functionally equivalent. In J.S. Carlson (Ed.), **Advances in cognition and educational practice: a research annual volume 1B** (pp.3-45). Connecticut: JAI Press.
- Butterworth, B. (2000). **The mathematical brain**. London: Papermac.
- Byrnes, J.P. (1996). **Cognitive development and learning in instructional contexts**. Boston: Allyn and Bacon.
- Cahan, S. & Noyman, A. (2001). The Kaufman ability battery for children mental processing scale: a valid measure of "pure" intelligence? **Educational and psychological measurement**, **61**(5), 827-840.
- Callaway, E. (1985). Event-related potentials and the biology of human information processing. **Behavioral and brain sciences**, **8**(2), 223-224.
- Calude, C.S. & Paun, G. (Eds.) (2000). **Finite versus Infinite: contributions to an eternal dilemma**. London: Springer-Verlag.
- Calvin, W.H. (1997). **How brains think: evolving intelligence, then and now**. London: Weidenfeld & Nicolson.
- Campbell, C. & Carlson, J.S. (1995). The dynamic assessment of mental abilities. In J.S. Carlson (Ed.), **European contributions to dynamic assessment volume 3** (pp.1-31). London: JAI.
- Campbell, D.T. (1986). Science's social system of validity-enhancing collective belief change and the problems of the social sciences. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.108-135). Chicago: University of Chicago Press.
- Campbell, D.T. (1989). History of psychological epistemology. In B. Gholsen, W.R. Shadish, R.A. Neimeyer, & A.C Houts, (Eds.). **Psychology of science: contributions to metascience (pp.21-46)**. Cambridge: Cambridge University Press.
- Campbell, F.A. & Ramey, C.T. (1990). The relationship between Piagetian cognitive development, mental test performance, and academic achievement in high-risk students with and without early educational experience. **Intelligence**, **14**(3), 293-308.
- Campbell, R.L. (1993). Epistemological problems for neo-Piagetians. **Monographs of the society for research in child development**, **58**(5-6), 168-191.
- Campbell, R.N. (2000). Expertise and cognitive development: seeking a connection. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.159-164). London: Routledge.
- Campione, J.C. (1989). Assisted assessment: a taxonomy of approaches and an outline of strengths and weaknesses. **Journal of learning disabilities**, **22**(3), 151-165.
- Campione, J.C. (1996). Assisted assessment: a taxonomy of approaches and an outline of strengths and weaknesses. In H. Daniels (Ed.), **An introduction to Vygotsky** (pp.219-250). London: Routledge.
- Campione, J.C. & Brown, A.L. (1987). Linking dynamic assessment with school achievement. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.82-109). New York: Guilford Press.
- Campione, J.C., Brown, A.L., Ferrara, R.A., Jones, R.S. & Steinberg, E. (1985). Breakdowns in flexible use of information: intelligence-related differences in transfer following equivalent learning performance. **Intelligence**, **9**, 297-315.
- Capaldi, E.J. & Proctor, R.W. (2000). Laudan's normative naturalism: a useful philosophy of science for psychology. **American journal of psychology**, **113**(3), 430-454.
- Capra, F. (1983). **The turning point: science, society and the rising culture**. London: Flamingo.
- Capra, F. (1997). **The web of life: a new synthesis of mind and matter**. London: Flamingo.
- Cardon, L.R. & Fulker, D.W. (1991). Sources of continuity in infant predictors of later IQ. **Intelligence**, **15**(3), 279-293.
- Carr, M. & Biddlecomb, B. (1998). Metacognition in mathematics from a constructivist perspective. In D.K. Hacker, J. Dunlosky & A.C. Graesser (Eds.), **Metacognition in educational theory and practice** (pp.69-92). New Jersey: Lawrence Erlbaum.
- Carlson, J.S. (Ed.), (1992). **Advances in cognition and educational practice: a research manual volume 1B**. Connecticut: JAI Press.
- Carlson, J.S. (1994). Dynamic assessment of mental abilities. In R. J. Sternberg (Ed.), **Encyclopedia of human intelligence vol 1** (pp.368-372). New York: Macmillan.
- Carlson, J.S. (Ed.), (1995). **European contributions to dynamic assessment. Advances in cognition and educational practice volume 3**. London: JAI Press.
- Carlson, J.S. (2002). Intelligence: perspectives and a model In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijssenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.11-27). Amsterdam: JAI.
- Carlson, J.S. & Wiedl, K.H. (1978). Use of testing-the-limits procedures in the assessment of intellectual capabilities in children with learning difficulties. **American journal of mental deficiency**, **82**(6), 559-564.
- Carlson, J.S. & Wiedl, K.H. (1992). The dynamic assessment of intelligence. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.167-186). New York: Springer-Verlag.
- Carlson, J.S. & Wiedl, K.H. (2000). The validity of dynamic assessment. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.681-712). New York: Elsevier.
- Carlstedt, B., Gustafsson, J-E. & Hautamäki, J. (2004). Intelligence: theory, research, and testing in the Nordic countries. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.49-78). **Cambridge: Cambridge University Press**.
- Carnap, R. (1983a). The logicist foundations of mathematics. In P.Benacerraf & H. Putnam (Eds.). (1983). **Philosophy of mathematics: selected readings** (pp.41-52). Cambridge: Cambridge University Press.



- Carnap, R. (1983b). Empiricism, semantics, and ontology. In P. Benacerraf & H. Putnam (Eds.), **Philosophy of mathematics: selected readings** (pp.241-257). Cambridge: Cambridge University Press.
- Carnap, R. (1993). Theories and nonobservables. In J.H. Fetzer (Ed.), **Foundations of philosophy of science: recent developments** (pp.96-109). New York: Paragon.
- Carney, T.F. (1972). **Content analysis: a technique for systematic inference from communications**. Winnipeg: University of Manitoba press.
- Carroll, J.B. (1994). Cognitive abilities: constructing a theory from data. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4. Theories of human intelligence** (pp.43-63). New Jersey: Ablex
- Carroll, J.B. (1995). Editorial - Reflections on Stephen Jay Gould's *The Mismeasure of Man* (1981): a retrospective review. **Intelligence**, **21**, 121-134.
- Carroll, J.B. (1996). A three-stratum theory of intelligence: Spearman's contribution. In I. Dennis & P. Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.1-17). New Jersey: Lawrence Erlbaum.
- Carroll, J.B. (1997a). Theoretical and technical issues in identifying a factor of general intelligence. In B. Devlin, S.E. Fienberg, D.P. Resnick & K. Roeder (Eds.), **Intelligence, genes, and success: scientists respond to *The Bell Curve*** (pp.125-156).
- Carroll, J.B. (1997b). The three-stratum theory of cognitive abilities. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.122-130). New York: Guilford.
- Carroll, J.B. (1998). What is intelligence? In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.51-54). New Jersey: Ablex.
- Case, K.E. & Fair, R.C. (1996a). **Principles of microeconomics**. New Jersey: Prentice-Hall.
- Case, K.E. & Fair, R.C. (1996b). **Principles of macroeconomics**. New Jersey: Prentice-Hall.
- Case, R. (1992). The role of the frontal lobes in the regulation of cognitive development. **Brain and cognition**, **20**, 51-73.
- Case, R. & Edelman, W. (1993). Introduction: structural approaches to individual differences. In R. Case & W. Edelman (Eds.), **The new structuralism in cognitive development: theory and research on individual pathways** (pp.1-10). Basel: Karger.
- Case, R. & Edelman, W. (Eds.), (1993). **The new structuralism in cognitive development: theory and research on individual pathways**. Basel: Karger.
- Case, R., Okamoto, Y., Henderson, B. & McKeough, A. (1993). Individual variability and consistency in cognitive development: new evidence for the existence of central conceptual structures. In R. Case & W. Edelman (Eds.), **The new structuralism in cognitive development: theory and research on individual pathways** (pp.71-100). Basel: Karger
- Castelijns, J., Van Werkhoven, W. & Stevens, L.M. (2002). The implementation of responsive teaching in kindergarten. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.293-306). Amsterdam: JAI.
- Casti, J.L. (2000). **Five more golden rules: knots, codes, chaos, and other great theories of 20th-century mathematics**. New York: John Wiley.
- Casti, J.L. (2001). **Paradigms regained: a further exploration of the mysteries of modern science**. New York: Perennial.
- Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. **Journal of educational psychology**, **54**(1), 1-22.
- Cattell, R.B. (1988a). Psychological theory and scientific method. In J.R. Nesselroade & R.B. Cattell (Eds.), **Handbook of multivariate experimental psychology** (pp.3-20). New York: Plenum.
- Cattell, R.B. (1988b). The principles of experimental design and analysis in relation to theory building. In J.R. Nesselroade & R.B. Cattell (Eds.), **Handbook of multivariate experimental psychology** (pp.21-67). New York: Plenum.
- Cattell, R.B. (1998). Where is intelligence? Some answers from the triadic theory. In J.J. McArdle & R.W. Woodcock (Eds.), **Human cognitive abilities in theory and practice** (pp.29-38). New Jersey: Lawrence Erlbaum.
- Ceci, S.J. & Bruck, M. (1994). The bio-ecological theory of intelligence: a developmental-contextual perspective. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4. Theories of human intelligence** (pp.65-84). New Jersey: Ablex
- Ceci, S.J., Rosenblum, T., De Bruyn, E. & Lee, D.Y. (1997). A bio-ecological model of intellectual development: moving beyond h^2 . In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.303-322). Cambridge: Cambridge University Press.
- Ceci, S.J., Rosenblum, T.B. & Kumpf, M. (1998). The shrinking gap between high- and low-scoring groups: current trends and possible causes. In U. Neisser (Ed.), **The rising curve: long-term gains in IQ and related measures** (pp.287-302). Washington: American Psychological Association.
- Cernovsky, Z.Z. (1997). A critical look at intelligence research. In D. Fox & I. Prilleltensky (Eds.), **Critical psychology: an introduction** (pp.121-133). London: Sage.
- Cervone, D. (2000). Theory, metatheory, and metaphor in the study of behavioral self-regulation. **Contemporary psychology**, **45**(1), 118-121.
- Chaitin, G. (2000). A Century of Controversy over the foundations of mathematics. In C. S. Calude & G. Paun (Eds.), **Finite versus Infinite: contributions to an eternal dilemma** (pp.75-100). London: Springer-Verlag.
- Chaitin, G. (2006). The limits of reason. **Scientific American**, **294**(3), 74-81.



- Chalmers, A.F. (1976). **What is this thing called science? An assessment of the nature and status of science and its methods 1st Ed.** Queensland: University of Queensland Press.
- Chalmers, A.F (1987). **What is this thing called science? An assessment of the nature and status of science and its methods 2nd Ed.** Queensland: University of Queensland Press.
- Chalmers, A.F (1999). **What is this thing called science? 3rd Ed.** Queensland: University of Queensland Press.
- Chalmers, D.J. (2003). Consciousness and its place in nature. In S.P.Stich & T.A.Warfield (Eds.), **The Blackwell guide to philosophy of mind** (pp.102-142). Massachusetts: Blackwell.
- Chalmers, M. & McGonigle, B. (2000). Capturing dynamic structuralism in the laboratory. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.183-200). London: Routledge.
- Chambers, E.A. (2004). An introduction to meta-analysis with articles from the journal of educational research (1992-2002). **The journal of educational research**, **98**(1), 35-45.
- Chan, W.Y, Ashman, A.F. & Van Kraayenoord, C.E. (2000). Science and biological classification: linking dynamic assessment to classroom curriculum. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.607-639). New York: Elsevier.
- Changeux, J-P. & Konishi, M. (Eds.), (1987). **The neural and molecular bases of learning**. Chichester: John Wiley.
- Changeux, J-P. & Konishi, M. (1987). Introduction. In J-P.Changeux & M. Konishi (Eds.), **The neural and molecular bases of learning** (pp.5-12). Chichester: John Wiley.
- Changeux, J-P. & Dehaene, S. (1994). Neuronal models of cognitive functions. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.363-402). Oxford: Blackwell.
- Charles, D.C. (1973). Explaining intelligence in adulthood: the role of life history. **Gerontologist**, **13**(4), 483-489.
- Chartier, D. (1996). An evaluation of two cognitive learning methods in adults on pre-qualification schemes: Logo and logical reasoning workshops (ARL). **European journal of psychology of education**, **11**(4), 443-457.
- Chater, N. & Vitányi, P.(2003). Simplicity: a unifying principle in cognitive science. **Trends in cognitive sciences**, **7**(1), 19-22.
- Chen, J-Q. & Gardner, H. (1997). Alternative assessment from a multiple intelligences theoretical perspective. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.105-121). New York: Guilford.
- Cherny, S.S., Fulker, D.W. & Hewitt, J.K. (1997). Cognitive development from infancy to middle childhood. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.463-482). Cambridge: Cambridge University Press.
- Chiew, V. & Wang, Y. (2003). From cognitive psychology to cognitive informatics. **Proceedings of the Second IEEE international conference on cognitive informatics (ICCI '03)**, 114-120.
- Chin, W.W. (1998). The partial least squares approach to structural equation modelling. In G.A. Marcoulides (Ed.), **Modern methods for business research** (pp.295-336). New Jersey: Lawrence Erlbaum.
- Chipman, S.F. (1986). Integrating three perspectives on learning. In S.L. Friedman, K.A. Klivington & R.W. Peterson (Eds.), **The brain, cognition, and education** (pp.203-229). Orlando: Academic Press.
- Chipman, S.F., Nichols, P.D. & Brennan, R.L. (1995). Introduction. In P.D. Nichols, S.F. Chipman & R.L. Brennan (Eds.), **Cognitively diagnostic assessment** (pp.1-18). New Jersey: Lawrence Erlbaum.
- Chipuer, H.M., Rovine, M.J. & Plomin, R. (1990). LISREL modelling: genetic and environmental influences on IQ revisited. **Intelligence**, **14**(1), 11-29.
- Chomsky, N. (1959). Verbal behavior (Book review). **Language**, **35**, 26-58.
- Chow, S.L. (1991). Some reservations about power analysis. **American psychologist**, **46**, 1088-1089.
- Chow, S.L. (1998a). The null-hypothesis significance-test procedure is still warranted. **Behavioral and brain sciences**, **21**(2), 228-235.
- Chow, S.L. (1998b). What statistical significance means. **Theory and psychology**, **8**(3), 323-330.
- Chow, S.L. (2002). Issues in statistical inference. **History and philosophy of psychology bulletin**, **14**(1), 30-41.
- Chown, M. (2006). The longest stake-out. **New scientist**, **189**(2543), 32-35.
- Chugani, H.T., Phelps, M.E. & Mazziotta, J.C. (1994). Positron emission tomography study of human brain functional development. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.125-143). Oxford: Blackwell.
- Church, A.T. & Katigbak, M.S. (1987). Ecocultural bias in culture-specific intelligence tests in an ecologically diverse culture: the Philippines. **Intelligence**, **11**(4), 371-389.
- Churchland, P.(2000). Matter and consciousness. In J. McErlean (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.485-493). California: Wadsworth.
- Churchland, P.M. & Hooker, C.A. (Eds.). (1985). **Images of science: essays on realism and empiricism, with a reply from Bas C. Van Fraassen**. Chicago: The university of Chicago Press.
- Churchland, P.S. (1986). **Neurophilosophy: toward a unified science of the mind-brain**. Cambridge: MIT Press.
- Churchland, P.S. & Sejnowski, T.J. (1996). **The computational brain**. Cambridge: MIT Press.
- Cicchetti, D. & Grove, W.M. (1991). (Eds.), **Thinking clearly about psychology volume 1: matters of public interest**. Minneapolis: University of Minnesota.
- Cilliers, F.P. (1990). The brain, the mental apparatus and the text: a post-structural neuropsychology. **South African journal of philosophy**, **9**(1), 1-8.

- Clapham, C. (1996). **Oxford concise dictionary of mathematics**. Oxford: Oxford University Press.
- Cliff, N. (1991a). Comments on "implications of a multidimensional latent trait model for measuring change". In L.M. Collins & J.L. Horn (Eds.), **Best methods for the analysis of change: recent advances, unanswered questions, future directions** (pp.198-201). Washington: American psychological association.
- Cliff, N. (1991b). Ordinal methods in the assessment of change. In L.M. Collins & J.L. Horn (Eds.), **Best methods for the analysis of change: recent advances, unanswered questions, future directions** (pp.34-46). Washington: American psychological association.
- Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. **Psychological science**, **3**(3), 186-190.
- Close, F., Marten, M. & Sutton, C. (2002). **The particle odyssey: a journey to the heart of matter**. Oxford: Oxford University press.
- Cobley, P. & Jansz, L. (1999). **Introducing semiotics**. Cambridge: Icon.
- Cockerham, W.C. (1998). **Medical sociology**. New Jersey: Prentice Hall.
- Coertze, P.J. (1977). Inleiding tot die algemene volkekunde. Johannesburg: Voortrekkerpers.
- Cohen, I.B. (2000). The notion of a balance. In J. McErlean (Ed.), **Philosophies of science: from foundations to contemporary issues** (pp.286-297). California: Wadsworth.
- Cohen, J. (1990). Things I have learned (so far). **American psychologist**, **45**(12), 1304-1312.
- Cohen, J. (1994). The earth is round ($p < .05$). **American psychologist**, **49**(12), 997-1003.
- Cohen, J.E. (2005). Human population grows up. **Scientific American**, **293**(3), 48-55.
- Cohen, M.R. & Nagel, E. (1949). **An introduction to logic and scientific method**. London: Routledge and Kegan Paul.
- Cole, E. (1996a). Multicultural psychological assessment: new challenges, improved methods. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.88-95). Ontario: Captus Press.
- Cole, E. (1996b). Immigrant and refugee children and families; supporting a new road travelled. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.35-42). Ontario: Captus Press.
- Cole, M. & Cole, S. (Eds.), (1979). **The making of mind: a personal account of Soviet psychology**. Cambridge: Harvard University Press.
- Cole, M., John-Steiner, V. Scribner, S. & Souberman, E. (Eds.), (1978). Vygotsky's "**Mind in society: the development of higher psychological processes**". Cambridge: Harvard University Press.
- Cole, M. & Scribner, S. (1978). Introduction. In M. Cole, V. John-Steiner, S. Scribner & E. Souberman (Eds.), Vygotsky's "**Mind in society: the development of higher psychological processes**" (pp.1-14). Cambridge: Harvard University Press.
- Cole, M. & Valsiner, J. (2005). Actualizing potentials: learning through psychology's recurrent crises. In D.B. Pillemer & S.H. White (Ed.), **Developmental psychology and social change: research, history, and policy** (pp.288-313). Cambridge: Cambridge University Press.
- Coles, G.S. (1982). Learning disabilities theory and Soviet Psychology: a comparison of basic assumptions. **Journal of clinical neuropsychology**, **4**(3), 269-283.
- Collins, A. (1996). Design issues for learning environments. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), **International perspectives on the design of technology-supported learning environments** (pp.347-361). New Jersey: Lawrence Erlbaum.
- Collins, L.M. (1991). Measurement in longitudinal research. In L.M. Collins & J.L. Horn (Eds.), **Best methods for the analysis of change: recent advances, unanswered questions, future directions** (pp.137-148). Washington: American psychological association.
- Collins, L.M. & Horn, J.L. (Eds.), (1991). **Best methods for the analysis of change: recent advances, unanswered questions, future directions**. Washington: American psychological association.
- Colom, R., Lluís-Font, J.M. & Andrés-Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. **Intelligence**, **33**(1), 83-91.
- Colvin, S.S. (1921). Intelligence and its measurement. **Journal of educational psychology**, **12**(3), 136-139.
- Converse, P.E. (1986). The social psychology of "Other Worlds". In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.42-60). Chicago: University of Chicago Press.
- Conway, A.R.A. (2005). Cognitive mechanisms underlying intelligence: defense of a reductionist approach. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.47-59). London: Sage
- Coombs, C.H. (1967). **A theory of data**. New York: John Wiley.
- Coombs, C.H., Dawes, R.M. & Tversky, A. (1970). **Mathematical psychology: an elementary introduction**. New Jersey: Prentice-Hall.
- Coon, H., Carey, G. & Fulker, D.W. (1992). Community influences on cognitive ability. **Intelligence**, **16**(2), 169-188.
- Cooper, H. & Hedges, L.V. (Eds.), (1994). **The handbook of research synthesis**. New York: Russell Sage Foundation.
- Cooper, H. & Hedges, L.V. (1994a). Research synthesis as a scientific enterprise. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.3-14). New York: Russell Sage Foundation.
- Cooper, H. & Hedges, L.V. (1994b). Potentials and limitations of research synthesis. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.521-529). New York: Russell Sage Foundation.
- Cooper, H.M. & Lemke, K.M. (1991). On the role of meta-analysis in personality and social psychology. **Personality and social psychology bulletin**, **17**(3), 245-251.

- Cooper, R. & Shallice, T. (1995). Soar and the case for unified theories of cognition. **Cognition**, **55**, 115-149.
- Copeland, J. (2002). The Church-Turing thesis. In J. Perry & E. Zalta (Eds.), **The Stanford encyclopedia of philosophy** (Retrieved 16 May 2006 from <http://Plato.stanford.edu/entries/church-turing/>).
- Copi, I.M. (1972). **Introduction to logic**. New York: Macmillan.
- Corballis, M.C. (1985). On being brainy. In K.B. Madsen & L.P. Mos (Eds.), **Annals of theoretical psychology. Volume 3** (pp.135-142). New York: Plenum Press.
- Cordova, D.I & Lepper, M.R. (1996). Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. **Journal of educational psychology**, **88**(4), 715-730.
- Cornwell, J. (2004). **Hitler's scientists: science, war and the Devil's pact**. London: Penguin.
- Cortada De Kohan, N. (1972). Test construction and standardization in different cultural settings. In L.J. Cronbach & P.J.D. Drenth (Eds.), **Mental tests and cultural adaptation** (pp.121-127). The Hague: Mouton.
- Cossons, N., Nahum, A. & Turvey, P.(1997). **Making of the modern world: milestones of science and technology**. London: John Murray in association with the Science Museum.
- Coulter, J. (1997). Neural Cartesianism: comments of the epistemology of the cognitive sciences. In D.M. Johnson, & C.E. Erneling (Eds.), **The future of the cognitive revolution** (pp.293-301). New York: Oxford.
- Coveney, P.& Highfield, R. (1995). **Frontiers of complexity: the search for order in a chaotic world**. New York: Fawcett Columbine.
- Cowan, N. (2005). Understanding intelligence: a summary and an adjustable-attention hypothesis. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.469-488). London: Sage.
- Coyle, T.R. (2006). Test-retest changes on scholastic aptitude tests are not related to *g*. **Intelligence**, **34**, 15-27.
- Craig, A. (1990). The study of change. In H. Van Niekerk (Ed.), **Cognitive development in the Southern African context: papers of the seminar on cognitive development, 1 November 1990** (pp.62-74). Pretoria: Human Sciences Research Council.
- Craig, A. (1991). Adult cognition and tertiary studies. **South African journal of higher education**, **5**(2), 137-144.
- Craig, A. (2000). Time and learning. **South African journal of psychology**, **30**(4), 6-16.
- Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: a framework for memory research. **Journal of verbal learning and verbal behavior**, (11), 671-684.
- Craver, C.F. (2002). Structure of scientific theories. In P.Machamer & M. Silberstein (Eds.), **The Blackwell guide to the philosophy of science** (pp.55-79). Massachusetts: Blackwell.
- Crawford, S.A.S. & Das, J.P. (1992). Teaching for transfer: a program for remediation in reading. In J.S. Carlson (Ed.), **Advances in cognition and educational practice: a research annual volume 1B** (pp.73-103). Connecticut: JAI Press.
- Creswell, J.W. (1998). **Qualitative inquiry and research design: choosing among five traditions**. London: Sage.
- Creswell, J.W. (2002). **Educational research: planning, conducting, and evaluating quantitative and qualitative research**. New Jersey: Merrill Prentice Hall.
- Crocker, L. & Algina, J. (1986). **Introduction to classical and modern test theory**. New York: Holt, Rinehart and Winston.
- Cronbach, L.J. (1986). Social inquiry by and for Earthlings. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.83-107). Chicago: University of Chicago Press
- Cronbach, L.J. (1990). **Essentials of psychological testing 5th Ed**. New York: Harper Collins.
- Cronbach, L.J. & Drenth, P.J.D. (Eds.), (1972). **Mental tests and cultural adaptation**. The Hague: Mouton.
- Cronbach, L.J. & Furby, L. (1970). How we should measure "change"- or should we? **Psychological bulletin**, **74**(1), 68-80.
- Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. **Psychological bulletin**, **52**, 281-302.
- Crosson, F.J. (1985). Psyche and the computer. In S. Koch & D.E. Leary (Eds.). **A century of psychology as science** (pp.437-451). New York: McGraw-Hill.
- Curry, H.B. (1983). Remarks on the definition and nature of mathematics. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.202-206). Cambridge: Cambridge University Press.
- D'Abro, A. (1953). **The rise of the new physics: its mathematical and physical theories**. Vol 1. New York: Dover.
- D'Agostino, S. (2000). **A history of the ideas of theoretical physics. Essays on the nineteenth and twentieth century physics**. Dordrecht: Kluwer Academic.
- Dalmedico, A.D. (1997). Mathematics in the twentieth century. In J. Krige & D. Pestre (Ed.). **Science in the twentieth century** (pp.651-667). Amsterdam: Harwood Academic.
- Damasio, A. (2003). **Looking for Spinoza: joy, sorrow and the feeling brain**. Orlando: Harcourt.
- Damasio, A. (2006). **Descartes'error: emotion, reason and the human brain**. London: Vintage.
- D'Andrade, R. (1986). Three scientific world views and the covering law model. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.19-41). Chicago: University of Chicago Press.
- Daniel, M.H. (1997). Intelligence testing: status and trends. **American psychologist**, **52**(10), 1038-1045).
- Daniel, M.H. (1999). Behind the scenes: using new measurement methods on the DAS and KAIT. In S.E. Embretson & S.L. Hershberger (Eds.), **The new rules of measurement: what every psychologist and educator should know** (pp. 37-63). New Jersey: Lawrence Erlbaum.
- Daniels, H. (Ed). (1996). **An introduction to Vygotsky**. London: Routledge.
- Dann, P.L., Irvine, S.H & Collis, J.M. (Eds.), (1991). **Advances in computer-based human assessment**. Dordrecht: Kluwer.

- Danthir, V., Roberts, R.D., Schulze, R. & Wilhelm, O. (2005). Mental speed: in frameworks, paradigms, and a platform for the future. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.27-46). London: Sage.
- Danziger, K. (1990). **Constructing the subject: historical origins of psychological research**. New York: Cambridge University Press.
- Darwin, C. (no date). *The origin of species by means of natural selection or the preservation of favoured races in the struggle for life*. New York: The Modern Library. (Originally published in 1859)
- Das, J.P. (1987). Forward. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.vii-xi). New York: The Guilford Press.
- Das, J.P. (1998). On definition of intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.55-56). New Jersey: Ablex.
- Das, J.P. & Conway, N.F. (1992). Reflections on remediation and transfer: A Vygotskian perspective. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.94-115). New York: Springer-Verlag.
- Das, J.P. & Naglieri, K. (1992). Assessment of attention, simultaneous-successive coding, and planning. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.207-232). New York: Springer-Verlag.
- Das, J.P., Naglieri, J.A. & Kirby, J.R. (1994). **Assessment of cognitive processes: the PASS theory of intelligence**. Boston: Allyn & Bacon.
- Das, J.P., Parilla, R.K. & Papadopoulos, T.C. (2000). Cognitive education and reading disability. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.274-291). Amsterdam: Pergamon.
- Daugman, J.G. (2001). Brain metaphor and brain theory. In W. Bechtel, P.Mandik, J. Mundale & R.S. Stufflebeam, (Eds.). **Philosophy and the neurosciences: a reader** (pp.23-36). Massachusetts: Blackwell.
- David, D., Miclea, M. & Opre, A. (2004). The information-processing approach to the human mind: basics and beyond. **Journal of clinical psychology, 60**(4), 353 - 368.
- Davidson, J.E. & Downing, C.L. (2004). Contemporary models of intelligence. In R.J. Sternberg (Ed.). **Handbook of intelligence** (pp.34-49). Cambridge: Cambridge University Press.
- Davidson, J.E. & Sternberg, R.J. (1998). Smart problem solving: how metacognition helps. In D.K. Hacker, J. Dunlosky & A.C. Graesser (Eds.), **Metacognition in educational theory and practice** (pp.47-68). New Jersey: Lawrence Erlbaum.
- Davis, H. & Anderson, M. (1999). Individual differences and development - one dimension or two? In M. Anderson (Ed.), **The development of intelligence** (pp.161-191). Sussex: Psychology Press.
- Dawes, R.M. (1971). A case study of graduate admissions: applications of three principles of human decision making. **American psychologist, 26**, 180-188.
- Dawes, R.M. (1989). Experience and validity of clinical judgment: the illusory correlation. **Behavioral sciences and the law, 7**, 457-467.
- Dawes, R.M. (1995). The nature of human nature: an empirical case for withholding judgment - perhaps indefinitely. **Political psychology, 16**, 81-97.
- Dawkins, R. (1999a). **Unweaving the rainbow: science, delusion and the appetite for wonder**. London: Penguin.
- Dawkins, R. (1999b). **The extended phenotype: the long reach of the gene**. Oxford: Oxford University Press.
- Dawkins, R. (2001). **River out of Eden: a Darwinian view of life**. London: Phoenix.
- Dawkins, R. (2004). **A devil's chaplain: selected essays**. London: Phoenix.
- Dawkins, R. (2005). **The ancestor's tale. A pilgrimage to the dawn of life**. London: Phoenix.
- Dawkins, R. (2006). **The god delusion**. London: Bantam Press.
- Dawson, M.R.W. (1995). The problems and prospects of comparative and noncomparative theoretical psychology: a response to Kukla. **New ideas in psychology, 13**(3), 219-222.
- Day, J.D. & Cordón, L.A. (1993). Static and dynamic measures of ability: an experimental comparison. **Journal of educational psychology, 85**(1), 75-82.
- Day, J.D., Engelhardt, J.L., Maxwell, S.E. & Bolig, E.E. (1997). Comparison of static and dynamic assessment procedures and their relation to independent performance. **Journal of educational psychology, 89**(2), 358-368.
- Deacon, T. (1997). **The symbolic species: the co-evolution of language and the human brain**. London: Penguin.
- Deary, I.J. (2001). Human intelligence differences: a recent history. **Trends in cognitive sciences, 5**(3), 127 - 130.
- Deary, I.J. (in press). Book review: Intelligence, destiny and education: the ideological roots of intelligence testing by J. White. *Intelligence*.
- Deary, I.J. & Smith, P. (2004). Intelligence research and assessment in the United Kingdom. In R.J. Sternberg (Ed). **International handbook of intelligence** (pp.1-48). Cambridge: Cambridge University Press.
- De Beer, M. (2000). **The construction and evaluation of a dynamic computerised adaptive test for the measurement of learning potential**. Unpublished D.Phil. thesis, University of South Africa.
- De Beer, N.A. (1991). Die toepassingspotensiaal Van die Feuerstein instrumentele verryingsprogram (FIE) vir leergestremde kinders binne die multikulturele Suid-Afrikaanse konteks. **Tydskrif vir geesteswetenskappe, 31**(3), 189-198.
- De Corte, E., Lodewijks, H., Parmentier, R. & Span, P. (Eds.), (1987). **Learning and instruction: European research in an international context volume 1**. Oxford: Pergamon.

- De Corte, E. (1996). Changing views of computer-supported learning environments for the acquisition of knowledge and thinking skills. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), **International perspectives on the design of technology-supported learning environments** (pp.129-145). New Jersey: Lawrence Erlbaum.
- DeFrance, J.F., Hymel, C., Degioanni, J., Calkin, D.S., Estes, S., Schweitzer, F.C. & Hymel, R. (1994). Assessment of cognitive function: exploration of memory processing to topographical mapping techniques. In C.R. Reynolds (Ed). **Cognitive assessment: a multidisciplinary perspective** (pp.57-71). New York: Plenum Press.
- De Grolier, E. (1990). From theories to concepts and from facts to words. **International social science journal**, **42**(2), 241-251.
- Dehaene-Lambertz, G. & Dehaene, S. (1997). In defense of learning by selection: neurobiological and behavioral evidence revisited. **Behavioral and brain sciences**, **20**(4), 560-561.
- De Jong, H.B., Bem, S. & Schouten, M. (2004). Theory in psychology: a review essay of Andre Kukla's *Methods of theoretical psychology*. **Philosophical psychology**, **17**(2), 275-295.
- De Leeuw, J. (1994). Statistics and the sciences. In I. Borg & P.Mohler (Eds.), **Trends and perspectives in empirical social research** (pp.139-148). Berlin: Walter De Gruyter.
- Delclos, V.R., Burns, M.S. & Kulewicz, S.J. (1997). Effects of dynamic assessment on teachers' expectations of handicapped children. **American educational research journal**, **24**, 325-336.
- Delclos, V.R., Vye, N.J., Burns, S.M., Bransford, J.D. & Hasselbring, T.S. (1992). Improving the quality of instruction: roles for dynamic assessment. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.317-331). New York: Springer-Verlag.
- Delius, C., Gatzemeier, M., Sertcan, D. & Wünsch, K. (2000). **The story of philosophy from antiquity to the present**. Cologne: Könemann.
- Demetriou, A. (2006). Dissecting *g*. **Behavioral and brain sciences**, **29**, 130-132.
- Demetriou, A. & Papadopoulos, T.C. (2004). Human intelligence: from local models to universal theory. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.445-474). **Cambridge: Cambridge University Press**.
- Demetriou, A. & Valanides, N. (1998). A three-level theory of the developing mind: basic principles and implications for instruction and assessment. In R.J. Sternberg & W.M. Williams (Eds.), **Intelligence, instruction, and assessment: theory into practice** (pp.149-199). New Jersey: Lawrence Erlbaum.
- De Mey, M. (1989). Cognitive paradigms and psychology of science. In B. Gholsen, W.R. Shadish, R.A. Neimeyer, & A.C Houts, (Eds.). **Psychology of science: contributions to metascience** (pp.275-295). Cambridge: Cambridge University Press.
- Denishoff, R.S., Callahan, O. & Levine, M.H. (Eds.). (1974). **Theories and paradigms in contemporary sociology**. Illinois: F.E. Peacock.
- Dennett, D.C. (1981). Reflections on "An epistemological nightmare". In D.R. Hofstadter & D.C. Dennett (Eds.). **The mind's I: fantasies and reflections on self and soul** (pp.415-426). London: Penguin.
- Dennett, D.C. (1993). **Consciousness explained**. London: Penguin.
- Dennett, D.C. (1995). **Darwin's dangerous idea: evolution and the meanings of life**. London: Penguin.
- Dennett, D.C. (1999). Afterword. In R. Dawkins. **The extended phenotype: the long reach of the gene** (pp265-268). Oxford: Oxford University Press.
- Dennett, D.C. (2004). **Freedom evolves**. London: Penguin.
- Dennis, I. & Tapsfield, P. (Eds.), (1996). **Human abilities: their nature and measurement**. New Jersey: Lawrence Erlbaum.
- Dennis, I. & Tapsfield, P. (1996). Introduction: the abilities of man revisited. In I.Dennis & P.Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.xi-xv). New Jersey: Lawrence Erlbaum.
- Dennis, M.A. (1997). Historiography of science. In J. Krige & D. Pestre (Ed.). **Science in the twentieth century** (pp.7-26). Amsterdam: Harwood Academic.
- Denzin, N.K. & Lincoln, Y.S. (Eds.), (2000). **Handbook of qualitative research**. London: Sage.
- Department of Education (2003). **Education statistics in South Africa at a glance in 2001**. Pretoria: EMIS Directorate, Department of Education.
- De Ribaupierre, A. (1993). Structural invariants and individual differences: on the difficulty of dissociating developmental and differential processes. In R. Case & W. Edelstein (Eds.), **The new structuralism in cognitive development: theory and research on individual pathways** (pp.11-32). Basel: Karger
- Derr, R.L. (1989). Insights on the nature of intelligence from ordinary discourse. **Intelligence**, **13**(2), 113-118.
- Desoete, A., Roeyers, H., Buysse, A. & De Clercq, A. (2002). Dynamic assessment of metacognitive skills in young children with mathematics-learning disabilities. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.307-333). Amsterdam: JAI.
- Detterman, D.K. (1979). A job half done: the road to intelligence testing in the year 2000. **Intelligence**, **3**, 295-306.
- Detterman, D.K. (1982). Does "g" exist? **Intelligence**, **6**, 99-108.
- Detterman, D.K. (1991). Reply to Deary and Pagliari: is *g* intelligence or stupidity? **Intelligence**, **15**(3), 251-255.
- Detterman, D.K. (Ed.) (1994a). **Current topics in human intelligence volume 4. Theories of human intelligence**. New Jersey: Ablex.

- Detterman, D.K. (1994b). Theoretical possibilities: the relation of human intelligence to basic cognitive abilities. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4. Theories of human intelligence** (pp.85-115). New Jersey: Ablex.
- Detterman, D.K. (1998a). Human intelligence is a complex system of separate processes. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.57-61). New Jersey: Ablex.
- Detterman, D.K. (1998b). Qualitative integration: the last word? In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.163-166). New Jersey: Ablex.
- Detterman, D.K. (2006). Editorial note on controversial papers. **Intelligence**, **34**(2), iv.
- Deutsch, R. M. (2003). Mediation from the perspective of the theory of mediated learning experience. **Journal of cognitive education and psychology**, **3**(1), 29-45.
- Deutsch, R. M. & Reynolds, Y. (2000). The use of dynamic assessment by educational psychologists in the UK. **Educational psychology in practice**, **16**(3), 311-331.
- De Villiers, A.B. (1999). **Disadvantaged students' academic performance: analysing the zone of proximal development**. Unpublished D.Phil. thesis, University of Cape Town.
- DeVries, R. (2000). Vygotsky, Piaget, and education: a reciprocal assimilation of theories and educational practices. **New ideas in psychology**, **18**, 187-213.
- Devlin, B., Fienberg, S.E., Resnick, D.P. & Roeder, K. (Eds.) (1997). **Intelligence, genes and success: scientists respond to *The Bell Curve***. New York: Springer-Verlag.
- De Weerdt, E.H. (1927). A study of the improvability of fifth grade school children in certain mental functions. **Journal of educational psychology**, **18**, 547-557.
- Diamond, A. (1994). Neuropsychological insights into the meaning of object concept development. In M.H. Johnson, (Ed) **Brain development and cognition: a reader** (pp 208-247). Oxford: Blackwell.
- Diamond, J. (2006). **Collapse: how societies choose to fail or survive**. London: Penguin.
- DiBello, L.V., Stout, W.F. & Roussos, L.A. (1995). Unified cognitive/psychometric diagnostic likelihood-based classification techniques. In P.D. Nichols, S.F. Chipman & R.L. Brennan (Eds.), **Cognitively diagnostic assessment** (pp.361-389). New Jersey: Lawrence Erlbaum.
- Dickersin, K. (1994). Research registers. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.71-83). New York: Russell Sage Foundation.
- DiLalla, L.F. (2000). Development of intelligence: current research and theories. **Journal of school psychology**, **38**(1), 3-7.
- Dillenbourg, P. (1996). Distributing cognition over humans and machines. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), **International perspectives on the design of technology-supported learning environments** (pp.165-183). New Jersey: Lawrence Erlbaum.
- Dillon, R.F. (1997). A new era in testing. In R.F. Dillon (Ed.), **Handbook on testing** (pp.1-19). Connecticut: Greenwood Press.
- Dillon, R.F. (Ed.), (1997). **Handbook on testing**. Connecticut: Greenwood Press.
- Dilthey, W. (2000). Introduction to the human sciences. In J. McErlean (Ed.), **Philosophies of science: from foundations to contemporary issues** (pp.309-322). California: Wadsworth.
- Doehring, D.G. (1982). Learning disabilities theory and Soviet psychology: a comparison of straw men and red herrings. **Journal of clinical neuropsychology**, **4**(3), 285-289.
- Domoter, Z. (1992). Measurement from empiricist and realist points of view. In C.W. Savage & P.Ehrlich (Eds.), **Philosophical and foundational issues in measurement theory** (pp.195-221). New Jersey: Lawrence Erlbaum.
- Donald, M. (1997). The mind considered from a historical perspective: human cognitive phylogensis and the possibility of continuing cognitive evolution. In D.M. Johnson, & C.E Erneling (Eds.), **The future of the cognitive revolution** (pp.356 - 365). New York: Oxford.
- Donnelly, R.A. (2004). **The complete idiot's guide to statistics**. New York: Alpha.
- Doré, F.Y. (1991). A neo-Piagetian theory can contribute to comparative studies of cognitive development. **Behavioral and brain sciences**, **14**(2), 368-372.
- Dorfman, W.I. & Hersen, M. (Eds.), (2001). **Understanding psychological assessment**. New York: Kluwer.
- Doyal, L. & Harris, R. (1986). **Empiricism, explanation and rationality. An introduction to the philosophy of science of the social sciences**. London: Routledge & Kegan Paul.
- Draaisma, D. (2000). **Metaphors of memory: a history of ideas about the mind**. Cambridge: Cambridge University Press.
- Draney, K.L., Pirolli, P. & Wilson, M. (1995). A measurement model for a complex cognitive skill. In P.D. Nichols, S.F. Chipman & R.L. Brennan (Eds.), **Cognitively diagnostic assessment** (pp.103-125). New Jersey: Lawrence Erlbaum.
- Drasgow, F. & Olson-Buchanan, J.B. (1999). Preface. In F. Drasgow & J.B. Olson-Buchanan (Eds.), **Innovations in computerized assessment** (pp.ix-xiv). New Jersey: Lawrence Erlbaum.
- Drasgow, F. & Olson-Buchanan, J.B. (Eds.), (1999). **Innovations in computerized assessment**. New Jersey: Lawrence Erlbaum.
- Du Bois, P.H. (1970). **A history of psychological testing**. Boston: Allyn and Bacon.
- Dummet, M. (1983). The philosophical basis of intuitionist logic. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.97-129). Cambridge: Cambridge University Press.

- Dumont, K. & Frindte, W. (2005). Content analysis of the homepages of academic psychologists. **Computers in human behavior, 21**, 73-83
- DuMouchel, W.H. (1992). Introduction to Edwards, Lindman, and Savage (1963) "Bayesian statistical inference for psychological research". In S. Kotz & N.L. Johnson (Eds.), **Breakthroughs in statistics volume 1: foundations and basic theory** (pp.519-530). New York: Springer-Verlag.
- Duncan, J. (2005). Frontal lobe function and general intelligence: why it matters. **Cortex, 41**(2), 215-217.
- Duncan, J., Seitz, R.J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F.N. & Emslie, H. (2000). A neural basis for general intelligence. **Science, 289**, 457-460.
- Dunlosky, J. & Hertzog, C. (1998). Training programs to improve learning in later adulthood: helping older adults educate themselves. D.K. Hacker, J. Dunlosky & A.C. Graesser (Eds.), **Metacognition in educational theory and practice** (pp.249-275). New Jersey: Lawrence Erlbaum.
- Du Toit, P.J.S. (1990). Kognisie en metakognisie: belangrike strategieë in die leerhandelinge en in remediërende onderwys. **Journal of technical and vocational education, 126**, 23-26.
- Duveen, G. (2000). Psychological development as a social process. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.67-90). London: Routledge.
- Dwyer, C.A. & Dwyer, F.M. (1987). Effect of depth of information processing on students' ability to acquire and retrieve information related to different instructional objectives. **Programmed learning and educational technology, 24**(4), 264-280.
- d'Ydewalle, G. (1987). Is it still worthwhile to investigate the impact of motivation on learning? In E. De Corte, H. Lodewijks, R. Parmentier & P.Span (Eds.), **Learning and instruction: European research in an international context volume 1** (pp.191-200). Oxford: Pergamon
- d'Ydewalle, G. & Denis, M. (2000). Cognitive psychology and its successful spin-offs. **International journal of psychology, 35**(6), 298-299.
- Eagly, A.H & Wood, W. (1994). Using research synthesis to plan future research. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.485-500). New York: Russell Sage Foundation.
- Eccles, J. (1989). Brain and mind: two or one? In C. Blakemore & S. Greenfield (Eds.), **Mindwaves: thoughts on intelligence, identity and consciousness** (pp.293-304). Oxford: Blackwell.
- Eddington, A. (1939). **The philosophy of physical science, Tarner Lectures, 1938**. Cambridge: Cambridge University Press.
- Eddy, D.M. & Hasselblad, V. (2005). **FAST*PRO**. Retrieved April 13 2005 from <http://bmj.bmjournals.com/content/vol316/issue7126/fulltext/supplemental/221/index.shtml>
- Edwards, W., Lindman, H. & Savage, L.J. (1963). Bayesian statistical inference for psychological research. **Psychological review, 70**, 193-242.
- Einstein, A. & Infeld, L. (1961). **The evolution of physics**. Cambridge: Cambridge University Press.
- Elkonin, B.D. (2001). L.S Vygotsky and D.B. Elkonin: symbolic mediation and joint action. **Journal of Russian and East European psychology, 39**(4), 9-20.
- Elkonin, D., Foxcroft, C., Roodt, G. & Astbury, G. (2001). The use of assessment measures in various applied contexts. In C. Foxcroft & G. Roodt (Eds.), **An introduction to psychological assessment in the South African context** (pp.265-294). Cape Town: Oxford University Press.
- Elliott, J. & Lauchlan, F. (1997). Assessing potential - the search for the philosopher's stone? **Educational and child psychology, 14**(4), 6-16.
- Elliott, J., Lauchlan, F. & Stringer, P.(1996). Dynamic assessment and its potential for educational psychologists: part 1 - theory and practice. **Educational psychology in practice, 12**(3), 152-160.
- Elliott, J.G. (2000). Dynamic assessment in educational contexts: purpose and promise. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.713-740). New York: Elsevier.
- Elliott, J.G. (2001). All testing is dynamic: a response to Sternberg and Grigorenko. **Issues in Education, 7**(2), 185-191.
- Elliott, J.G. (2003). Dynamic assessment in educational settings: realising potential. **Educational review, 55**(1), 15-32.
- Elliott, R. (1988). Tests, abilities, race, and conflict. **Intelligence, 12**(4), 333-350.
- Ellis, B. (1966). **Basic concepts of measurement**. Cambridge: Cambridge University Press.
- Ellis, B. (1992). Conventionalism in measurement theory. In C.W. Savage & P. Ehrlich (Eds.), **Philosophical and foundational issues in measurement theory** (pp.167-180). New Jersey: Lawrence Erlbaum.
- Ellis, B.B. (1990). Assessing intelligence cross-nationally: a case for differential item functioning detection. **Intelligence, 4**(1), 61-78.
- Embretson, S.E. (1983). Construct validity: construct representation versus nomothetic span. **Psychological bulletin, 93**(1), 179-197.
- Embretson, S.E. (1987). Toward development of a psychometric approach. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.141-170). New York: The Guilford Press.
- Embretson, S.E. (1991a). A multidimensional latent trait model for measuring learning and change. **Psychometrika, 56**, 495-515.



- Embretson, S.E. (1991b). Implications of a multidimensional latent trait model for measuring change. In L.M. Collins & J.L. Horn (Eds.), **Best methods for the analysis of change: recent advances, unanswered questions, future directions** (pp.184-197). Washington: American psychological association.
- Embretson, S. E. (1992). Measuring and validating cognitive modifiability as an ability: a study in the spatial domain. **Journal of educational measurement**, **29**(1), 25-50.
- Embretson, S.E. (1994). Applications of cognitive design systems to test development. In C.R. Reynolds (Ed). **Cognitive assessment: a multidisciplinary perspective** (pp.107-135). New York: Plenum Press.
- Embretson, S.E. (1995). A measurement model for linking individual learning to processes and knowledge: application to mathematical reasoning. **Journal of educational measurement**, **32**(3), 277-294.
- Embretson, S.E. (1996). Multidimensional latent trait models in measuring fundamental aspects of intelligence. In I.Dennis & P.Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.117-132). New Jersey: Lawrence Erlbaum.
- Embretson, S.E. (1997a). Structured ability models in test designed from cognitive theory. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.223-236). London: Ablex.
- Embretson, S.E. (1997b). Measurement principles for the new generation of tests: a quiet revolution. In R.F. Dillon (Ed.), **Handbook on testing** (pp.20-38). Connecticut: Greenwood Press.
- Embretson, S.E. (1999). Issues in the measurement of cognitive abilities. In S.E. Embretson & S.L. Hershberger (Eds.), **The new rules of measurement: what every psychologist and educator should know** (pp.1-15). New Jersey: Lawrence Erlbaum.
- Embretson, S.E. (2000). Multidimensional measurement from dynamic tests: abstract reasoning under stress. **Multivariate behavioral research**, **35**(4), 505-542.
- Embretson, S.E. (2004). The second century of ability testing: some predictions and speculations. **Measurement**, **2**(1), 1-32.
- Embretson, S.E. (2005). Measuring human intelligence with artificial intelligence: adaptive item generation. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.251-267). Cambridge: Cambridge University Press.
- Embretson, S.E. & Hershberger, S.L. (Eds.), (1999). **The new rules of measurement: what every psychologist and educator should know**. New Jersey: Lawrence Erlbaum.
- Embretson, S.E. & Hershberger, S.L. (1999). Summary and future of psychometric methods in testing. In S.E. Embretson & S.L. Hershberger (Eds.), **The new rules of measurement: what every psychologist and educator should know** (pp.243-254). New Jersey: Lawrence Erlbaum.
- Embretson, S. E. & McCollam, K.M. (2004). Psychometric approaches to understanding and measuring intelligence. In R.J. Sternberg (Ed.), **Handbook of intelligence** (pp.423-444). Cambridge: Cambridge University Press.
- Embretson, S.E. & Prenovost, L.K. (2000). Dynamic cognitive testing: what kind of information is gained by measuring response time and modifiability? **Educational and psychological measurement**, **60**(6), 837-863.
- Embretson, S.E. & Reise, S.P. (2000). **Item response theory for psychologists**. New Jersey: Lawrence Erlbaum.
- Emihovich, C. & Lima, E.S. (1995). The many facets of Vygotsky: a cultural historical voice from the future. **Anthropology and Education quarterly**, **26**(4), 375-383.
- Emsley, J. (2003). **Nature's building blocks: an A-Z guide to the elements**. London: Oxford.
- Engelbrecht, M. (1999). **Leerpotensiaal as voorspeller Van akademiese Van universiteitstudente**. Unpublished D.Phil. thesis, Potchefstroom University for Christian Higher Education.
- Engelbrecht, P. (1996). Cognitive control therapy for South African children with learning difficulties. In P.Engelbrecht, S. Kriegler & M. Booysen (Eds.), **Perspectives on learning difficulties: international concerns and South African realities** (pp.199-206). Pretoria: Van Schaik.
- Engelbrecht, P.Kriegler, S. & Booysen, M. (Eds.), (1996). **Perspectives on learning difficulties: international concerns and South African realities**. Pretoria: Van Schaik.
- Entwistle, N. (1987). Explaining individual differences in school learning. In E. De Corte, H. Lodewijks, R. Parmentier & P.Span (Eds.), **Learning and instruction: European research in an international context volume 1** (pp.69-88). Oxford: Pergamon
- Ergang, R. (1967). **Europe: from the Renaissance to Waterloo**. Massachusetts: D.C. Heath.
- Ernest, P. (1994). (Ed.). **Mathematics, education and philosophy: an international perspective**. London: Farmer Press.
- Ernest, P.(1994). The dialogical nature of mathematics. In P.Ernest (Ed.). **Mathematics, education and philosophy: an international perspective** (pp.33-48). London: Farmer Press.
- Ernest, P.(1998). **Social constructivism as a philosophy of mathematics**. New York: New York Press.
- Erwin, E. (1985). The use of theory in psychology. In K.B. Madsen & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 3** (pp.73-85). New York: Plenum Press.
- Eskola, A. & Weckroth, K. (1996). Societal changes and psychological theory: the Russian case. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.85-98). London: Greenwood Press.
- Estany, A. (2001). The thesis of theory-laden observation in the light of cognitive psychology. **Philosophy of science**, **68**(2), 203 - 217.

- Esters, I.G. & Ittenbach, R.F. (1999). Contemporary theories and assessments of intelligence: a primer. **Professional school counselling**, 2(5), 373-376.
- Estes, D. / Bartsch, K. (1997). Constraining the brain: the role of developmental psychology on developmental cognitive neuroscience. **Behavioral and brain sciences**, 20(4), 562-563.
- Estes, W.K. (1975). Some targets for mathematical psychology. **Journal of mathematical psychology**, 12, 263-282.
- Estes, W.K. (1997). Significance testing in psychological research: some persisting issues. **Psychological science**, 8(1), 18-19.
- Estes, W.K. (1998). Where is intelligence? In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.63-67). New Jersey: Ablex.
- Ettin, M.F. (1984). Sound theories and theory soundings. In J.R. Royce & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 2** (pp.211-223). New York: Plenum Press.
- Everitt, B.S. & Wykes, T. (1999). **Dictionary of statistics for psychologists**. London: Oxford.
- Eves, H. & Newsom, C.V. (1965). **An introduction to the foundations and fundamental concepts of mathematics**. New York: Holt, Rinehart and Winston.
- Eysenck, H.J. (1985). The place of theory in a world of facts. In K.B. Madsen & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 3** (pp.17-72). New York: Plenum Press.
- Eysenck, H.J. (1986). Toward a new model of intelligence. **Personality and individual differences**, 7(5), 731-736.
- Eysenck, H.J. (1987). "There is nothing more practical than a good theory" (Kurt Lewin). True or false? In W.J. Baker., M.E. Hyland., H. Van Rappard. & A.W. Staats (Eds.), **Current issues in theoretical psychology: selected/edited proceedings of the founding conference of the international society for theoretical psychology held in Plymouth, U.K., 30 August - 2 September, 1985** (pp.49-64). Amsterdam: North-Holland.
- Eysenck, H.J. (1988). Is intelligence? In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.69-72). New Jersey: Ablex.
- Eysenck, H.J. (1994). A biological theory of intelligence. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4. Theories of human intelligence** (pp.117-149). New Jersey: Ablex.
- Eysenck, H.J. (1995). Meta-analysis squared - does it make sense? **American psychologist**, 50, 110-111.
- Eysenck, H.J. & Schoenthaler, S.J. (1997). Raising IQ level by vitamin and mineral supplementation. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.363-392). Cambridge: Cambridge University Press.
- Eysenck, M.W. & Keane, M.T. (1993). **Cognitive psychology: a student's handbook**. Sussex: Lawrence Erlbaum.
- Falik, L.H. (2000). Mediated learning experience and the counseling process. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.309-323). Amsterdam: Pergamon.
- Farmelo, G. (Ed.). (2002a). **It must be beautiful: great equations of modern science**. London: Granta.
- Farmelo, G. (2002b). Foreword: it must be beautiful. In G. Farmelo (Ed.). **It must be beautiful: great equations of modern science** (pp.ix-xvi). London: Granta.
- Faure, M. & Venter, A. (1993). Karl Popper's critical rationalism. In J. Snyman (Ed.), **Conceptions of social inquiry** (pp.29-52). Pretoria: Human Sciences Research Council.
- Faust, D. & Meehl, P.(1992). Using scientific methods to resolve questions in the history and philosophy of science: some illustrations. **Behavior therapy**, 23, 195-211.
- Faust, D. & Meehl, P.(2002). Using meta-scientific studies to clarify or resolve questions in the philosophy and history of science. **Philosophy of science**, 69, S185-S196.
- Fawcett, J. & Downs, F.S. (1992). **The relationship of theory and research**. Philadelphia: F.A. Davis.
- Fearn, N. (2001). **Zeno and the tortoise: how to think like a philosopher**. London: Atlantic Books.
- Feist, G.J. (1995). Psychology of science and history of psychology: putting behavioural generalization to the test. **Psychological inquiry**, 6(2), 119-123.
- Feldman, D.H. & Fowler, R.C. (1997). The nature(s) of developmental change: Piaget, Vygotsky, and the transition process. **New ideas in psychology**, 15(3), 195-210.
- Feldman, J.A. (1997). Radical empiricism is not constructive. **Behavioral and brain sciences**, 20(4), 563564).
- Fernández-Ballesteros, R. & Calero, M.D. (2000). The assessment of learning potential: the EPA instrument. C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.293-323). New York: Elsevier.
- Fernández-Ballesteros, R. & Calero, M.D. (2001). Not all testing is dynamic testing. **Issues in education**, 7(2), 172-175.
- Fernández-Ballesteros, R. & Colom, R. (2004). The psychology of human intelligence in Spain. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.79-103). Cambridge: Cambridge University Press.
- Ferrara, R.A., Brown, A.L. & Campione, J.C. (1986). Children's learning and transfer of inductive reasoning rules: studies of proximal development. **Child development**, 57(5), 1087-1099.
- Fetzer, J.H. (1993). **Foundations of philosophy of science: recent developments**. New York: Paragon.
- Fetzer, J.H. (2002). **Computers and cognition: why minds are not machines**. Dordrecht: Kluwer Academic.
- Feuerstein, R. (1972). Cognitive assessment of the socioculturally deprived child and adolescent. In L.J. Cronbach & P.J.D. Drenth (Eds.), **Mental tests and cultural adaptation** (pp.265-275). The Hague: Mouton.
- Feuerstein, R. (1994). Learning potential assessment device. In R. J. Sternberg (Ed.), **Encyclopedia of human intelligence vol 2** (pp.660-665). New York: Macmillan.



- Feuerstein, R. & Feuerstein, R.S. (2001). Is dynamic assessment compatible with the psychometric model? In A.S. Kaufman & N.L. Kaufman (Eds.), **Specific learning disabilities and difficulties in children and adolescents: psychological assessment and evaluation** (pp.218-246). Cambridge: Cambridge University Press.
- Feuerstein, R., Feuerstein, R.S., Falik, L.H. & Rand, Y. (Eds.), (2002). **The dynamic assessment of cognitive modifiability**. Israel: The ICELP Press.
- Feuerstein, R., Feuerstein, R. & Gross, S. (1997). The learning potential assessment device. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.297-313). New York: Guilford.
- Feuerstein, R. & Kozulin, A. (1995). The Bell Curve: getting the facts straight. **Educational leadership**, 52(7), 1-6.
- Feuerstein, R., Rand, Y., Jensen, M.R., Kaniel, S. & Tzuril, D. (1987). Prerequisites for assessment of learning potential: The LPAD model. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.35-51). New York: Guilford Press.
- Feuerstein, R.; Rand, Y. & Hoffman, M. (1979). **The dynamic assessment of retarded performers: the learning potential assessment device, theory, instruments and techniques**. Baltimore: University Park Press.
- Feuerstein, R.; Rand, Y. & Hoffman, M. & Miller, R. (1980). **Instrumental enrichment: an intervention program for cognitive modifiability**. Illinois: Scott, Foresman and Company.
- Feuerstein, R.S. (2000). Dynamic cognitive assessment and the instrumental enrichment program: origins and development. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.147-165). Amsterdam: Pergamon.
- Feuerstein, S. (2000). Questioning as a form of mediation. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.134-144). Amsterdam: Pergamon.
- Fienberg, S.E. & Resnick, D.P.(1997). Re-examining the Bell Curve. In B. Devlin, S.E. Fienberg, D.P.Resnick & K. Roeder (Eds.), **Intelligence, genes and success: scientists respond to The Bell Curve** (pp.3 - 19). New York: Springer-Verlag.
- Fikes, T.G. (2001). Evolutionary psychology as computational theory in the cognitive sciences. **Journal of psychology and theology**, 29(4), 340-360.
- Filatova, N.E. (1996). Understanding the other order of things. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.116-123). London: Greenwood Press.
- Finlay, B.L. (1997). So many problems, so little time: evolution and the dendrite. **Behavioral and brain sciences**, 20(4), 564-565.
- Finn, G.P.T. (2000). Piaget, Vygotsky and the social dimension. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.121-128). London: Routledge.
- Fischer, G.H. (1996). Models for an "Objective" assessment of treatment effects based on item response data. In I.Dennis & P.Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.133-150). New Jersey: Lawrence Erlbaum.
- Fischer, G.H. (1997). Structural Rasch models: some theory, applications and software. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.185-207). London: Ablex.
- Fischer, G.H. & Laming, D. (Eds.), (1994). **Contributions to mathematical psychology, psychometrics, and methodology**. New York: Springer-Verlag.
- Fischer, G.H. & Tanzer, N. (1994). Some LBTL and LLTM relationships. In G.H Fischer & D. Laming (Eds.), **Contributions to mathematical psychology, psychometrics, and methodology** (pp.278-303). New York: Springer-Verlag.
- Fischer, K.W. (1980). A theory of cognitive development: the control and construction of hierarchies of skills. **Psychological review**, 87(6), 477-531.
- Fischer, K.W. (1993). Analyzing diversity in developmental pathways: methods and concepts. In R. Case & W. Edelman (Eds.), **The new structuralism in cognitive development: theory and research on individual pathways** (pp.33-56). Basel: Karger
- Fischer, K.W. & Rose, S.P. (1998). **Educational leadership**, 56(3), 56-61
- Fisher, R.A. (1956). **Statistical methods and scientific inference**. Edinburgh: Oliver and Boyd.
- Fisher, R.A. (1958). **Statistical methods for research workers**. Edinburgh: Oliver and Boyd.
- Fiske, D.W. (1986). Specificity of method and knowledge in social science. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.61-82). Chicago: University of Chicago Press.
- Fiske, D.W. & Shweder, R.A. (Eds.), (1986). **Metatheory in social science: pluralisms and subjectivities**. Chicago: University of Chicago Press.
- Flanagan, D.P., Andrews, T.J. & Genshaft, J.L. (1997). The functional utility of intelligence tests with special education populations. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.457-483). New York: Guilford.
- Flanagan, D.P., Genshaft, J.L. & Harrison, P.L. (Eds.), (1997). **Contemporary intellectual assessment: theories, tests and issues**. New York: Guilford.
- Flanagan, D.P. & McGrew, K.S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: narrowing the gap between practice and cognitive science. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.314-325). New York: Guilford.
- Flanagan, D.P. & McGrew, K.S. (1998). Interpreting intelligence tests from contemporary *GfGc* theory: joint confirmatory factor analysis of the WJ-R and KAIT in a non-white sample. **Journal of school psychology**, 36(2), 151-182.

- Fletcher, R. (1991). Intelligence, equality, character, and education. **Intelligence**, **15**(2), 139-149.
- Flynn, J.R. (1998). IQ gains over time: toward finding the causes. In U. Neisser (Ed.), **The rising curve: long-term gains in IQ and related measures** (pp.25-66). Washington: American Psychological Association.
- Foley, D.E. (1997). Deficit thinking models based on culture: the anthropological protest. In R.R. Valencia (Ed.), **The evolution of deficit thinking: educational thought and practice** (pp.113-131). London: Falmer Press.
- Folman, R. & Budoff, M. (1971). Learning potential and vocational aspirations of retarded adolescents. **Exceptional children**, **38**(2), 121-131.
- Forrest-Pressley, D.L., MacKinnon, G.E. & Waller, T.G. (1985). (Eds.), **Metacognition, cognition and human performance: theoretical perspectives volume 1**. Orlando: Academic Press.
- Fox, D. (2000). The critical psychology project: transforming society and transforming psychology. In T. Sloan (Ed.), **Critical psychology: voices for change** (pp.21-45). London: Macmillan.
- Fox, D & Prilleltensky, I. (Eds.) (1997). **Critical psychology: an introduction**. London: Sage.
- Foxcroft, C. & Roodt, G. (2001). **An introduction to psychological assessment in the South African context**. Cape Town: Oxford University.
- Fraser, D.A.S. (1992). Introduction to Bartlett (1937) "Properties of sufficiency and statistical tests". In S. Kotz & N.L. Johnson (Eds.), **Breakthroughs in statistics volume 1: foundations and basic theory** (pp.109-126). New York: Springer-Verlag.
- Freeman, F.N. (1921). Intelligence and its measurement. **Journal of educational psychology**, **12**(3), 133-136.
- Freeman, L. & Miller, A. (2001). Norm-referenced, criterion-referenced, and dynamic assessment: what exactly is the point? **Educational psychology in practice**, **17**(1), 3-16.
- Freeze, G.L. (2000). From Stalinism to stagnation, 1953-1985. In G.L. Freeze (Ed.), **Russia: a history** (pp.347-382). Oxford: Oxford University Press.
- Freeze, G.L. (Ed.), (2000). **Russia: a history**. Oxford: Oxford University Press.
- Frege, G. (1983). The concept of number. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.130-159). Cambridge: Cambridge University Press.
- Friedman, S.L. & Cocking, R.R. (1986). Instructional influences on cognition and on the brain. In S.L. Friedman, K.A. Klivington & R.W. Peterson (Eds.), **The brain, cognition, and education** (pp.319-343). Orlando: Academic Press.
- Friedman, S.L., Klivington, K.A. & Peterson, R.W. (Eds.), (1986). **The brain, cognition, and education**. Orlando: Academic Press.
- Frisby, C.L. (2001). Academic achievement. In L.A. Suzuki, J.G. Ponterotto & P.J. Meller (Eds.), **Handbook of multicultural assessment: clinical, psychological, and educational applications** (pp.541-568). San Francisco: Jossey-Bass.
- Frisby, C.L. & Braden, J.P.(1992). Feuerstein's dynamic assessment approach: a semantic, logical and empirical critique. **Journal of special education, (Electronic version)** **26**(3), 281-300.
- Fuchs, W.R. (1967). **Modern mathematics**. London: Weidenfeld & Nicolson.
- Fuster, J.M. (2005). The cortical substrate of general intelligence. **Cortex**, **41**(2), 228-229.
- Galbraith, R.C. (1983). Individual differences in intelligence: a reappraisal of the confluence model. **Intelligence**, **7**, 185-194.
- Galperin, P.Y. & Talyzina, N.F. (1961). Formation of elementary geometrical concepts and their dependence on directed participation by the pupils. In N. O'Connor (Ed.), **Recent Soviet psychology** (pp.247-272). Oxford: Pergamon.
- Galton, F. (1981). Hereditary genius: An inquiry into its laws and consequences. In W.S. Sahakian (Ed), **History of psychology: a source book in systematic psychology** (pp.223-230). Illinois: F.E. Peacock.
- Gamlin, P. (1996). Issues in dynamic assessment/instruction. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.79-87). Ontario: Captus Press.
- Gamlin, P., Elmpak, M., Franke-Wagar, P., Lotfabadi, H., Nourani, K. & Wang, L-C. (1996). Teaching number concepts to young children: a dynamic approach for linking relational amount concepts to a formal number system. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.272-297). Ontario: Captus Press.
- Garbers, J.G. (Ed.), (1996). **Effective research in the human sciences**. Pretoria: J.L. Van Schaik.
- Gardner, H. (1986). Notes of cognitive development: recent trends, new directions. In S.L. Friedman, K.A. Klivington & R.W. Peterson (Eds.), **The brain, cognition, and education** (pp.259-285). Orlando: Academic Press.
- Gardner, H. (1998). The waning of intelligence tests. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.73-76). New Jersey: Ablex.
- Gardner, H., Hatch, T. & Torff, B. (1997). A third perspective: the symbols systems approach. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.243-268). Cambridge: Cambridge University Press.
- Gardner, H., Kornhaber, M.L. & Wake, W.K. (1996). **Intelligence: multiple perspectives**. New York: Harcourt Brace College.
- Gardner, M. (2003). **Are universes thicker than blackberries? Discourses on Gödel, magical hexagrams, little red riding hood, and other mathematical and pseudoscientific topics**. New York: W.W. Norton.
- Garner, R. (1988). **Metacognition and reading comprehension**. New Jersey: Ablex.
- Gauch, H.G. (2006). Winning the accuracy game. **American scientist**, **94**(2), 133-141.
- Gaydon, V.P.(1988). **Predictors of performance of disadvantaged adolescents on the Soweto/Alexandra gifted child programme**. Unpublished M.Ed. dissertation, University of the Witwatersrand.
- Gaylord-Ross, R. (Ed.), (1990). **Issues and research in special education volume 1**. New York: Teachers College.

- Gazzaniga, M.S. (1995). The puzzle of averages. J. Brockman & K. Matson (Eds.). **How things are: a science tool-kit for the mind** (pp.165-168). London: Weidenfeld and Nicolson.
- Gazzaniga, M.S. (2005). What's on your mind? **New Scientist**, **186**(2503), 48-50.
- Geisser, S. (1992). Introduction to Fisher (1922) "On the mathematical foundations of theoretical statistics". In S. Kotz & N.L. Johnson (Eds.), **Breakthroughs in statistics volume 1: foundations and basic theory** (pp.1-10). New York: Springer-Verlag.
- Geldenhuis, E.G. & Waterston, M. (1998). To determine intellectual ability is to talk about fair evaluation. **Salut**, **5**(2), 38-39.
- Gell-Mann, M. (1998). **The quark and the jaguar: adventures in the simple and the complex**. London: Abacus.
- Gelman, R. (2000). Domain specificity and variability in cognitive development. **Child development**, **71**(4), 854-856.
- Gelman, R. & Au, T.K-F. (Eds.), (1996). **Perceptual and cognitive development**. San Diego: Academic Press.
- Gentner, D. & Jeziorski, M. (1989). Historical shifts in scientific analogy. In B. Gholson, W.R. Shadish, R.A. Neimeyer, & A.C. Houts, (Eds.). **Psychology of science: contributions to metascience (pp.296-325)**. Cambridge: Cambridge University Press.
- Gerber, M.M. (2000). Dynamic assessment for students with learning disabilities: lessons in theory and design. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.263-292). New York: Elsevier.
- Gerber, M.M. (2001). All teachers are dynamic tests. **Issues in education**, **7**(2), 193-200.
- Gergen, K.J. (1986). Correspondence versus autonomy in the language of understanding human action. In D.W. Fiske & R.A. Shweder (Eds.). **Metatheory in social science: pluralisms and subjectivities** (pp.136-162). Chicago: University of Chicago Press.
- Gergen, K.J. (1987a). Introduction: toward metapsychology. In H.J. Stam, T.B. Rogers, & K.J. Gergen. (Eds.). **The analysis of psychological theory: metapsychological perspectives** (pp.1-21). Cambridge: Hemisphere.
- Gergen, K.J. (1987b). The language of psychological understanding. In H.J. Stam, T.B. Rogers, & K.J. Gergen. (Eds.). **The analysis of psychological theory: metapsychological perspectives** (pp.115-129). Cambridge: Hemisphere.
- Gerstlé, J. (1989). Concepts, theories and research programmes. **International social science journal**, **41**(4), 607-615.
- Gewer, A. (1998). **Uncovering potential: dynamic assessment of non-verbal reasoning ability in educationally disadvantaged children**. Unpublished M.Ed. dissertation, University of the Witwatersrand.
- Ghesquière, P. (2002). Perspectives in learning potential assessment: conclusive remarks. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.385-388). Amsterdam: JAI.
- Ghiselli, E.E., Campbell, J.P. & Zedeck, S. (1981). **Measurement theory for the behavioral sciences**. San Francisco: W.H. Freeman.
- Gholson, B., Shadish, W.R., Neimeyer, R.A. & Houts, A.C. (Eds.). (1989). **Psychology of science: contributions to metascience**. Cambridge: Cambridge University Press.
- Giere, R.N. (2001a). Cognitive approaches to science. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.41-43). Massachusetts: Blackwell.
- Giere, R.N. (2001b). Theories. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.520-524). Massachusetts: Blackwell.
- Gifford, B.R. & O'Connor, M.C. (Eds.), (1993). **Changing assessments: alternative views of aptitude, achievement and instruction**. Boston: Kluwer.
- Gigerenzer, G. (1991). From tools to theories: a heuristic of discovery in cognitive psychology. **Psychological review**, **98**(2), 254-267.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. & Kruger, L. (1990). **The empire of chance: how probability changed science and everyday life**. Cambridge: Cambridge University Press.
- Gignac, G.E. (2005). Openness to experience, general intelligence and crystallized intelligence: a methodological extension. **Intelligence**, **33**, 161-167.
- Gignac, G.E. (2006). Evaluating sub-test *g* saturation levels via the single trait-correlated uniqueness (STCU) SEM approach: evidence in favor of crystallized subtests as the best indicators of *g*. **Intelligence**, **34**, 29-46.
- Gilgen, A.R. (2000). Common prescriptions for psychology derived from dialectical materialism and chaos theory. **Psychological reports**, **86**, 487-492.
- Gilgen, A.R. & Gilgen, C.K. (1996). Historical background, analytical overview and glossary. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.3-53). London: Greenwood Press.
- Gillam, R. & McFadden, T.U. (1994). Redefining assessment as a holistic discovery process. **Journal of childhood communication disorders**, **16**(1), 36-40.
- Gillam, R.B., Pena, E.D. & Miller, L. (1999). Dynamic assessment narrative and expository discourse. **Topics in language disorders**, **20**(1), 33-47.
- Gillen, J. (2000). Versions of Vygotsky. **British journal of educational studies**, **48**(2), 183-198.
- Gindis, B. (1995a). Viewing the disabled child in the sociocultural milieu: Vygotsky's quest. **School psychology international**, **16**(2), 155-166.
- Gindis, B. (1995b). A voice from the future. **School psychology international**, **16**(2), 99-103.
- Gipps, C.V. (1994). **Beyond testing: towards a theory of educational assessment**. London: Farmer Press.

- Giorgi, A. (1985). Toward the articulation of psychology as a coherent discipline. In S. Koch & D.E. Leary (Eds.). **A century of psychology as science** (pp.46-59). New York: McGraw-Hill.
- Glas, C.A.W. (1997). Testing the generalized partial credit model. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.237-260). London: Ablex.
- Glaser, R. (1987). Learning theory and theories of knowledge. In E. De Corte, H. Lodewijks, R. Parmentier & P.Span (Eds.), **Learning and instruction: European research in an international context volume 1** (pp.397-414). Oxford: Pergamon
- Glaser, R. (1998). Intelligence as acquired proficiency. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.77-83). New Jersey: Ablex.
- Glaser, R. & Bassok, M. (1989). Learning theory and the study of instruction. **Annual review of psychology**, **40**, 631-666.
- Glaser, R., Fegusson, E.L. & Vosniadou, S. (1996). Introduction: cognition and the design of environments for learning. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), **International perspectives on the design of technology-supported learning environments** (pp.1-9). New Jersey: Lawrence Erlbaum.
- Glaspey, A.M. & Stoel-Gammon, C. (2005). Dynamic assessment in phonological disorders: the scaffolding scale of stimulability. **Topics in language disorders**, **25**(3), 220-230.
- Glass, G., McGraw, B. & Smith, M.L. (1981). **Meta-analysis in social research**. Beverly Hills: Sage.
- Gleick, J. (1991). **Chaos: making a new science**. London: Cardinal.
- Gleick, J. (Ed.). (2000). **The best American science writing 2000**. New York: Ecco Harper Collins.
- Gleitman, H. (1985). Some trends in the study of cognition. In S. Koch & D.E. Leary (Eds.). **A century of psychology as science** (pp.420-436). New York: McGraw-Hill.
- Glutting, J.J. & McDermott, P.A. (1990). Childhood learning potential as an alternative to traditional ability measures. **Psychological assessment: a journal of consulting and clinical psychology**, **2**(4), 398-403.
- Glymour, C. (1997). Social statistics and genuine inquiry: reflections on *The Bell Curve*. In B. Devlin, S.E. Fienberg, D.P. Resnick & K. Roeder (Eds.), **Intelligence, genes, and success: scientists respond to The Bell Curve** (pp.257-280).
- Goddard, H.H. (1912). **The Kallikak family: a study in the heredity of feeble-mindedness**. New York: Macmillan.
- Gödel, K. (1983a). Russell's mathematical logic. In P. Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.447-469). Cambridge: Cambridge University Press.
- Gödel, K. (1983b). What is Cantor's continuum problem?. In P. Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.470-485). Cambridge: Cambridge University Press.
- Goikoetxea, E. & Gondra, J.M. (1995). The performance of Gypsy children on the Raven's coloured progressive matrices after receiving additional instructions. **Journal of cognitive education**, **4**(2-3), 55-68.
- Goldberg, E. (1982). Learning disabilities and Soviet Psychology: are there basic assumptions? **Journal of clinical neuropsychology**, **4**(3), 291-295.
- Goldberg, M.F. (1991). Portrait of Reuven Feuerstein. **Educational leadership**, **49**(1), 37-40.
- Goldman, A.I. (1985). The relation between epistemology and psychology. **Synthese**, **64**, 29-68.
- Goldman-Rakic, P.S. (1986). Setting the stage: neural development before birth. In S.L. Friedman, K.A. Klivington & R.W. Peterson (Eds.), **The brain, cognition, and education** (pp.233-258). Orlando: Academic Press.
- Good, I.J. (1992). Comments on Cox (1957) statistical frontiers. In S. Kotz & N.L. Johnson (Eds.), **Breakthroughs in statistics volume 1: foundations and basic theory** (pp.xxxvii). New York: Springer-Verlag.
- Goodnow, J.J. (1998). A social view of intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.85-89). New Jersey: Ablex.
- Gordon, E.W. & Lemons, M.P. (1997). An interactionist perspective on the genesis of intelligence. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.323-340). Cambridge: Cambridge University Press.
- Gould, S.J. (1994). **Eight little piggies: reflections in natural history**. London: Penguin.
- Gould, S.J. (1997a). **The mismeasure of man**. London: Penguin.
- Gould, S.J. (1997b). **Dinosaur in a haystack: reflections in natural history**. London: Penguin.
- Gould, S.J. (1998). **Questioning the millennium: a rationalist's guide to a precisely arbitrary countdown**. London: Vintage.
- Gould, S.J. (2000a). More things in heaven and earth. In H. Rose & S. Rose (Eds.). **Alas, poor Darwin: arguments against evolutionary psychology** (pp.85-105). London: Jonathan Cape.
- Gould, S.J. (2000b). **Wonderful life: the Burgess shale and the nature of history**. London: Vintage.
- Gould, S.J. (2001a). **Bully for brontosaurus: reflections in natural history**. London: Vintage.
- Gould, S.J. (2001b). **The lying stones of Marrakech: penultimate reflections in natural history**. London: Vintage.
- Gould, S.J. (2002). **The structure of evolutionary theory**. Cambridge: Belknap.
- Gould, S.J. (2003). **I have landed: splashes and reflections in natural history**. London: Vintage.
- Gould, S.J. (2004). **The hedgehog, the fox, and the magister's pox: mending and minding the misconceived gap between science and the humanities**. London: Vintage.
- Gouws, A. (1997). Vygotsky se defektologiese idees: ou wyn in 'n nuwe sak? **South African journal of education**, **17**(1), 21-24.
- Govier, T. (2000). Operationalization. In J. McErlean (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.259-260). California: Wadsworth.
- Grace, R.C. & Farreras, I.G. (1998). Are theoretical consultants needed? **American psychologist**, **53**(1), 67-68.

- Graneheim, U.H. & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. **Nurse education today**, **24**, 105-112.
- Gratzer, W. (2002). **Eurekas and euphorias: the Oxford book of scientific anecdotes**. Oxford: Oxford University Press.
- Gray, J. (1989). The mind-brain identity theory as a scientific hypothesis: a second look. In C. Blakemore & S. Greenfield (Eds.), **Mindwaves: thoughts on intelligence, identity and consciousness** (pp.461-483). Cambridge: Basil-Blackwell.
- Gray, J.A. (1966). Attention, consciousness and voluntary control of behaviour in Soviet psychology: philosophical roots and research branches. In N. O'Connor (Ed.), **Present-day Russian psychology: a symposium by seven authors** (pp.1-38). Oxford: Pergamon.
- Gredler, G.R. (1988). Assessment of learning ability: a current appraisal. In R.M. Gupta & P.Coxhead (Eds.), **Cultural diversity and learning efficiency: recent developments in assessment** (pp.39-63). New York: St Martin's Press.
- Green, B. (2004). **The fabric of the cosmos: space, time, and the texture of reality**. London: Penguin.
- Green, C.D. (1995). Logic of abstract theory confirmation or logic of scientific practice? A response to Kukla. **New ideas in psychology**, **13**(3), 223-228.
- Green, C.D. (2001). Scientific models, connectionist networks, and cognitive science. **Theory and psychology**, **11**(1), 97-117.
- Greenberg, K.H. (2000). Inside professional practice: a collaborative, systems orientation to linking dynamic assessment and intervention. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.489-519). New York: Elsevier.
- Greenberg, K.H. & Williams, L. (2002). Reciprocity and mutuality in dynamic assessment: asking uncomfortable questions. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.91-110). Amsterdam: JAI.
- Greenfield, P.M. (1987). Electronic technologies, education and cognitive development. In D.E. Berger, K. Pezdek & W. Banks. (Eds.), **Applications of cognitive psychology: problem solving, education and computing** (pp17-32). New Jersey: Lawrence Erlbaum.
- Greenfield, P.M. (1997). You can't take it with you: why ability assessments don't cross cultures. **American psychologist**, **52**(10), 1115-1124.
- Greenfield, P.M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), **The rising curve: long-term gains in IQ and related measures** (pp.81-123). Washington: American Psychological Association.
- Greenfield, S. (1997). **The human brain: a guided tour**. London: Weidenfeld and Nicolson.
- Greenfield, S. (2001). **The private life of the brain**. London: Penguin.
- Greenough, W.T., Black, J.E. and Wallace, C.S. (1994). Experience and brain development. In M.H. Johnson, (Ed) **Brain development and cognition: a reader** (pp.290-332). Oxford: Blackwell.
- Greenspan, S. & Driscoll, J. (1997). The role of intelligence in a broad model of personal competence. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.131-150). New York: Guilford.
- Gregory, R.J. (2004). **Psychological testing: history, principles and applications**. Boston: Pearson.
- Gribbin, J. (Ed.). (1998). **A brief history of science**. London: Weidenfeld & Nicolson.
- Gribbin, J. (2003). **Science: a history**. London: Penguin.
- Grieshaber, S. & Ashby, G. (1997). Cognition, culture and curricula: an early childhood perspective. **Journal of cognitive education**, **6**(1), 39-51.
- Grieve, K.W. & Van Staden, F.J. (1988). A cross-cultural study of children's cognitive maps. **South African journal of psychology**, **18**(3), 91-95.
- Griffey, S. & Claxton, G. (1997). Unconscious mental resources: implications for the ideal learning environment. **Journal of cognitive education**, **6**(3), 13-22.
- Grigorenko, E.L. (1998). Russian "defectology": anticipating *Perestroika* in the field. **Journal of learning disabilities**, **31**(2), 193-207.
- Grigorenko, E.L. (2004a). Heritability and intelligence. In R.J. Sternberg (Ed.). **Handbook of intelligence** (pp.53-91). Cambridge: Cambridge University Press.
- Grigorenko, E.L. (2004b). Is it possible to study intelligence without using the concept of intelligence? In R.J. Sternberg (Ed.), **International Handbook of intelligence** (pp.170-211). Cambridge: Cambridge University Press.
- Grigorenko, E.L. & Kornilova, T.V. (1997). The resolution of the nature-nurture controversy by Russian psychology: culturally biased or culturally specific? In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.393-439). Cambridge: Cambridge University Press.
- Grigorenko, E.L. & Sternberg, R.J. (1998). Dynamic Testing. **Psychological Bulletin** **124** (1), 75-111.
- Groome, D. (1999). **An introduction to cognitive psychology: processes and disorders**. Sussex: Psychology Press.
- Gross, A. (2002). Testing the validity of a meta-reading model. **Journal of cognitive education and psychology**, **2**(2), 182-183.
- Grossberg, S. (1997). Neural models of development and learning. **Behavioral and brain sciences**, **20**(4), 566.
- Grossberg, S. & Gutowski, W.E. (1987). Neural dynamics of decision making under risk: affective balance and cognitive-emotional interactions. **Psychological review**, **94**(3), 300-318.
- Guilford, J.P. (1982). Cognitive psychology's ambiguities: some suggested remedies. **Psychological review**, **89**(1), 48-59.



- Gulgoz, S. & Kagitcibasi, C. (2004). Intelligence and intelligence testing in Turkey. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.248-269). Cambridge: Cambridge University Press.
- Gupta, R.M. & Coxhead, P. (Eds.), (1988). **Cultural diversity and learning efficiency: recent developments in assessment**. New York: St Martin's Press.
- Gupta, R.M. & Coxhead, P.(1988a). Introduction. In R.M. Gupta & P.Coxhead (Eds.), **Cultural diversity and learning efficiency: recent developments in assessment** (pp.xii-xvii). New York: St Martin's Press.
- Gupta, R.M. & Coxhead, P. (1988b). Why assess learning potential? In R.M. Gupta & P.Coxhead (Eds.), **Cultural diversity and learning efficiency: recent developments in assessment** (pp.1-21). New York: St Martin's Press.
- Guthke, J. (1982). The learning test concept - an alternative to the traditional static intelligence test. **German journal of psychology**, **6**(4), 306-324.
- Guthke, J. (1992). Learning tests - the concept, main research findings, problems and trends. **Learning and individual differences**, **4**(2), 137-151.
- Guthke, J. (1993a). Current trends in theories and assessment of intelligence. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.13-18). Amsterdam: Swets & Zeitlinger.
- Guthke, J. (1993b). Developments in learning potential assessment. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.43-67). Amsterdam: Swets & Zeitlinger.
- Guthke, J. & Beckmann, J.F. (2000a). Learning test concepts and dynamic assessment. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.175-190). Amsterdam: Pergamon.
- Guthke, J. & Beckmann, J.F. (2000b). The learning test concept and its application to practice. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.17-69). New York: Elsevier.
- Guthke, J., Beckmann, J.F. & Dobat, H. (1997). Dynamic testing - problems, uses, trends and evidence of validity. **Educational and child psychology**, **14**(4), 17-32.
- Guthke, J., Beckmann, J.F. & Stein, H. (1995). Recent research evidence on the validity of learning tests. In J.S. Carlson (Ed.), **Advances in cognition and educational practice vol 3: European contributions to dynamic assessment** (pp.117-143). London: JAI Press.
- Guthke, J. & Wingenfeld, S. (1992). The learning test concept: origins, state of the art, and trends. In H.C. Haywood & D. Tzurriel (Eds.), **Interactive assessment** (pp.64-93). New York: Springer-Verlag.
- Gutiérrez-Clellen, V.F. & Pena, E. (2001). Dynamic assessment of diverse children: a tutorial. **Language, speech, and hearing services in schools**, **32**, 212-224.
- Guttman, L. & Levy, S. (1991). Two structural laws for intelligence tests. **Intelligence**, **15**, 79-103.
- Hacker, D.K., Dunlosky, J. & Graesser, A.C. (Eds.), (1998). **Metacognition in educational theory and practice**. New Jersey: Lawrence Erlbaum.
- Hacking, I. (2000). Experimentation and scientific realism. In J. McErlean (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.189-200). California: Wadsworth.
- Hadji, C. (2000). Science, pedagogy and ethics in Feuerstein's theory and applications. A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.21-33). Amsterdam: Pergamon.
- Haenen, J. (2000). Gal'perian instruction in the ZPD. **Human development**, **43**, 93-98.
- Haenen, J. (2003). Outlining the teaching-learning process: Piotr Gal'perin's contribution. **Learning and instruction**, **11**, 157-170.
- Haertel, G.D., Walberg, H.J. & Weinstein, T. (1983). Psychological models of educational performance: a theoretical synthesis of constructs. **Review of educational research**, **53**(1), 75-91.
- Haith, G. (1997). Processing limitations can help neural growth build hierarchical representations. **Behavioral and brain sciences**, **20**(4), 566-567.
- Hájek, A. & Hall, N. (2002). Induction and probability. In P.Machamer & M. Silberstein (Eds.), **The Blackwell guide to the philosophy of science** (pp.149-172). Massachusetts: Blackwell.
- Halford, G. (1999). The development of intelligence includes capacity to process relations of greater complexity. In M. Anderson (Ed.), **The development of intelligence** (pp.193-213). Sussex: Psychology Press.
- Halford, G.S. & Mccredden, J.E. (1998). Cognitive science questions for cognitive development: the concepts of learning, analogy, and capacity. **Learning and instruction**, **8**(4), 289-308.
- Halford, G.S. & Wilson, W.H. (1980). A category theory approach to cognitive development. **Cognitive psychology**, **12**, 356-411.
- Hall, J.A., Tickle-Degnen, L., Rosenthal, R. & Mosteller, F. (1994). Hypotheses and problems in research synthesis. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp17-28). New York: Russell Sage Foundation.
- Hall, S.S. (2000). Journey to the center of my mind. In J. Gleick (Ed.). **The best American science writing 2000** (pp.83-94). New York: Ecco Harper Collins.
- Halpern, D.F. (1987). Analogies as a critical thinking skill. In D.E. Berger, K. Pezdek & W. Banks. (Eds.), **Applications of cognitive psychology: problem solving, education and computing** (pp75-86). New Jersey: Lawrence Erlbaum.

- Hambleton, R.K. & Rogers, H.J. (1991). Evaluation of the plot method for identifying potentially biased test items. In S.H. Irvine & S.E. Newstead (Eds.), **Intelligence and cognition: contemporary frames of reference** (pp.307-330). Dordrecht: Martinus Nijhoff.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991). **Fundamentals of item response theory**. London: Sage.
- Hambrick, D.Z. (2005). The role of domain knowledge in higher-level cognition. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.361-372). London: Sage.
- Hambrick, D.Z., Kane, M.J. & Engle, R.W. (2005). The role of working memory in higher-level cognition: domain-specific versus domain-general perspectives. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.104-121). Cambridge: Cambridge University Press.
- Hamers, J.H.M., Hessels, M.G.P. & Pennings, A.H. (1996). Learning potential in ethnic minority children. **European journal of psychological assessment**, **12**(3), 183-192.
- Hamers, J.H.M., Hessels, M.G.P. & Tissink, J. (1995). Research on learning potential assessment. In J.S. Carlson (Ed.), **Advances in cognition and educational practice vol 3: European contributions to dynamic assessment** (pp.145-183). London: JAI Press.
- Hamers, J.H.M. & Resing, W.C.M. (1993). Learning potential assessment: introduction. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsseenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.23-41). Amsterdam: Swets & Zeitlinger.
- Hamers, J.H.M. & Sijtsma, K. (1993). Learning potential assessment: epilogue. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsseenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.365-376). Amsterdam: Swets & Zeitlinger.
- Hamers, J.H.M. & Sijtsma, K. (1995). Trends in learning potential assessment. In J.S. Carlson (Ed.), **Advances in cognition and educational practice vol 3: European contributions to dynamic assessment** (pp.83-116). London: JAI Press.
- Hamers, J.H.M., Sijtsma, K & Ruijsseenaars, A.J.J.M. (Eds.). (1993). **Learning Potential assessment: theoretical, methodological and practical issues**. Amsterdam: Swets & Zeitlinger.
- Haney, M.R. & Evans, J.G. (1999). National survey of school psychologists regarding use of dynamic assessment and other non-traditional assessment techniques. **Psychology in the schools**, **36**(4), 295-304.
- Hansen, A. (2003). On mediation from Magne Nyborg's perspective. **Journal of cognitive education and psychology**, **3**(1), 54-70.
- Harding, S. (2001). Feminist justificatory strategies. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.427-443). Massachusetts: Blackwell.
- Harris, C.W. (Ed.), (1967). **Problems in measuring change**. London: University of Wisconsin Press.
- Harlow, L.L. (1997). Significance testing introduction and overview. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.1-17). New Jersey: Lawrence Erlbaum.
- Harlow, L.L., Mulaik, S.A. & Steiger, J.H. (Eds.), (1997). **What if there were no significance tests?** New Jersey: Lawrence Erlbaum.
- Harnish, R.M. (2002). **Minds, brains, computers: an historical introduction to the foundations of cognitive science**. Oxford: Blackwell.
- Harré, R. (1988). **The philosophies of science**. Oxford: Oxford university press.
- Harré, R. (1997). "Berkeleyan" arguments and the ontology of cognitive science. In D.M. Johnson, & C.E. Erneling (Eds.), **The future of the cognitive revolution** (pp.335-352). New York: Oxford.
- Harré, R. (2001). Dispositions and powers. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.97-101). Massachusetts: Blackwell.
- Harré, R. (2002). **Great scientific experiments. Twenty experiments that changed our view of the world**. New York: Dover.
- Harris, R.J. (1997). Reforming significance testing via three-valued logic. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.145-174). New Jersey: Lawrence Erlbaum.
- Harrison, P.L., Flanagan, D.P. & Genshaft, J.L. (1997). An integration and synthesis of contemporary theories, test, and issues in the field of intellectual assessment. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.533-561). New York: Guilford.
- Harrison, R., Singer, J., Budoff, M. & Folman, R. (1972). Level of aspiration as a function of learning-potential status in the educable mentally retarded. **Psychological reports**, **30**, 47-57.
- Haviland, W.A. (1991). **Anthropology**. Fort Worth: Holt, Rinehart & Winston.
- Hawking, S. (2003). **On the shoulders of giants: the great works of physics and astronomy**. London: Penguin.
- Hawkins, M. (1997). **Hunting down the universe: the missing mass, primordial black holes and other dark matters**. London: Abacus.
- Hay, S. (1999). The developmental genetics of intelligence. In M. Anderson (Ed.), **The development of intelligence** (pp.75-104). Sussex: Psychology Press.
- Haywood, H.C. (2000). Cognitive-developmental therapy: overview. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.292-308). Amsterdam: Pergamon.
- Haywood, H.C. (2001a). What is dynamic testing? A response to Sternberg and Grigorenko. **Issues in Education**, **7**(2), 201-210.

- Haywood, H.C. (2001b). Clinical applications of cognitive education. **Journal of cognitive education and psychology**, 2(1), 22-36.
- Haywood, H.C. (2003). Mediation within a neo-Piagetian framework. **Journal of cognitive education and psychology**, 3(1), 71-81.
- Haywood, H.C. & Miller, M.B. (2003). Dynamic assessment of adults with traumatic brain injuries. **Journal of cognitive education and psychology**, 3(1), 137-163.
- Haywood, H.C. & Switzky, H.N. (1992). Ability and modifiability: what, how, and how much? In J.S. Carlson (Ed.), **Advances in cognition and educational practice: a research annual volume 1A** (pp.25-85). Connecticut: JAI Press.
- Haywood, H.C. & Tzuriel, D. (Eds.), (1992). **Interactive assessment**. New York: Springer-Verlag.
- Haywood, H.C., Tzuriel, D. & Vaught, S. (1992). Psychoeducational assessment from a transactional perspective. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.38-63). New York: Springer-Verlag.
- Healy, R. (2001). Quantum mechanics. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.376-384). Massachusetts: Blackwell.
- Hébert, J-P.(2005). Divertimento for strings. **American Scientist**, 93(6), 488-490.
- Heck, R.H. (1998). Factor analysis: exploratory and confirmatory approaches. In G.A. Marcoulides (Ed.), **Modern methods for business research** (pp.177-215). New Jersey: Lawrence Erlbaum.
- Hedges, L.V. (1994). Statistical considerations. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.29-38). New York: Russell Sage Foundation.
- Hegarty, S. (1988). Learning ability and psychometric practice. In R.M. Gupta & P.Coxhead (Eds.), **Cultural diversity and learning efficiency: recent developments in assessment** (pp.22-38). New York: St Martin's Press.
- Heitz, R.P., Unsworth, N. & Engle, R.W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.61-77). London: Sage.
- Helms, J.E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? **American psychologist**, 47(9), 1083-1101.
- Helms, J.E. (1997). The triple quandary of race, culture, and social class in standardized cognitive ability testing. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.517-532). New York: Guilford.
- Hempel, C. (1970). **Aspects of scientific explanation and other essays in the philosophy of science**. New York: Free Press.
- Hempel, C. (1983). On the nature of mathematical truth. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.377-393). Cambridge: Cambridge University Press
- Hempel, C. (1993). Principles of definition. In J.H. Fetzer (Ed.). **Foundations of philosophy of science: recent developments (pp.7-16)**. New York: Paragon.
- Hempel, C. (2000). Theoretical reduction. In J. McErlean (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.470-476). California: Wadsworth.
- Hempel, C. & Oppenheim, P.(2000). Studies in the logic of explanation. In J. McErlean (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.57-67). California: Wadsworth.
- Henley, S.J. (1989). **An investigation of Feuerstein's theory of mediated learning experience with a disadvantaged community**. Unpublished M.A. dissertation, University of the Witwatersrand.
- Henning, E., Van Rensburg, W. & Smit, B. (2004). **Finding your way in qualitative research**. Pretoria: Van Schaik.
- Henry, P.J., Sternberg, R.J. & Grigorenko, E.L. (2005). Capturing successful intelligence through measures of analytic, creative, and practical skills. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.295-311). London: Sage.
- Herrington, J. & Standen, P. (1999). Moving from an instructivist to a constructivist multimedia learning environment. **Journal of educational multimedia and hypermedia**, 9(3), 195-205.
- Hersh, R. (1994). Fresh breezes in the philosophy of mathematics. In P.Ernest (Ed.). **Mathematics, education and philosophy: an international perspective** (pp.11-20). London: Farmer Press.
- Hershberger, S.L. (1999). Introduction to personality measurement. In S.E. Embretson & S.L. Hershberger (Eds.), **The new rules of measurement: what every psychologist and educator should know** (pp.153-158). New Jersey: Lawrence Erlbaum.
- Hertz, N. (2002). **The silent takeover: global capitalism and the death of democracy**. London: Arrow.
- Hertzog, C. & Robinson, A.E. (2005). Metacognition and intelligence. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.101-123). London: Sage.
- Hessels, M.G.P. (1996). Ethnic differences in learning potential test scores: research into item and test bias in the learning potential test for ethnic minorities. **Journal of cognitive education**, 5(2), 133-153.
- Hessels, M.G.P. (2000). The learning potential test for ethnic minorities (LEM): a tool for standardized assessment of children in kindergarten and the first years of primary school. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.109-131). New York: Elsevier.
- Hessels, M.G.P. & Hamers, J.H.M. (1993). The learning potential test for ethnic minorities. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.285-311). Amsterdam: Swets & Zeitlinger.

- Hessels, M.G.P. & Hessels-Schlatter, C. (2002). Learning potential in immigrant children in two countries: the Dutch and the Swiss-French version of the learning potential test for ethnic minorities. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsseenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.227-245). Amsterdam: JAI.
- Hessels-Schlatter, C. (2002a). Moderate mental retardation and learning capacity: the analogical reasoning learning test. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsseenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.249-271). Amsterdam: JAI.
- Hessels-Schlatter, C. (2002b). The analogical reasoning learning test: theoretical and empirical foundation of a diagnostic tool for individuals with moderate mental retardation. **Journal of cognitive education and psychology**, 2(2), 177-179.
- Hey, T. & Walters, P.(1997). **Einstein's mirror**. Cambridge: Cambridge University Press.
- Heyes, C.M. (1989). Relationship between psychology and epistemology. In B. Gholson, W.R. Shadish, R.A. Neimeyer, & A.C Houts, (Eds.). **Psychology of science: contributions to metascience (pp.115-137)**. Cambridge: Cambridge University Press.
- Heyting, A. (1983a). The intuitionist foundations of mathematics. In P.Benacerraf & H. Putnam (Eds.). (1983). **Philosophy of mathematics: selected readings** (pp.52-61). Cambridge: Cambridge University Press.
- Heyting, A. (1983b). Disputation. In P.Benacerraf & H. Putnam (Eds.), (1983). **Philosophy of mathematics: selected readings** (pp.66-76). Cambridge: Cambridge University Press.
- Hilbert, D. (1983). On the infinite. In P.Benacerraf & H. Putnam (Eds.), **Philosophy of mathematics: selected readings** (pp.183-201). Cambridge: Cambridge University Press.
- Hilderbrandt, C. & Oliver, J. (2000). The mind as black box: a simulation of theory building in psychology. **Teaching of psychology**, 27(3), 195-197.
- History of Bayesian probability**. (2006). Retrieved 6 January 2006 from http://en.wikipedia.org/wiki/Bayesian_probability
- Hodapp, R.M. & Zigler, E. (1999). Intellectual development and mental retardation - some continuing controversies. In M. Anderson (Ed.), **The development of intelligence** (pp.295-308). Sussex: Psychology Press.
- Höfding, H. (1955). **A history of modern philosophy: a sketch of the history of philosophy from the close of the Renaissance to our own day**. Vol 1. New York: Dover.
- Hoffenberg, S.R. (1988). **Effectiveness of the learning potential assessment device with high achieving adolescents from an advantaged community**. Unpublished M.Ed. dissertation, University of the Witwatersrand.
- Hofstadter, D.R. (1980). **Gödel, Escher, Bach: an eternal golden braid. A metaphorical fugue on minds and machines in the spirit of Lewis Carroll**. London: Penguin.
- Hofstadter, D.R. (2000). Analogy as the core of cognition. In J. Gleick (Ed.). **The best American science writing 2000** (pp.116-144). New York: Ecco Harper Collins.
- Hofstadter, D.R. & Dennett, D.C. (Eds.) (1981). **The mind's I: fantasies and reflections on self and soul**. London: Penguin
- Holsti, O.R. (1969). **Content analysis for the social sciences and humanities**. Philippines: Addison-Wesley.
- Hooker, C.A. (2001). The unity of science. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.540-549). Massachusetts: Blackwell.
- Hookway, C. (2001). Holism. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.162-164). Massachusetts: Blackwell.
- Hopper, B. & Hurry, P. (2000). Learning the MI way: the effects on students' learning of using the theory of multiple intelligences. **Pastoral care in education**, 18(4), 26-32.
- Horgan, J. (1994). Can science explain consciousness? **Scientific America**, 271(1), 88-94.
- Horgan, J. (2000). **The undiscovered mind: how the brain defies explanation**. London: Phoenix.
- Horn, G. (1994). Brain mechanisms of memory and predispositions: interactive studies of cerebral function and behaviour. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.481-509). Oxford: Blackwell.
- Horn, J. (1988). Thinking about human abilities. In J.R. Nesselroade & R.B. Cattell (Eds.), **Handbook of multivariate experimental psychology** (pp.645-685). New York: Plenum.
- Horn, J. (1998a). A basis for research on age differences in cognitive capabilities. In J.J. McArdle & R.W. Woodcock (Eds.), **Human cognitive abilities in theory and practice** (pp.57-91). New Jersey: Lawrence Erlbaum.
- Horn, J. (1998b). Some thoughts about intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.91-96). New Jersey: Ablex.
- Horn, J.L. & Noll, J. (1994). A system for understanding cognitive capabilities: a theory and the evidence on which it is based. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4. Theories of human intelligence** (pp.151-203). New Jersey: Ablex.
- Horn, J.L. (1979). Some correctable defects in research on intelligence. **Intelligence**, 3, 307-322.
- Horwich, P.(1978). An appraisal of Glymour's confirmation theory. **Journal of philosophy**, 75, 98-113.
- Houts, A.C. (1989). Psychology of science and metascience. In B. Gholson, W.R. Shadish, R.A. Neimeyer, & A.C Houts, (Eds.). **Psychology of science: contributions to metascience (pp.47-88)**. Cambridge: Cambridge University Press.
- Howard, M.C. (1989). **Contemporary cultural anthropology**. Burnaby: Harper Collins.
- Howard, R.W. (1998). Every psychologist needs to be a theorist. **American psychologist**, 53(1), 69-70.



- Howe, M.J.A. (1987). Motivation, cognition and individual achievements. In E. De Corte, H. Lodewijks, R. Parmentier & P.Span (Eds.), **Learning and instruction: European research in an international context volume 1** (pp.133-146). Oxford: Pergamon
- Howe, M.J.A. (1996). Concepts of ability. In I.Dennis & P.Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.39-48). New Jersey: Lawrence Erlbaum.
- Howell, D.C. (1987). **Statistical methods for psychology**. Boston: Duxbury.
- Howie, D. (2003). The assessment and training of decision-making and self-advocacy skills. **Journal of cognitive education and psychology**, 3(1), 1-26.
- Howson, C. (2001a). Evidence and confirmation. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.108-116). Massachusetts: Blackwell.
- Howson, C. (2001b). Induction and the uniformity nature. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.181-183). Massachusetts: Blackwell.
- Huber, F. (2002). **Assessing theories: the problem of a quantitative theory of confirmation**. Unpublished D. Phil dissertation, University of Erfurt.
- Huberty, C.J. (1993). Historical origins of statistical testing practices: the treatment of Fisher versus Neyman-Pearson views in textbooks. **Journal of experimental education**, 61(4), 317-333.
- Hughes, D.M. (2000). Significant differences: the construction of knowledge, objectivity, and dominance. In J. McErlean (Ed.), **Philosophies of science: from foundations to contemporary issues** (pp.443-455). California: Wadsworth.
- Hughes, J. (1981). **The philosophy of social research**. London: Longman.
- Hughes, M. (2001). Berkeley. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.12-15). Massachusetts: Blackwell.
- Hume, D. (1981). A treatise on human nature . In W.S. Sahakian (Ed.). **History of psychology: a source book in systematic psychology** (pp.39-43). Illinois: F.E. Peacock.
- Humphreys, L.G. (1988). Trends in levels of academic achievement of blacks and other minorities. **Intelligence**, 12(3), 231-260.
- Humphreys, L.G. (1994). Intelligence from the standpoint of a (pragmatic) behaviourist. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4. Theories of human intelligence** (pp.205-232). New Jersey: Ablex.
- Humphreys, L.G. (1998). Describing the elephant. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.97-100). New Jersey: Ablex.
- Hunt, E. (1992). Why is it hard to improve mental competence? A cognitive science perspective. In J.S. Carlson (Ed.), **Advances in cognition and educational practice: a research annual volume 1A** (pp.3-24). Connecticut: JAI Press.
- Hunt, E. (1994). Theoretical models for the study of intelligence. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4. Theories of human intelligence** (pp.233-256). New Jersey: Ablex
- Hunt, E. (1995). Where and when to represent students this way and that way: an evaluation of approaches to diagnostic assessment. In P.D. Nichols, S.F. Chipman & R.L. Brennan (Eds.), **Cognitively diagnostic assessment** (pp.411-429). New Jersey: Lawrence Erlbaum.
- Hunt, E. (1997a). The concept and utility of intelligence. In B. Devlin, S.E. Fienberg, D.P.Resnick & K. Roeder (Eds.), **Intelligence, genes, and success: scientists respond to The Bell Curve** (pp.157-176).
- Hunt, E. (1997b). Nature vs. nurture: the feeling of vujà dé. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.531-551). Cambridge: Cambridge University Press.
- Hunt, E. (1998). The heffalump of intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.101-107). New Jersey: Ablex.
- Hunt, E. (2005). Information processing and intelligence: where we are and where we are going. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.1-25). Cambridge: Cambridge University Press.
- Hunt, E. & Sternberg, R.J. (2006). Sorry, wrong numbers: an analysis of a study of a correlation between skin colour and IQ. **Intelligence**, 34(2), 131-137.
- Hunter, J.E. (1997). Needed: a ban on the significance test. **Psychological science**, 8(1), 3-7.
- Hunter, J.E., Schmidt, F.L. & Jackson, G.B. (1982). **Meta-analysis: cumulating research findings across studies**. London: Sage.
- Hurtford, J., Joseph, S., Kirby, S. & Reid, A. (1997). Evolution might select constructivism. **Behavioral and brain sciences**, 20(4), 567-568.
- Huttenlocher, P.R. (1994). Morphometric study of human cerebral cortex development. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp 112-124). Oxford: Blackwell.
- Huysamen, G.K. (1980). **Psychological testing in South Africa**. Pretoria: Academica.
- Huysamen, G.K. (2005). Null hypothesis significance testing: ramifications, ruminations and recommendations. **South African journal of psychology**, 35(1), 1-20.
- Hydén, L-C. (1988). **The conceptual structure of Soviet psychology in Vygotskij's, Leontjev's and Rubiņštejn's theories**. Stockholm: Stockholms Universitet.
- Ifrah, G. (2001). **The universal history of computing: from the abacus to the quantum computer**. New York: John Wiley.
- Innocenti, G.M. (1997). "Differentiationism" can reconcile selectionism and constructivism. **Behavioral and brain sciences**, 20(4), 568-569.

- Ippel, M.J. (1991). An information-processing approach to item equivalence. In S.H. Irvine & S.E. Newstead (Eds.), **Intelligence and cognition: contemporary frames of reference** (pp.377-396). Dordrecht: Martinus Nijhoff.
- Irtel, H. (1994). The uniqueness structure of simple latent trait models. In G.H Fischer & D. Laming (Eds.), **Contributions to mathematical psychology, psychometrics, and methodology** (pp.265-275). NewYork: Springer-Verlag.
- Irvine, S.H. (1987). Functions and constants in mental measurement: a taxonomic approach. In S.H. Irvine & S.E. Newstead (Eds.), **Intelligence and cognition: contemporary frames of reference** (pp.1-25). Dordrecht: Martinus Nijhoff.
- Irvine, S.H. & Dann, P.L (1991). Challenges of computer-based human assessment. In P.L. Dann, S.H. Irvine & J.M. Collis (Eds.), **Advances in computer-based human assessment** (pp.3-25). Dordrecht: Kluwer.
- Irvine, S.H. & Newstead, S.E. (Eds.) (1987). **Intelligence and cognition: contemporary frames of reference**. Dordrecht: Martinus Nijhoff.
- Ittenbach, R.F., Esters, I.G. & Wainer, H. (1997). The history of test development. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.17-31). New York: Guilford.
- Jacobs, E.L. (2001). The effects of adding dynamic assessment components to a computerized preschool language screening test. **Communication disorders quarterly**, **22**(4), 217-226.
- Jaffe, A. & Quinn, F.S. (1993). "Theoretical mathematics": toward a cultural synthesis of mathematics and theoretical physics. **Bulletin of the American mathematical society**, **29**(1), 1-13.
- Jansen, M.G.H. (1994). Parameters of the latent distribution in Rasch's Poisson counts model. In G.H Fischer & D. Laming (Eds.), **Contributions to mathematical psychology, psychometrics, and methodology** (pp.319-326). NewYork: Springer-Verlag.
- Jardine, L. (1999). **Ingenious pursuits: building the scientific revolution**. London: Little, Brown and company.
- Jasso, G. (1993). Building the theory of comparison processes: construction of postulates and derivation of predictions. In J. Berger & M. Zelditch (Eds.). **Theoretical research programs: studies in the growth of theory** (pp.213-264). California: Stanford University Press
- Jencks, C. (2000). EP, Phone home. In H. Rose & S. Rose (Eds.). **Alas, poor Darwin: arguments against evolutionary psychology** (pp.28-46). London: Jonathan Cape.
- Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? **Harvard educational review**, **39**(1), 1-123.
- Jensen, A.R. (1980). **Bias in mental testing**. New York: The Free Press.
- Jensen, A.R. (1982a). The debunking of scientific fossils and straw persons. **Contemporary education review**, **1**(2), 121-135.
- Jensen, A.R. (1982b). The chronometry of intelligence. In R.J. Sternberg (Ed.), **Advances in the psychology of human intelligence volume 1** (pp.255-310). New Jersey: Lawrence Erlbaum.
- Jensen, A.R. (1994). Phlogiston, animal magnetism, and intelligence. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4. Theories of human intelligence** (pp.257-284). New Jersey: Ablex.
- Jensen, A.R. (1997). The puzzle of nongenetic variance. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.42-88). Cambridge: Cambridge University Press.
- Jensen, A.R. (1998a). Intelligence: "definition", measurement, and future research. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.109-112). New Jersey: Ablex.
- Jensen, A.R. (1998b). **The g factor: the science of mental ability**. Connecticut: Praeger.
- Jensen, A.R. (2005). Mental chronometry and the unification of differential psychology. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.26-50). Cambridge: Cambridge University Press.
- Jensen, A.R. (2006). Comments on correlations of IQ with skin colour and geographic-demographic variables. **Intelligence**, **34**(2), 121-139.
- Jensen, M.R. (1992). Principles of change models in school psychology and education. In J.S. Carlson (Ed.), **Advances in cognition and educational practice: a research annual volume 1B** (pp.47-72). Connecticut: JAI Press.
- Jensen, M.R. (2000). The mindladder model: using dynamic assessment to help students learn to assemble and use knowledge. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.187-227). New York: Elsevier.
- Jensen, M.R., & Feuerstein, R. (1987). The learning potential assessment device: from philosophy to practice. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.379-402). New York: Guilford Press.
- Jensen, M.R., Feuerstein, R., Rand, Y., Kaniel, S. & Tzuril, D. (1988). Cultural difference and cultural deprivation: a theoretical framework for differential intervention. In R.M. Gupta & P.Coxhead (Eds.), **Cultural diversity and learning efficiency: recent developments in assessment** (pp.64-88). New York: St Martin's Press.
- Jepsen, R.H. (2000). Dynamic assessment of learners with severe developmental disabilities. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.577-605). New York: Elsevier.
- John, R. (1980). Theory construction in sociology: the competing approaches. **Mid-American review of sociology**, **5**(1), 15-36.
- John-Steiner, V. & Soubberman, E. (1978). Afterword. In M. Cole, V. John-Steiner, S. Scribner & E. Soubberman (Eds.), **Vygotsky's "Mind in society: the development of higher psychological processes"** (pp.121-133). Cambridge: Harvard University Press.
- Johnson, B.T. (2005). **DSTAT**. Retrieved 13 April 2005 from www.erlbaum.com

- Johnson, D. H. (1999). The insignificance of statistical significance testing. **Journal of Wildlife Management** 63(3), 763-772.
- Johnson, D.M. & Erneling C.E. (Eds.), (1997). **The future of the cognitive revolution**. New York: Oxford.
- Johnson, M.H. (1994a). Cortical maturation and the development of visual attention in early infancy. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.167-194). Oxford: Blackwell.
- Johnson, M.H. (1994b). Constraints on cortical plasticity. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.703-721). Oxford: Blackwell.
- Johnson, M.H. (Ed). (1994). **Brain development and cognition: a reader**. Oxford: Blackwell.
- Johnson, M.H., Bates, L., Karmiloff-Smith, A. & Plunkett, K. (1997). Constraints on the construction of cognition. **Behavioral and brain sciences**, 20(4), 569-570.
- Johnson, M.H. & Gilmore, R.O. (1996). Developmental cognitive neuroscience: a biological perspective on cognitive change. In R. Gelman & T.K-F. Au (Eds.), **Perceptual and cognitive development** (pp.333-372). San Diego: Academic Press.
- Johnston, V.S., Partridge, D. & Lopez, P.D. (1983). A neural theory of cognitive development. **Journal of theoretical biology**, 100, 485 - 509.
- Joravsky, D. (1989). **Russian psychology: a critical history**. Oxford: Basil Blackwell.
- Jordaan, W.J. (1989). What do we do before we do theory? **South African journal of psychology**, 19(2), 53-64.
- Jordaan, W.J. (1993). Cognitive science: from information-processing to acts of meaning. **South African journal of philosophy**, 12(4), 91-102.
- Jorna, R.J. (1995). The semiotic (sign-oriented) aspects of connectionism. In I. Lubek, R. Van Hezewijk., G. Pheterson & C.W. Tolman (Eds.), **Trends and issues in theoretical psychology** (pp.233-238). New York: Springer.
- Jungk, R. (1960). **Brighter than a thousand suns: a personal history of the atomic scientists**. London: Penguin.
- Kail, R. & Pellegrino, J.W. (1985). **Human intelligence: perspectives and prospects**. New York: W.H. Freeman.
- Kahn, R.J. (2000). Dynamic assessment of infants and toddlers. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.325-373). New York: Elsevier.
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. **Psychological review**, 80(4), 237-251.
- Kaku, M. (1999). **Hyperspace: a scientific odyssey through parallel universes, time warps, and the tenth dimension**. Oxford: Oxford University Press.
- Kaku, M. (2006). **Parallel worlds: the science of alternative universes and our future in the cosmos**. London: Penguin.
- Kalaian, H.A. & Raudenbush, S.W. (1996). A multivariate mixed linear model for meta-analysis. **Psychological methods**, 1(3), 227-235.
- Kalat, J.W. (2004). **Biological psychology**. California: Wadsworth.
- Kamphaus, R.W., Petoskey, M.D. & Morgan, A.W. (1997). A history of intelligence test interpretation. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.32-47). New York: Guilford.
- Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. **Psychological review**, 111(2), 512-523.
- Kandel, E.R. & Squire, L.R. (2000). Neuroscience: breaking down scientific barriers to the study of brain and mind. **Science**, 290(5494), 1113-1120.
- Kane, M.J. (2005). Full frontal fluidity? Looking in on the neuroimaging of reasoning and intelligence. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.141-163). London: Sage.
- Kanevsky, L. & Geake, J. (2004). Inside the zone of proximal development: validating a multifactor model of learning potential with gifted students and their peers. **Journal for the education of the gifted**, 28(2), 182-217.
- Kaniel, S. (2000). A metacognitive decision-making model for dynamic assessment and intervention. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.643-679). New York: Elsevier.
- Kaniel, S. (2001). Not all testing is dynamic testing. **Issues in education**, 7(2), 211-224.
- Kaniel, S. & Tzurriel, D. (1992). Mediated learning experience approach in the assessment and treatment of borderline psychotic adolescents. In H.C. Haywood & D. Tzurriel (Eds.), **Interactive assessment** (pp.399-). New York: Springer-Verlag.
- Kanselaar G. & Erkens, G. (1996). Interactivity in cooperative problem solving with computers. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), **International perspectives on the design of technology-supported learning environments** (pp.185-202). New Jersey: Lawrence Erlbaum.
- Karmiloff-Smith, A. (1994). Self-organization and change. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.592-618). Oxford: Blackwell.
- Karmiloff-Smith, A. (2000). Why babies' brains are not Swiss army knives. In H. Rose & S. Rose (Eds.). **Alas, poor Darwin: arguments against evolutionary psychology** (pp.144-156). London: Jonathan Cape.
- Karpov, Y. (2003). Vygotsky's concept of mediation. **Journal of cognitive education and psychology**, 3(1), 46-53.
- Karpov, Y.V. & Bransford, J.D. (1995). L.S. Vygotsky and the doctrine of empirical and theoretical learning. **Educational psychologist**, 30(2), 61-66.
- Karpov, Y.V. & Gindis, B. (2000). Dynamic assessment of the level of internalisation of elementary school children's problem-solving activity. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.133-154). New York: Elsevier.
- Karpov, Y.V. & Haywood, H.C. (1998). Two ways to elaborate Vygotsky's concept of mediation: implications for instruction. **American psychologist**, 53(1), 27-36.



- Katz, N., Kizony, R. & Parush, S. (2002). Visiomotor organization and thinking operations performance of school-age Ethiopian, Bedouin, and mainstream Israeli children. **Occupational therapy journal of research**, 22(1), 34-43.
- Kaufman, A.S. (2004). Tests of intelligence. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.445-476). Cambridge: Cambridge University Press.
- Kaufman, A.S. & Kaufman, N.L. (Eds.), (2001). **Specific learning disabilities and difficulties in children and adolescents: psychological assessment and evaluation**. Cambridge: Cambridge University Press.
- Kaufman, J.C. & Kaufman, A.S. (2006). Some considerations concerning neurological development and psychometric assessment. **Behavioral and brain sciences**, 29, 137-138.
- Kazdin, A.E. (Ed.), (1992). **Methodological issues and strategies in clinical research**. Washington: American Psychological Association.
- Keane, K.J. (1987). Assessing deaf children. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.360-376). New York: Guilford Press.
- Keane, K.J., Tannenbaum, A.J. & Krapf, G.F. (1992). Cognitive competence: reality and potential in the deaf. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.300-316). New York: Springer-Verlag.
- Keith, T.Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.373-402). New York: Guilford.
- Kellert, S.H. (2000). In the wake of chaos. In J. McErlean (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.511-525). California: Wadsworth.
- Kendall, M.G. (1978). Where shall the history of statistics begin? In E.S. Pearson & M.G. Kendall (Eds.). **Studies in the history of statistics and probability volume 1** (pp.45-46). London: Charles Griffin.
- Kendall, M.G. & Plackett, R.L. (Eds.), (1977). **Studies in the history of statistics and probability volume 2**. London: Charles Griffin.
- Kendler, H.H. (2002). Unified knowledge: fantasy or reality? **Contemporary psychology**, 47(5), 501-503.
- Kennedy, H. & Dehay, C. (1997). Cortical development: a progressive and selective mesh, with or without constructivism. **Behavioral and brain sciences**, 20(4), 570-571.
- Kenny, D.A. (2003). META-analysis: easy to answer. Version III. Retrieved March 10, 2004, from <http://davidakenny.net/meta.htm>
- Kenrick, D.T. (2001). Evolutionary psychology, cognitive science, and dynamical systems: building an integrative paradigm. **Current directions in psychological science**, 10(1), 13-17.
- Kerlinger, F.N. (1981). **Foundations of behavioral research**. New York: Holt, Rinehart and Winston.
- Kerr, S.T. (1997, November). **Why Vygotsky? The role of theoretical psychology in Russian education reform**. Paper presented at the Annual meeting of the American Association for the Advancement of Slavic Studies, Seattle, Washington, United States of America.
- Kester, E.S., Pena, E.D. & Gillam, R.B. (2001). Outcomes of dynamic assessment with culturally and linguistically diverse students: a comparison of three teaching methods within a test-teach-retest framework. **Journal of cognitive education and psychology**, 2(1), 42-59.
- Kevles, D.J. (1997). From eugenics to genetic manipulation. In J. Krige & D. Pestre (Eds.), **Science in the twentieth century** (pp.301-318). Amsterdam: Harwood Academic.
- Khalidi, M. (2001). Incommensurability. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.172-180). Massachusetts: Blackwell.
- Khani, M.S. & Gamlin, P. (1996). Validating the similarity thinking instrument with preschool children: an investigation of the ability of young children to generalize knowledge. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.127-145). Ontario: Captus Press.
- Killeen, P.R. (2005). An alternative to null-hypothesis significance tests. **Psychological science**, 16(5), 345-353.
- Kincaid, H. (2002). Social sciences. In P.Machamer & M. Silberstein (Eds.), **The Blackwell guide to the philosophy of science** (pp.290-311). Massachusetts: Blackwell.
- Kingsbury, G.C. & Houser, R.L. (1999). Developing computerised adaptive-tests for school children. In F. Drasgow & J.B. Olson-Buchanan (Eds.), **Innovations in computerized assessment** (pp.93-115). New Jersey: Lawrence Erlbaum.
- Kirkwood, M.W., Weiler, M.D., Bernstein, J.H., Forbes, P.W. & Waber, D.P.(2001). Sources of poor performance on the Rey-Osterrieth complex figure test among children with learning difficulties: a dynamic assessment approach. **Clinical neuropsychologist**, 15(3), 345-356.
- Kirschenbaum, R.J. (1998). Dynamic assessment and its use with underserved gifted and talented populations. **Gifted child quarterly**, 42(3), 140-147.
- Klauer, K.J. (1993). Learning potential testing: the effect of retesting. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsseenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.135-152). Amsterdam: Swets & Zeitlinger.
- Klauer, K.J. (2002). A new generation of cognitive training for children: a European perspective. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsseenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.147-174). Amsterdam: JAI.

- Klauer, K.J. & Phye, G.D. (1994). **Cognitive training for children: a developmental program of inductive reasoning and problem solving**. Seattle: Hogrefe & Huber.
- Kleene, S.C. (1967). **Mathematical logic**. New York: John Wiley and Sons.
- Klein, P.S. (1992a). Assessing cognitive modifiability of infants and toddlers: observations based on mediated learning experience. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.233-250). New York: Springer-Verlag.
- Klein, P.S. (1992b). Cognitive and emotional interplay in early development: mediational role of parents. In J.S. Carlson (Ed.), **Advances in cognition and educational practice: a research annual volume 1A** (pp.169-194). Connecticut: JAI Press.
- Klein, P.S. (2000). A mediational approach to early intervention. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.240-256). Amsterdam: Pergamon.
- Kleiter, G.D. (1994). Natural sampling: rationality without base rates. In G.H Fischer & D. Laming (Eds.), **Contributions to mathematical psychology, psychometrics, and methodology** (pp.375-388). New York: Springer-Verlag.
- Kliegl, R., Smith, J. & Baltes, P.B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. **Developmental psychology**, **25**(2), 247-256.
- Kline, P. (1998). **The new psychometrics: science, psychology and measurement**. London: Routledge.
- Kline, R.B. (2004). **Beyond significance testing: reforming data analysis methods in behavioral research**. Washington: American Psychological Association.
- Kline, T.J.B. (2005). **Psychological testing: a practical approach to design and evaluation**. London: Sage.
- Klvington, K.A. (1986). Building bridges among neuroscience, cognitive psychology, and education. In S.L. Friedman, K.A. Klvington & R.W. Peterson (Eds.), **The brain, cognition, and education** (pp.3-15). Orlando: Academic Press.
- Koch, S. (Ed.) (1959-1963). **Psychology: a study of a science** (vols 1 - 6). New York: McGraw-Hill.
- Koch, S. (1981). The nature and limits of psychological knowledge. **American psychologist**, **36**(3), 257-269.
- Koch, S. (1985). The nature and limits of psychological knowledge: lessons of a century qua "science". In S. Koch & D.E. Leary (Eds.), **A century of psychology as science** (pp.75-99). New York: McGraw-Hill.
- Koch, S. & Leary, D.E. (Eds.). (1985). **A century of psychology as science**. New York: McGraw-Hill.
- Kolb, B. (1994). Brain development, plasticity, and behaviour. In M.H. Johnson, (Ed) **Brain development and cognition: a reader** (pp.338-356). Oxford: Blackwell.
- Koltsova, V.A. (1996). Ideological and scientific discourse in Soviet psychological science. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.60-69). London: Greenwood Press.
- Koltsova, V.A., Oleinik, Y.N., Gilgen, A.R. & Gilgen, C.K. (Eds.), **Post-Soviet Perspectives on Russian Psychology**. London: Greenwood Press.
- Kordig, C.R. (1971). The comparability of scientific theories. **Philosophy of science**, **38**, 467-485.
- Kirmann, A. & Sporer, S.L. (1983). Learning tests - concepts and critical evaluation. **Studies in educational evaluation**, **9**(1), 169-184.
- Koslow, A. (1992). Quantitative nonnumerical relations in science: Eudoxus, Newton, and Maxwell. In C.W. Savage & P.Ehrlich (Eds.), **Philosophical and foundational issues in measurement theory** (pp.139-165). New Jersey: Lawrence Erlbaum.
- Kotelova, Y.V. (1972). From the history of Soviet labor psychology. In J. Brožek & D.I. Slobin (Eds.), **Psychology in the USSR: An historical perspective** (pp.197-210). New York: IASP.
- Kotz, S. & Johnson, N.L. (Eds.), (1992). **Breakthroughs in statistics volume 1: foundations and basic theory**. New York: Springer-Verlag.
- Kostyuk, G.S. (1972). The problem of child development in Soviet psychology. In J. Brožek & D.I. Slobin (Eds.), **Psychology in the USSR: An historical perspective** (pp.123-143). New York: IASP.
- Kozorovitskiy, Y., Gross, C.G., Kopil, C., Battaglia, L., McBreen, M., Stranahan, A.M & Gould, E. (2005). Experience induces structural and biochemical changes in the adult primate brain. **Proceedings of the national academy of sciences**, **102**, 17478-17482.
- Kozulin, A. (1986). The concept of activity in Soviet psychology: Vygotsky, his disciples and critics. **American psychologist**, **41**(3), 264-274.
- Kozulin, A. (1987). Social contexts misconstrued: the case of Soviet developmental psychology. **Human development**, **30**, 336-340.
- Kozulin, A. (1990). **Vygotsky's psychology: a biography of ideas**. Cambridge: Harvard University Press
- Kozulin, A. (1999). Cognitive learning in younger and older immigrant students. **School psychology international**, **20**(2), 177-190.
- Kozulin, A. (2000). The diversity of instrumental enrichment applications. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.257-273). Amsterdam: Pergamon.
- Kozulin, A. (2002a). Mediated learning experience - from theory to classroom applications. **School psychology international**, **23**(1), 5-6.
- Kozulin, A. (2002b). Sociocultural theory and the mediated learning experience. **School psychology international**, **23**(1), 7-35.

- Kozulin, A. (2005). Learning potential assessment: where is the paradigm shift? In D.B. Pillemer & S.H. White (Ed.), **Developmental psychology and social change: research, history, and policy** (pp.352-367). Cambridge: Cambridge University Press.
- Kozulin, A. & Garb, E. (2002). Dynamic assessment of EFL test comprehension. **School psychology international**, 23(1), 112-127.
- Kozulin, A., Gindis, B., Ageyev, V.S & Miller, S.M. (Eds.), (2003). **Vygotsky's educational theory in cultural context**. Cambridge: Cambridge University Press.
- Kozulin, A. & Presseisen, B.Z. (1995). Mediated learning experience and psychological tools: Vygotsky's and Feuerstein's perspectives in a study of student learning. **Educational psychologist**, 30(2), 67-75.
- Kozulin, A. & Rand, Y. (Eds.), (2002). **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology**. Amsterdam: Pergamon.
- Krantz, D.H. (1964). Conjoint measurement: the Luce-Tukey axiomatization and some extensions. **Journal of mathematical psychology**, 1, 248-277.
- Krantz, D.H. (1999). The null hypothesis testing controversy in psychology. **Journal of the American statistical association**, 94(448), 1372-1381.
- Krauss, L. (1997). **The physics of Star Trek**. London: Flamingo.
- Krech, D. (1970). Comment. In J.R. Royce (Ed.). **Toward unification in psychology: the first Banff conference on theoretical psychology** (pp.42-43). Toronto: University of Toronto Press.
- Krechevsky, M. & Gardner, H. (1994). Multiple intelligences in multiple contexts. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4: theories of intelligence** (pp.285-305). New Jersey: Ablex.
- Kreisel, G. (1983). Hilbert's programme. In P. Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.207-238). Cambridge: Cambridge University Press.
- Krementsov, N. (1997). Russian science in the twentieth century. In J. Krige & D. Pestre (Eds.), **Science in the twentieth century** (pp.777-795). Amsterdam: Harwood Academic.
- Kriegler, S.M. & Skuy, M. (1996). Perspectives on psychological assessment in South African schools. In P. Engelbrecht, S. Kriegler & M. Booysen (Eds.), **Perspectives on learning difficulties: international concerns and South African realities** (pp.109-122). Pretoria: Van Schaik.
- Krige, J. & Pestre, D. (Eds.) (1997). **Science in the twentieth century**. Amsterdam: Harwood Academic.
- Krippendorff, K. (1980). **Content analysis: an introduction to its methodology**. Beverly Hills: Sage.
- Krishnamurty, G.B., Kasovia-Schmitt, P. & Ostroff, D.J. (1995). **Statistics: an interactive text for the health and life sciences**. London: Jones and Bartlett.
- Kristensen, K.B., Slife, B.D. & Yanchar, S.C. (2000). On what basis are evaluations possible in a fragmented psychology? An alternative to objectivism and relativism. **The journal of mind and behavior**, 21(3), 273-288.
- Krueger, J. (2001). Null hypothesis significance testing: on the survival of a flawed method. **American psychologist**, 56(1), 16-26.
- Kuhn, D. (2000). The view from giants' shoulders. In L. Smith, J. Dockrell & P. Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.246-259). London: Routledge.
- Kuhn, T. S. (1962). **The structure of scientific revolutions**. Chicago: The University of Chicago Press.
- Kuhn, T.S. (1977). **The essential tension: selected studies in scientific tradition and change**. Chicago: The University of Chicago Press.
- Kulik, J.A. & Kulik, C.C. (1989). Meta-analysis in education. **International journal of educational research**, 13, 221-340.
- Kukla, A. (1989). Nonempirical issues in psychology. **American psychologist**, 44(5), 785-794.
- Kukla, A. (1990a). Clinical versus statistical theory appraisal. **Psychological inquiry**, 1(20), 160-161.
- Kukla, A. (1990b). Theoretical psychology, artificial intelligence, and empirical research. **American psychologist**, 45(6), 780-781.
- Kukla, A. (1992). Unification as a goal for psychology. **American psychologist**, 47(8), 1054-1055.
- Kukla, A. (1995a). Amplification and simplification as modes of theoretical analysis in psychology. **New ideas in psychology**, 13(3), 201-217.
- Kukla, A. (1995b). On the metatheory of psychology: a reply to Dawson, Green, MacKay and Rozeboom. **New ideas in psychology**, 13(3), 247-257.
- Kukla, A. (2001). **Methods of theoretical psychology**. London: MIT Press.
- Kyburg, H.E. (1992). Measuring errors of measurement. In C.W. Savage & P. Ehrlich (Eds.), **Philosophical and foundational issues in measurement theory** (pp.75-91). New Jersey: Lawrence Erlbaum.
- Kyllonen, P.C. (1996). Is working memory capacity Spearman's *g*? In I. Dennis & P. Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.49-75). New Jersey: Lawrence Erlbaum.
- Lakatos, I. (1993). History of science and its rational reconstruction. In J.H. Fetzer (Ed.). **Foundations of philosophy of science: recent developments (pp.381-413)**. New York: Paragon.
- Lamal, P.A. (1988). Radical behaviorism and selected these concerning theory appraisal. **The psychological record**, 38, 175-185.
- Lashley, K.S. (1930). Basic neural mechanisms in behavior. **Psychological review**, 37, 1-24.
- Lategan, F.E. (1991). Die selfkonsep en die ontplooiing Van potensiaal. **Lyra academica**, 6(1), 90-105.

- Lauchlan, F. & Elliott, J. (2001). The psychological assessment of learning potential. **British journal of educational psychology**, **71**(4), 647-665.
- Laughon, P. (1990). The dynamic assessment of intelligence: a review of three approaches. **School psychology review**, **19**(4), 459-471.
- Lautrey, J. (1993). Structure and variability: a plea for a pluralistic approach to cognitive development. In R. Case & W. Edelman (Eds.), **The new structuralism in cognitive development: theory and research on individual pathways** (pp.101-114). Basel: Karger
- Lautrey, J. & De Ribaupierre, A. (2004). Psychology of human intelligence in France and French-speaking Switzerland. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.104-134). **Cambridge: Cambridge University Press.**
- Lawler, E.J. & Ford, R. (1993). Metatheory and friendly competition in theory growth: the case of power processes in bargaining. In J. Berger & M. Zelditch (Eds.), **Theoretical research programs: studies in the growth of theory** (pp.172-210). California: Stanford University Press
- Lazarsfeld, P.F. (1977). Notes on the history of quantification in sociology - trends, sources and problems. In M.G. Kendall & R.L. Plackett (Eds.), **Studies in the history of statistics and probability volume 2** (pp.231-269). London: Charles Griffin.
- Leahey, T.H. (2000). **A history of psychology: main currents in psychological thought**. New Jersey: Prentice Hall.
- Leary, D.E. (Ed). (1990a). **Metaphors in the history of psychology**. Cambridge: Cambridge University Press.
- Leary, D.E. (1990b). Psyche's muse: the role of metaphor in the history of psychology. In D.E. Leary (Ed). **Metaphors in the history of psychology (pp.1-78)** Cambridge: Cambridge University Press.
- Lee, C.D. (2003). Cultural modelling: CHAT as a lens for understanding instructional discourse based on African American English discourse patterns. In A. Kozulin, B. Gindis, V.S. Ageyev & S.M. Miller (Eds.), **Vygotsky's educational theory in cultural context** (pp.393-410). Cambridge: Cambridge University Press.
- Lehman, H. (1979). **Introduction to the philosophy of mathematics**. London: Basil Blackwell.
- Lehman, J.F., Laird, J. & Rosenbloom, P. (2006). A gentle introduction to SOAR, an architecture for human cognition: 2006 update. <http://www.eecs.umich.edu/~soar/docs/Gentle.pdf>
- Lehmann, E.L. (1992). Introduction to Neyman and Pearson (1933) "On the problem of the most efficient tests of statistical hypotheses". In S. Kotz & N.L. Johnson (Eds.), **Breakthroughs in statistics volume 1: foundations and basic theory** (pp.67-72). New York: Springer-Verlag.
- Leiman, M. (1992). The concept of sign in the work of Vygotsky, Winnicott and Bakhtin: further integration of object relations theory and activity theory. **British journal of medical psychology**, **65**, 209-221.
- Leontiev, A.A. (1981). Sign and activity. In J.V. Wertsch (Ed.), **The concept of activity in Soviet psychology** (pp.242-255). New York: M.E. Sharpe.
- Leontiev, A.N. (1961). Learning as a problem in psychology. In N. O'Connor (Ed.), **Recent Soviet psychology** (pp.227-246). Oxford: Pergamon.
- Leplin, J. (2001). Realism and instrumentalism. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.393-401). Massachusetts: Blackwell.
- Levenson, T. (1997). **Measure for measure: how music and sciences together have explored the universe**. Oxford: Oxford University Press.
- Levitin, K. (Ed.), (1982). **One is not born a personality: profiles of Soviet education psychologists**. Moscow: Progress.
- Lewis, G.N. (1959). Methods of science; numbers. Schwartz, G. & Bishop, P.W. (1959a). **Moments of discovery: the origins of science**. (pp.85-91). Vol 1. New York: Basic books.
- Li, R. (1996). **A theory of conceptual intelligence: thinking, learning, creativity, and giftedness**. London: Praeger.
- Li, S-C. & Kunzmann, U. (2004). Research on intelligence in German-speaking countries. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.135-169). **Cambridge: Cambridge University Press.**
- Lichtenstein, P.E. (1980). Theoretical psychology: where is it headed? **The psychological record**, **30**, 447-458.
- Lidz, C.S. (1981). **Improving assessment of schoolchildren**. San Francisco: Jossey-Bass.
- Lidz, C.S. (Ed.), (1987). **Dynamic assessment: an interactional approach to evaluating learning potential**. New York: The Guilford Press.
- Lidz, C.S. (1987a). Historical perspectives. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.3-32). New York: Guilford Press.
- Lidz, C.S. (1987b). Cognitive deficiencies revisited. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.444-475). New York: Guilford Press.
- Lidz, C.S. (1992a). The extent of incorporation of dynamic assessment into cognitive assessment courses: a national survey of school psychology trainers. **Journal of special education**, **26**(3), 325-331.
- Lidz, C.S. (1992b). Dynamic assessment: some thoughts on the model, the medium, and the message. **Learning and individual differences**, **4**(2), 125-136.
- Lidz, C.S. (1995). Dynamic assessment and the legacy of L.S. Vygotsky. **School psychology international**, **16**(2), 143-153.
- Lidz, C.S. (1997) Dynamic assessment approaches. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.281-296). New York: Guilford Press.



- Lidz, C.S. (2000a). Theme and some variations on the concepts of mediated learning experience and dynamic assessment. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.166-174). Amsterdam: Pergamon.
- Lidz, C.S. (2000b). The application of cognitive functions scale (ACFS): an example of curriculum-based dynamic assessment. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.407-439). New York: Elsevier.
- Lidz, C.S. (2001). Multicultural issues and dynamic assessment. In L.A. Suzuki, J.G. Ponterotto & P.J. Meller (Eds.), **Handbook of multicultural assessment: clinical, psychological, and educational applications** (pp.523-539). San Francisco: Jossey-Bass.
- Lidz, C.S. (2002). Mediated Learning Experience (MLE) as a basis for an alternative approach to assessment. **School psychology international**, **23**(1), 68-84.
- Lidz, C.S. (2003). **Early childhood assessment**. New Jersey: John Wiley & Sons.
- Lidz, C.S. (2004). Assessment procedure with deaf students between the ages of four and eight years. **Educational and child psychology**, **21**(1), 59-74.
- Lidz, C.S. & Elliott, J.G. (Eds.), (2000a). **Dynamic assessment: prevailing models and applications. Advances in cognition and educational practice volume 6**. New York: JAI - Elsevier Science.
- Lidz, C.S. & Elliott, J.G. (2000b). Introduction. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.3-13). New York: Elsevier.
- Lidz, C.S. & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, B. Gindis, V.S. Ageyev & S.M. Miller (Eds.), **Vygotsky's educational theory in cultural context** (pp.99-116). Cambridge: Cambridge University Press.
- Lidz, C.S. & Pena, E. D. (1996). Dynamic assessment: the model, its relevance as a nonbiased approach, and its application to Latino American preschool children. **Language, speech and hearing services in schools**, **27** (4), 367-372.
- Lidz, C.S. & Thomas, C. (1987). The preschool learning assessment device: extension of a static approach. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.288-326). New York: Guilford Press.
- Likelihood function**. (2006). Retrieved 29 January 2006 from <http://en.wikipedia.org/wiki/Likelihood>
- Lillegard, N. (2004). Philosophy of social science. **The internet encyclopaedia of philosophy**. Retrieved October 30, 2005, from <http://www.iep.utm.edu/s/socscien.htm>.
- Lindkvist, K. (1981). Approaches to textual analysis. In K.E. Rosengren (Ed.), **Advances in content analysis** (pp.23-41). Beverly Hills: Sage.
- Lindley, D.V. (1992). Comments on Cox (1957) statistical frontiers. In S. Kotz & N.L. Johnson (Eds.), **Breakthroughs in statistics volume 1: foundations and basic theory** (pp.xxxviii). New York: Springer-Verlag.
- Lindzey, G. (Ed.), (1989). **A history of psychology in autobiography volume 8**. Stanford: Stanford University Press.
- Lipson, L.E. (1992). **Relationship of static and dynamic measures to scholastic achievement of black pupils**. Unpublished M.Ed. dissertation, University of the Witwatersrand.
- Lipton, P.(2001). Inference to the best explanation. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.184-1193). Massachusetts: Blackwell,
- Livio, M. (2003). **The golden ratio: the story of phi, the extraordinary number of nature, art and beauty**. London: Review.
- Lloyd, F. & Pidgeon, D.A. (1961). An investigation into the effects of coaching on non-verbal test material with European, Indian and African children. **The British Journal of educational psychology**, **31**(2), 145-151.
- Lloyd, P. (1995). **Psychology: cognition and language development**. Leicester: British psychological society.
- Locurto, C. (1990). The malleability of IQ as judged from adoption studies. **Intelligence**, **14**(3), 275-292.
- Loehlin, J.C. (1992). Should we do research on race differences in intelligence? **Intelligence**, **16**(1), 1-4.
- Loehlin, J.C. (2004). Group differences in intelligence. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.176-193). **Cambridge: Cambridge University Press**.
- Loehlin, J.C., Horn, J.M. & Willerman, L. (1997). Heredity, environment, and IQ in the Texas adoption project. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.105-125). Cambridge: Cambridge University Press.
- Loftus, G.R. (1996). Psychology will be a much better science when we change the way we analyze data. **Current directions in psychological science**, **5**(6), 161-171.
- Lohman, D.F. (1996). Spatial ability and *g*. In I. Dennis & P. Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.97-116). New Jersey: Lawrence Erlbaum.
- Lohman, D.F. (1997a). Lessons from the history of intelligence testing. **International journal of educational research**, **27**(5), 359-377.
- Lohman, D.F. (1997b). The history of intelligence testing in context: the impact of personal, religious, and scientific beliefs on the development of theories and tests of human abilities. In R.F. Dillon (Ed.), **Handbook on testing** (pp.82-106). Connecticut: Greenwood Press.
- Lohman, D.F. (2005). Reasoning abilities. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.225-250). Cambridge: Cambridge University Press.



- Looren, De Jong, H. (1995). Representations: naturalist versus computationalist. In I. Lubek, R. Van Hezewijk., G. Pheterson & C.W. Tolman (Eds.), **Trends and issues in theoretical psychology** (pp.239-245). New York: Springer.
- Looren De Jong, H. (2003). Evolution and the many faces of psychology. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.416-425). Ontario: Captus University Publications.
- Looren De Jong, H., Bem, S. & Schouten, M. (2004). Theory in psychology: a review essay of Andre Kukla's "Methods of theoretical psychology". **Philosophical psychology**, **17**(2), 275-295.
- Lopes, A., Roodt, G. & Mauer, R. (2001). The predictive validity of the APIL-B in a financial institution. **Journal of industrial psychology**, **27** (1), 61-69.
- Louw, J. (2002). Psychology, history, and society. **South African journal of psychology**, **32**(1), 1 - 8.
- Lord, F.M. & Novick, M.R. (1968). **Statistical theories of mental test scores**. Reading: Addison Wesley.
- Losardo, A. & Notari-Syverson, A. (2001). **Alternative approaches to assessing young children**. Baltimore: Paul H. Brooks.
- Lövdén, M. & Lindenberger, U. (2005). Development of intellectual abilities in old age: from age gradients to individuals. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.203-221). London: Sage.
- Lozano, A.M. & Kalia, S.K. (2005). New movement in Parkinson's. **Scientific American**, **293**(1), 68-75.
- Lubek, I., Van Hezewijk, R., Pheterson, G. & Tolman, C.W. (Eds.), (1995). **Trends and issues in theoretical psychology**. New York: Springer.
- Luce, R.D. (1997). Some unresolved conceptual problems of mathematical psychology. **Journal of mathematical psychology**, **41**, 79-87.
- Luce, R.D. (1999). Where is mathematical modelling in psychology headed? **Theory and psychology**, **9**(6), 723-737.
- Luce, R.D. (2000). Empirical quantification is the key to measurement - including psychological measurement. **Contemporary psychology**, **45**(5), 499-502.
- Luce, R.D. & Tukey, J.W. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. **Journal of mathematical psychology**, **1**, 1-27.
- Ludwig, K. (2003). The mind-body problem: an overview. In S.P.Stich & T.A.Warfield (Eds.), **The Blackwell guide to philosophy of mind** (pp.1-46). Massachusetts: Blackwell.
- Lunz, M.E. & Lincare, J.M. (1998). Measurement designs using multifacet Rasch modelling. In G.A. Marcoulides (Ed.), **Modern methods for business research** (pp.47-77). New Jersey: Lawrence Erlbaum.
- Luria, A.R. (1979). **The making of mind: a personal account of Soviet psychology** (edited by M. Cole & S. Cole). Cambridge: Harvard University Press.
- Luria, A.R. (1994). The problem of the cultural behaviour of the child. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.46-72). Oxford: Blackwell.
- Lussier, C.M. & Swanson, L. (2002). Dynamic assessment: a selective synthesis of the experimental literature. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology** (pp.65-87). Amsterdam: JAI.
- Luther, M., Cole, E. & Gamlin, P.(Eds.), (1996). **Dynamic assessment for instruction: from theory to application**. Ontario: Captus Press.
- Luther, M. & Wyatt, F. (1996). A comparison of Feuerstein's method of (LPAD) assessment with conventional IQ testing on disadvantaged North York high school students. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.168-181). Ontario: Captus Press.
- Luther, M., Wylie, B. & Rosenthal, E. (1996). Feuerstein's thinking skills program with children, adolescents, and adults in "special needs" schools. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.254-269). Ontario: Captus Press.
- Lycan, W.G. (2003). The mind-body problem. In S.P.Stich & T.A. Warfield (Eds.), **The Blackwell guide to philosophy of mind** (pp.47-64). Massachusetts: Blackwell.
- Lykken, D.T. (1991). What's wrong with psychology anyway? In D. Cicchetti & W.M. Grove (Eds.), **Thinking clearly about psychology volume 1: matters of public interest** (pp.3-39). Minneapolis: University of Minnesota.
- Lynn, R. & Vanhanen, T. (2002). **IQ and the wealth of nations**. Connecticut: Praeger.
- Machamer, P.(2002). A brief historical introduction to the philosophy of science. In P.Machamer & M. Silberstein (Eds.), **The Blackwell guide to the philosophy of science** (pp.1-17). Massachusetts: Blackwell.
- Machamer, P.& Silberstein, M. (Eds.), (2002). **The Blackwell guide to the philosophy of science**. Massachusetts: Blackwell.
- Machlup, F. (2000). Are the social sciences really inferior? In J. McErlean (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.263-275). California: Wadsworth.
- MacKay, D.G. (1995). Empirical hypotheses versus theoretical analysis - a much needed distinction in psychology: a response to Kukla. **New ideas in psychology**, **13**(3), 229-235.
- Maddox, J. (1998). **What remains to be discovered: mapping the secrets of the universe, the origins of life. And the future of the human race**. London: Macmillan
- Madsen, K.B. (1968). **Theories of motivation: a comparative study of modern theories of motivation 4th Ed.**. Ohio: Kent State University Press.
- Madsen, K.B. (1971). Humanistic psychology and the philosophy of science. **Journal of humanistic psychology**, **11**(1), 1-10).

- Madsen, K.B. (1973). Theories of motivation. In B. Wolman (Ed.), **Handbook of general psychology** (pp.673-706). New Jersey: Prentice-Hall.
- Madsen, K.B. (1976). Meta-meta-teori: sammenligning af metateorier. **Skolepsykologi**, **13**(5), 367-401.
- Madsen, K.B. (1984a). The hypothesis quotient: a quantitative estimation of the testability of a theory. In J.R. Royce & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 2** (pp.185-202). New York: Plenum Press.
- Madsen, K.B. (1984b). The hypothesis quotient: a reply to commentators. In J.R. Royce & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 2** (pp.235-240). New York: Plenum Press.
- Madsen, K.B. (1985). Psychological metatheory: an introduction to volume 3. In K.B. Madsen & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 3** (pp.1-16). New York: Plenum Press.
- Madsen, K.B. (1987a). Theoretical psychology: a definition and systematic classification. In W.J. Baker., M.E. Hyland., H. Van Rappard. & A.W. Staats (Eds.), **Current issues in theoretical psychology: selected/edited proceedings of the founding conference of the international society for theoretical psychology held in Plymouth, U.K., 30 August - 2 September, 1985** (pp.165-174). Amsterdam: North-Holland.
- Madsen, K.B. (1987b). Systematologi: en sammenlignende videnskabsteori. **Psyke & Logos**, **8**, 261-286.
- Madsen, K.B. (1988). **A history of psychology in metascientific perspective**. Amsterdam: North-Holland.
- Madsen, K.B. (1990). Teoretisk psykologi. **Psyke & Logos**, **11**, 347-351.
- Madsen, K.B. (1992). Psychological metatheory: the historical development and the present situation of psychological theorizing. **Psyke and Logos**, **13**, 497-517.
- Madsen, K.B. & Mos, L.P.(Eds.). (1985). **Annals of theoretical psychology. Volume 3**. New York: Plenum Press.
- Mahoney, M.J. (1989). Participatory epistemology and science. In B. Gholson, W.R. Shadish, R.A. Neimeyer, & A.C Houts, (Eds.). **Psychology of science: contributions to metascience (pp.138-164)**. Cambridge: Cambridge University Press.
- Mahoney, M.S. (1997). Computer science: the search for a mathematical theory. In J. Krige & D. Pestre (Ed.). **Science in the twentieth century** (pp.617-634). Amsterdam: Harwood Academic.
- Maiers, W. (2003). The bogus claim of evolutionary psychology. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.426-435). Ontario: Captus University Publications.
- Majumdar, D. (2004). An axiomatic characterization of Bayes' rule. **Mathematical social sciences**, **47**, 261-273.
- Malofeev, N.N. (1998). Special education in Russia: historical aspects. **Journal of learning disabilities**, **31**(2), 181-185.
- Mandel, R. (2002). The effects of intervention in seriation from the "Bright Start" program on the development of seriation-mathematical thinking of Israeli kindergarteners of Ethiopian and Russian origin. **Journal of cognitive education and psychology**, **2**(2), 180-181.
- Mandler, G. (2002). Origins of the cognitive (r)evolution. **Journal of history of the behavioral sciences**, **38**(4), 339-353.
- Marais, H.C. & Bondesio, M.J. (1996). Research management. In J.G. Garbers (Ed.), **Effective research in the human sciences** (pp.111-140). Pretoria: J.L. Van Schaik.
- Maraun, M.D. (1998). Measurement as a normative practice: implications of Wittgenstein's philosophy for measurement in psychology. **Theory and psychology**, **8**(4), 435-461.
- Marcoulides, G.A. (Ed.), (1998). **Modern methods for business research**. New Jersey: Lawrence Erlbaum.
- Marcoulides, G.A. (1998). Applied generalizability theory models. In G.A. Marcoulides (Ed.), **Modern methods for business research** (pp.1-21). New Jersey: Lawrence Erlbaum.
- Marcoulides, G.A. (1999). Generalizability theory: picking up where the Rasch IRT model leaves off? In S.E. Embretson & S.L. Hershberger (Eds.), **The new rules of measurement: what every psychologist and educator should know** (pp.129-152). New Jersey: Lawrence Erlbaum.
- Marcoulides, G.A. & Drezner, Z. (1997). A method for analysing performance assessments. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.261-277). London: Ablex.
- Maree, D.J.F. (1995). **Cognition and consciousness: developing a conceptual framework**. Unpublished DPhil thesis, University of Pretoria.
- Maree, D.J.F. (1998). **Models used in validity studies: comparison of classical discriminant analysis and a specific neural network** - final internal report. Pretoria: Human Sciences Research Council.
- Maree, D.J.F. (2003, September). **Critical realism, objectivity and psychology**. Paper presented at The Psychological Society of South Africa (PsySSa) Annual Conference, Johannesburg, South Africa.
- Maree, D.J.F. & Maree, M. (2005, August). **Assessment of hope - the process of constructing a gender-sensitive scale for hope within a South African context**. Paper presented at the "Hope: probing the boundaries" Conference, Prague, Czech Republic.
- Mareschal, D. & Schultz, T.R. (1997). From neural constructivism to children's cognitive development: bridging the gap. **Behavioral and brain sciences**, **20**(4), 571-572.
- Martí, E. (1996). Mechanisms of internalisation and externalisation of knowledge in Piaget's and Vygotsky's theories. In A. Tryphon & J. Vonèche (Eds.), **Piaget-Vygotsky: the social genesis of thought** (pp.57-83). Sussex: Psychology Press.
- Martin, J. & VenLehn, K. (1995). A Bayesian approach to cognitive assessment. In P.D. Nichols, S.F. Chipman & R.L. Brennan (Eds.), **Cognitively diagnostic assessment** (pp.141-165). New Jersey: Lawrence Erlbaum.
- Martinez, J.L. & Kesner, R.P. (Eds.), **Neurobiology of learning and memory**. San Diego: Academic Press.
- Mason, B. (2005). Getting the message. **New Scientist**, **187**(2514), 8.

- Maslin, K.T. (2001). **An introduction to the philosophy of mind**. Cambridge: Polity.
- Mastergeorge, A.M. (2001). Guided participation in sociocultural learning: intervention and apprenticeship. **Topics in language disorders**, 22(4), 74-87.
- Matarazzo, J.D. (1992). Biological and physiological correlates of intelligence. **Intelligence**, 16, 257-258.
- Matthews, G., Zeidner, M. & Roberts, R.D. (2005). Emotional intelligence: an elusive ability? In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.79-99). London: Sage.
- Matthews, R. (2006). **Bayesian Critique of Statistics in Health: the great health hoax**. Retrieved 6 January 2006 from <http://www.cs.ucsd.edu/users/goguen/courses/275f00/stat.html>
- Matusov, E. & Hayes, R. (2000). Sociocultural critique of Piaget and Vygotsky. **New ideas in psychology**, 18, 215-239.
- Mautner, T. (Ed.). (2000). **The penguin dictionary of philosophy**. London: Penguin.
- Mautone, P.D. & Mayer, R.E. (2001). Signaling as a cognitive guide in multimedia learning. **Journal of educational psychology**, 93(2), 377-389.
- Maxwell, S.E. & Delaney, H.D. (1985). Measurement and statistics: an examination of construct validity. **Psychological bulletin**, 97(1), 85-93.
- Mayer, R.E. (1997). Multimedia learning: are we asking the right questions? **Educational psychologist**, 32(1), 1-19.
- Mayer, R.E. (2001). **Multimedia learning**. Cambridge: Cambridge University Press.
- Mayer, R.E. (2002). Cognitive theory and the design of multimedia instruction: an example of the two-way street between cognition and instruction. **New directions for teaching and learning**, 89, 55-71.
- Mayer, R.E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. **Learning and instruction**, 13(2), 125-139.
- Mayer, R.E. & Anderson, R.B. (1992). The instructive animation: helping students build connections between words and pictures in multimedia learning. **Journal of educational psychology**, 84(4), 444-452.
- Mayer, R.E. & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. **Journal of educational psychology**, 90(2), 312-320.
- Mayer, R.E. & Moreno, R. (2002a). Aids to computer-based multimedia learning. **Learning and instruction**, 12, 107-119.
- Mayer, R.E. & Moreno, R. (2002b). Animation as an aid to multimedia learning. **Educational psychology review**, 14(1), 87-99.
- Mayer, R.E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. **Educational psychologist**, 38(1), 43-52.
- Mayer, R.E., Dow, G.T. & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: what works in the design of agent-based microworlds? **Journal of educational psychology**, 95(4), 806-813.
- Mayer, R.E., Fennell, S., Farmer, L. & Campbell, J. (2004). A personalization effect in multimedia learning: students learn better when words are in conversational style rather than formal style. **Journal of educational psychology**, 96(2), 389-395.
- Mayer, R.E., Heiser, J. & Lonn, S. (2001). Cognitive constraints on multimedia learning: when presenting more material results in less understanding. **Journal of educational psychology**, 93(1), 187-198.
- Mayer, R.E. & Sims, V.K. (1994). For whom is a picture worth a thousand words? Extensions of dual-coding theory of multimedia learning. **Journal of educational psychology**, 86(3), 389-401.
- Mayer, R.E., Sobko, K. & Mautone, P.D. (2003). Social cues in multimedia learning: role of speakers voice. **Journal of educational psychology**, 95(2), 419-425 .
- Mays, W. (1966). Translator's introduction. In E.W. Beth & J. Piaget. **Mathematical epistemology and psychology** (pp.xiii-xix) (W. Mays, Trans.). Dordrecht: D. Reidel. (Original work published in 1961).
- McAllister, J.W. (2001). Relativism. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.405-407). Massachusetts: Blackwell.
- McArdle, J.J. & Prescott, C.A. (1997). Contemporary models for the biometric genetic analysis of intellectual abilities. D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.403-436). New York: Guilford.
- McArdle, J.J. & Woodcock, R.W. (Eds.), (1998). **Human cognitive abilities in theory and practice**. New Jersey: Lawrence Erlbaum.
- McArthur, D. (1987). Developing computer tools to supports performing and learning complex cognitive skills. In D.E. Berger, K. Pezdek & W. Banks. (Eds.), **Applications of cognitive psychology: problem solving, education and computing** (pp.183-200). New Jersey: Lawrence Erlbaum.
- McCallum, R.S. & Bracken, B.A. (1997). The universal nonverbal intelligence test. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.268-280). New York: Guilford.
- McCauley, R.N. (1998). Levels of explanation and cognitive architectures. In W. Bechtel & G. Graham, (Eds.). **A companion to cognitive science** (pp.611-624). Massachusetts: Blackwell.
- McCauley, R.N. (2001). Explanatory pluralism and the co-evolution of theories in science. In W. Bechtel, P.Mandik, J. Mundale & R.S. Stufflebeam, (Eds.). **Philosophy and the neurosciences: a reader** (pp.431-456). Massachusetts: Blackwell.
- McCollam, K.M. (1998). Latent trait and latent class models. In G.A. Marcoulides (Ed.), **Modern methods for business research** (pp.23-46). New Jersey: Lawrence Erlbaum.
- McCollum, G. (1997). More mathematics: bodily-kinaesthetic intelligence. **Behavioral and brain sciences**, 20(4), 572.

- McCormick, C.B., Miller, G.E. & Pressley, M. (Eds.), (1989). **Cognitive strategy research: from basic research to educational implications**. London: Springer-Verlag.
- McDonald, R.P. (1997). Goodness of approximation in the linear model. In L.L. Harlow, S.A. Mulaik & J.H, Steiger (Eds.), **What if there were no significance tests?** (pp.199-219). New Jersey: Lawrence Erlbaum.
- McDonald, R.P. (1999). **Test theory: a unified treatment**. New Jersey: Lawrence Erlbaum.
- McErlean, J. (Ed.) (2000). **Philosophies of science: from foundations to cotemporary issues**. California: Wadsworth.
- McGrath, R.E. (2005). Rethinking psychosocial constructs: reply to comments by Barrett, Kagan, and Maraun and Peters. **Journal of personality assessment, 85**(2), 141-145.
- McGrew, K.S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive *Gf - Gc* framework. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.151-179). New York: Guilford.
- McGuire, W.J. (1989). Perspectivism in programmatic scientific research. In B. Gholson, W.R. Shadish, R.A. Neimeyer, & A.C Houts, (Eds.). **Psychology of science: contributions to metascience** (pp.214-245). Cambridge: Cambridge University Press.
- McLeish, J. (1975). **Soviet psychology: history, theory, content**. London: Methuen.
- McManus, I.C. & Bryden, M.P. (1994). The neurobiology of handedness, language, and cerebral dominance: a model for the molecular genetics of behaviour. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.679-702). Oxford: Blackwell.
- McMullin, E. (1978). Philosophy of science and its rational reconstructions. In G. Radnitzky & G. Andersson (Eds.), **Progress and rationality in science. Boston studies in the philosophy of science volume 58** (pp.221-252). Dordrecht: D.Reidel.
- McMullin, E. (1990). Can theory appraisal be quantified? **Psychological inquiry, 1**(2), 164-166.
- McNally, R.J. (1992). Disunity in psychology: chaos or speciation? **American psychologist, 47**(8), 1054.
- Mearig, J.S. (1987). Assessing the learning potential of kindergarten and primary-age children. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.237-267). New York: Guilford Press.
- Medvedev, A. (1996). Ethnic psychology in Soviet Russia. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.285-296). London: Greenwood Press.
- Meehl, P.E. (1967). Theory-testing in psychology and physics: a methodological paradox. **Philosophy of science, 34**(2), 103-115.
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. **Journal of consulting and clinical psychology, 46**(4), 806-834.
- Meehl, P.E. (1986). What social scientists don't understand. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.315-338). Chicago: University of Chicago Press.
- Meehl, P.E. (1989). Paul Meehl. In G. Lindzey (Ed.), **A history of psychology in autobiography volume 8** (pp.337-389). Stanford: Stanford University Press.
- Meehl, P.E. (1990). Why summaries on research on psychological theories are often uninterpretable. **Psychological reports, 1**(66), 195-244.
- Meehl, P.E. (1992). Cliometric metatheory: the actuarial approach to empirical, history-based philosophy of science. **Psychological reports, 71**, 339-467.
- Meehl, P.E. (1997). The problem is epistemology, not statistics: replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L.L. Harlow, S.A. Mulaik & J,H, Steiger (Eds.), **What if there were no significance tests?** (pp.393-425). New Jersey: Lawrence Erlbaum.
- Meehl, P.E. (1998, May). **The power of quantitative thinking**. Speech delivered upon receipt of the James McKeen Cattell Fellow award at the meeting of the American Psychological Society, Washington, D.C., United States of America.
- Meehl, P.E. (2002). Cliometric metatheory: II. Criteria scientists use in theory appraisal and why it is rational to do so. **Psychological reports, 91**, 339-404.
- Meier, S.T. (1994). **The chronic crisis in psychological measurement and assessment: a historical survey**. California: Academic Press.
- Meijer, J. (1993). Learning potential, personality characteristics and test performance. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.341-362). Amsterdam: Swets & Zeitlinger.
- Meijer, J. & Elshout, J.J. (2001). The predictive and discriminant validity of the zone of proximal development. **British journal of educational psychology, 71**(1), 93-113.
- Meijer, J., Oostdam, R. & Van der Sluis, S. (2002). A feasibility study into incremental predictive validity of dynamic assessment of verbal-reasoning ability. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.273-291). Amsterdam: JAI.
- Menchinskaya, N.A. (1972). Fifty years of the Soviet Psychology of learning. In J. Brožek & D.I. Slobin (Eds.), **Psychology in the USSR: An historical perspective** (pp.181-196). New York: IASP.

- Mendelsohn, P. (1996). Mapping models of cognitive development to design principles of learning environments. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), **International perspectives on the design of technology-supported learning environments** (pp.323-344). New Jersey: Lawrence Erlbaum.
- Meng, H, Smith, S.D., Hager, K., Held, M., Liu, J., Olson, R.K., Penington, B.F., DeFries, J.C., Gelernter, J., O'Reilly-Pol, T., Somlo, S., Skudlarski, P., Shaywitz, S.E., Shaywitz, B.A., Marchione, K., Wang, Y., Paramasivam, M., LoTurco, J.L., Page, G.P. & Gruen, J.R. (2005). DCDC2 is associated with reading disability and modulates neuronal development in the brain. **Proceedings of the national academy of sciences**, **102**(47), 17053-17058.
- Menzel, R. & Bicker, G. (1987). Plasticity in neuronal circuits and assemblies of invertebrates. In J-P.Changeux & M. Konishi (Eds.), **The neural and molecular bases of learning** (pp.433-472). Chichester: John Wiley.
- Merzenich, M.M. (1987). Dynamic neocortical process and the origins of higher brain functions. In J-P.Changeux & M. Konishi (Eds.), **The neural and molecular bases of learning** (pp.337-358). Chichester: John Wiley.
- Meshi, D., Drew, M.R., Saxe M., Ansorge, M.S., David, D., Santarelli, L., Malapani, C., Moore, H. & Hen, R. (2006). Hippocampal neurogenesis is not required for behavioral effects of environmental enrichment. **Nature Neuroscience**, **9**, 729 - 731.
- Messerer, J., Hunt, E., Meyers, G. & Lerner, J. (1984). Feuerstein's instrumental enrichment: a new approach
- Messick, S. (1996). Human abilities and modes of attention: the issues of stylistic consistencies in cognition. In I. Dennis & P.Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.77-96). New Jersey: Lawrence Erlbaum.
- Meta Analysis Unit (The), Universidad De Murcia.** (2005). Retrieved April 13 2005 from www.um.es/facpsi/metaanalysis/software.php
- Metaxis (Vers. 1), Update Software Ltd.** (2005). Retrieved April 13 2005 from www.update-software.com/metaxis/metaxis-frame.html
- Meyer, J.H.F., Cliff, A.F. & Dunne, T.T. (1994). Impressions of disadvantage: II - Monitoring and assisting the student at risk. **Higher education**, **27**(1), 95-117.
- Meyers, J. (1987). The training of dynamic assessors. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.403-425). New York: Guilford Press.
- Michalos, A.C. (1971). Theory appraisal and the growth of scientific knowledge. **Studies in history and philosophy of science**, **1**(4), 353-362.
- Michell, J. (1990). **An introduction to the logic of psychological measurement**. New Jersey: Erlbaum.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. **British journal of psychology**, **88**, 355-383.
- Michell, J. (1999). **Measurement in psychology: critical history of a methodological concept**. Cambridge: Cambridge University Press.
- Michell, J. (2000). Normal science, pathological science and psychometrics. **Theory and psychology**, **10**(5), 639-667.
- Michell, J. (2001). Teaching and misteaching measurement in psychology. **Australian psychologist**, **36**(3), 211-217.
- Michell, J. (2002). Stevens's theory of scales of measurement and its place in modern psychology. **American journal of psychology**, **54**(2), 99-104.
- Michell, J. (2003a, May). **The relation between the Rasch model and additive conjoint measurement**. Paper presented at the colloquium of Sciences sociales et modes De mesure: alternatives contemporaines, Paris, France.
- Michell, J. (2003b). The quantitative imperative: positivism, naïve realism and the place of qualitative methods in psychology. **Theory and psychology**, **13**(1), 5-31.
- Michell, J. (2004). Item response models, pathological science and the shape of error. **Theory and psychology**, **14**(1), 121-129.
- Michell, J. (2005). The meaning of the quantitative imperative: a response to Niaz. **Theory and psychology**, **15**(2), 257-263.
- Midgley, M. (1999). One world, but a big one. In S. Rose (Ed.). **From brains to consciousness: essays on the new sciences of the mind** (pp.246-270). London: Penguin.
- Midgley, M. (2000). Why memes? In H. Rose & S. Rose (Eds.). **Alas, poor Darwin: arguments against evolutionary psychology** (pp.67-84). London: Jonathan Cape.
- Millar, D.; Millar, I.; Millar, J. & Millar, M. (1996). **The Cambridge dictionary of scientists**. Cambridge: Cambridge University Press.
- Miller, A.I. (2002). Erotica, aesthetics and Schrödinger's wave equation. In G. Farmelo (Ed.). **It must be beautiful: great equations of modern science** (pp.80-101). London: Granta.
- Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. **Psychological review**, **63**(2), 81-97.
- Miller, G.A. (1985). The constitutive problem of psychology. In S. Koch & D.E. Leary (Eds.). **A century of psychology as science** (pp.40-45). New York: McGraw-Hill.
- Miller, J.G. (1997). A cultural-psychology perspective on intelligence. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.269-302). Cambridge: Cambridge University Press.
- Miller, L.T. (1990). Psychometric and information processing approaches to measuring cognitive abilities: paradigms in military testing. **Canadian psychology**, **40**(3), 241-254.
- Miller, L.T. & Vernon, P.A. (1992). The general factor in short-term memory, intelligence, and reaction time. **Intelligence**, **16**, 5-29.
- Miller, M.B. (2003a). The meaning of mediation: an introduction. **Journal of cognitive education and psychology**, **3**(1), 27-28.

- Miller, M.B. (2003b). The meaning of mediation: discussion of varying perspectives. **Journal of cognitive education and psychology**, *3*(1), 82-89.
- Miller, M.B. & Yager, H. (2001). Homage to Reuven Feuerstein: a review of Kozulin and Rand's edited volume. Experience of mediated learning: an impact of Feuerstein's theory on education and psychology. **Journal of cognitive education and psychology**, *2*(1), 88-91.
- Miller, P.H. (1989). **Theories of developmental psychology**. New York: W.H. Freeman.
- Miller, R. (1989). Conceptual issues in theorising about cognition. **South African journal of higher education**, *3*(1), 154-160.
- Miller, R. (1998). A follow-up study of the academic performance of English first and second language students. **South African journal of higher education**, *12*(2), 167-175.
- Miller, R. & Bradbury, J. (1999). Academic performance of English first and second language students: selection criteria. **South African journal of science**, *95*(1), 30-34.
- Miller, W.G. (2004). Statistics and measurement: using the free OpenStat package. Retrieved February 5 2006 from <http://www.statpages.org/miller/openstat/>
- Millstein, R. L. (2002). Evolution. In P.Machamer & M. Silberstein (Eds.), **The Blackwell guide to the philosophy of science** (pp.227-251). Massachusetts: Blackwell.
- Minick, N. (1987). Implications of Vygotsky's theories for dynamic assessment. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.116-140). New York: The Guilford Press.
- Minnaert, A. (2002). Alternative assessment of students' domain-specific learning competencies in the transition of secondary to higher education. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.335-351). Amsterdam: JAI.
- Minnaert, A. & Janssen, P.J. (1996). How general are the effects of domain-specific prior knowledge on study expertise as compared to general thinking skills? In M. Birenbaum & F.J.R.C. Dochy (Eds.), **Alternatives in assessment of achievements, learning processes and prior knowledge** (pp.265-281). Boston: Kluwer.
- Minsky, M. (1988). **The society of mind**. New York: Simon and Schuster.
- Misak, C. (2001). Peirce. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.335-339). Massachusetts: Blackwell.
- Mislevy, R.J. (1995). Probability-based inference in cognitive diagnosis. In P.D. Nichols, S.F. Chipman & R.L. Brennan (Eds.), **Cognitively diagnostic assessment** (pp.43-71). New Jersey: Lawrence Erlbaum.
- Missiuna, C. (1996). Assessing preschool children: application of dynamic assessment techniques in the classroom. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.158-167). Ontario: Captus Press.
- Molenaar, P.C.M. & Raijmakers, M.E.J. (2000). A causal interpretation of Piaget's theory of cognitive development: reflections on the relationship between epigenesis and nonlinear dynamics. **New ideas in psychology**, *18*, 41-55.
- Moll, I. (1981). What is mediation and do we need it? A reply to Tony Morphet. In W. Morrow (Ed.), **The proceedings of the Kenton conference 1985** (pp.48-51). Bellville: University of the Western Cape.
- Moll, I. (1989). Roots and disputes of cognitive developmental conceptions of teaching. **South African journal of education**, *9*(4), 714-721.
- Moll, I. & Slonimsky, L. (1989). Towards an understanding of cognition and learning in the academic support context. **South African journal of higher education**, *3*(1), 160-166.
- Montuschi, E. (2001). Metaphor in science. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.277-282). Massachusetts: Blackwell.
- Moore, D.S. (1997). Bayes for beginners? Some reasons to hesitate. **The American statistician**, *51*, 254-261.
- Moreno, R. & Mayer, R.E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. **Journal of educational psychology**, *91*(2), 358-368.
- Moreno, R. & Mayer, R.E. (2000). A coherence effect in multimedia learning: the case for minimizing irrelevant sounds in the design of multimedia instructional messages. **Journal of educational psychology**, *92*(1), 117-125.
- Moreno, R. & Mayer, R.E. (2002). Verbal redundancy in multimedia learning: when reading helps listening. **Journal of educational psychology**, *94*(1), 156-163.
- Morgan, H. (1996). An analysis of Gardner's theory of multiple intelligence. **Roeper review**, *18*(4), 263-269.
- Moring, G.F. (2002). **The complete idiot's guide to theories of the universe**. Indianapolis: Alpha.
- Morphet, T. (1986). Who is Feuerstein and do we need him? In W. Morrow (Ed.), **The proceedings of the Kenton conference 1985** (pp.29-47). Bellville: University of the Western Cape.
- Morrison, T. (2001). **Actionable learning: a handbook for capacity building through case based learning**. Tokyo: Asian Development Bank Institute.
- Morrow, W. (Ed.), (1986). **The proceedings of the Kenton conference 1985**. Bellville: University of the Western Cape.
- Mouton, J. (1993a). Positivism. In J. Snyman (Ed.), **Conceptions of social inquiry** (pp.1-28). Pretoria: Human Sciences Research Council.
- Mouton, J. (1993b). Scientific realism and realism in the social sciences. In J. Snyman (Ed.), **Conceptions of social inquiry** (pp.219-239). Pretoria: Human Sciences Research Council.

- Mouton, J. & Marais, H.C. (1988). **Basic concepts in the methodology of the social sciences**. Pretoria: Human Sciences Research Council.
- Mpofu, E. (2004). Being intelligent with Zimbabweans: a historical and contemporary view. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.364-390). Cambridge: Cambridge University Press.
- Muir, H. (2005). Particle smasher gets a super-brain. **New Scientist**, **86** (2500), 10-11.
- Muir, H. (2006). The cosmic controller. **New Scientist**, **190**(2556), 47-49.
- Mulaik, S.A. (1988). Confirmatory factor analysis. In J.R. Nesselroade & R.B. Cattell (Eds.), **Handbook of multivariate experimental psychology** (pp.259-288). New York: Plenum.
- Mulaik, S.A., Raju, N.S. & Harshman, R.A. (1997). There is a time and a place for significance testing. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.65-115). New Jersey: Lawrence Erlbaum.
- Mullins, J. (2005). Whatever happened to machines that think? **New Scientist**, **186**(2), 32-37.
- Mundale, J. (2001). Neuroanatomical foundations of cognition: connecting the neuronal level with the study of higher brain areas. In W. Bechtel, P.Mandik, J. Mundale & R.S. Stoffelbeam, (Eds.). **Philosophy and the neurosciences: a reader** (pp.37-54). Massachusetts: Blackwell.
- Murphy, K.R. & Davidshofer, C.O. (1998). **Psychological testing: principles and applications**. New Jersey: Prentice Hall.
- Murphy, R. (2002). **A review of South African research in the field of dynamic assessment**. Unpublished MA dissertation. University of Pretoria. (available online from: <http://upetd.up.ac.za/thesis/available/etd-05042002-161239/>)
- Murphy, R. & Maree, D.J.F. (2006). A review of South African research in the field of dynamic assessment. **South African journal of psychology**, **36**(1), 1038-1061.
- Murray, C. (2004). **Human accomplishment: the pursuit of excellence in the arts and sciences 800BC to 1950**. New York: Perennial.
- Murray, D.A. (1988). **Effectiveness of Feuerstein's learning potential assessment device in a South African context**. Unpublished M.Ed. dissertation, University of the Witwatersrand.
- Musgrave, A. (1985). Realism versus constructive empiricism. In P.M. Churchland & C.A. Hooker (Eds.), **Images of science: essays on realism and empiricism, with a reply from Bas C. Van Fraassen** (pp.197-221). Chicago: The university of Chicago Press.
- Myers, D.G. (1991). Union is strength: a consumer's view of meta-analysis. **Personality and social psychology bulletin**, **17**(3), 265-266.
- Mynhardt, G.L. (1995). **The "social" in cognitive development: a clarification of L.S. Vygotsky's theory and its implications for education**. Unpublished M.Ed. dissertation, University of the Witwatersrand.
- Nagel, E. (1979). **The structure of science: problems in the logic of scientific explanation**. Oxford: Routledge.
- Nagel, E. (2000). The instrumentalist view of theories. In J. McErlean (Ed.), **Philosophies of science** (pp.88-95). California: Wadsworth.
- Nagel, T. (1981). What is it like to be a bat? In D.R. Hofstadter & D.C. Dennett (Eds.), **The mind's I: fantasies and reflections on self and soul** (pp.391-403). London: Penguin.
- Naglieri, J.A. (1997). Planning, attention, simultaneous, and successive theory and the cognitive assessment system: a new theory-based measure of intelligence. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.247-267). New York: Guilford.
- Narens, L. & Luce, R.D. (1986). Measurement: the theory of numerical assignment. **Psychological bulletin**, **99**(2), 166-180.
- Narrol, H.G. (1996). Computers, thinking, and education. In M. Luther, E. Cole & P.Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.221-224). Ontario: Captus Press.
- Necka, E. & Orzechowski, J. (2005). Higher-order cognition and intelligence. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.122-141). Cambridge: Cambridge University Press.
- Neisser, U. (1979). The concept of intelligence. **Intelligence**, **3**, 217-227.
- Neisser, U. (1997). The future of cognitive science: an ecological analysis. In D.M. Johnson & C.E. Erneling (Eds.), **The future of the cognitive revolution** (pp.247-260). New York: Oxford.
- Neisser, U. (1998). Introduction: rising test scores and what they mean. In U. Neisser (Ed.), **The rising curve: long-term gains in IQ and related measures** (pp.3-22). Washington: American Psychological Association.
- Neisser, U. (Ed.), (1998). **The rising curve: long-term gains in IQ and related measures**. Washington: American Psychological Association.
- Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J. & Urbina, S. (1996). Intelligence: knowns and unknowns. **American psychologist**, **51** (2), 77-101.
- Neisworth, J.T. & Bagnato, S.J. (1992). The case against intelligence testing in early intervention. **Topics in early childhood special education**, **12**(1), 1-20.
- Nel, A. (1997). **Die voorspelling Van akademiese sukses binne konteks Van 'n alternatiewe universiteitstoelatingsbeleid**. Unpublished M.A. dissertation, Rand Afrikaans University.
- Nelkin, D. (2000). Less selfish than sacred? Genes and the religious impulse in evolutionary psychology. In H. Rose & S. Rose (Eds.). **Alas, poor Darwin: arguments against evolutionary psychology** (pp.14-27). London: Jonathan Cape.
- Nelson, T.O. (1998). Foreward: metacognitive food for thought in educational theory and practice. In D.K. Hacker, J. Dunlosky & A.C. Graesser (Eds.), **Metacognition in educational theory and practice** (pp.ix-xii). New Jersey: Lawrence Erlbaum.



- Nesselroade, J.R. & Cattell, R.B. (Eds.), (1988). **Handbook of multivariate experimental psychology**. New York: Plenum.
- Nester, M.R. (1996). An applied statistician's creed. **Applied statistics**, **45**(4), 401-410.
- Netchine-Grynberg, G. (1995). The functionality of cognition according to Cassirer, Meyerson, Vygotsky, and Wallon: toward the roots of the concept of cognitive tool. In I. Lubek, R. Van Hezewijk., G. Pheterson & C.W. Tolman (Eds.), **Trends and issues in theoretical psychology** (pp.207-213). New York: Springer.
- Nettelbeck, T. (1999). Savant syndrome - rhyme without reason. In M. Anderson (Ed.), **The development of intelligence** (pp.247-273). Sussex: Psychology Press.
- Neubauer, A.C. & Fink, A. (2005). Basic information processing and the psychophysiology of intelligence. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.68-87). Cambridge: Cambridge University Press.
- Neufeld, R.W.J. (2002). Trajectory of cognitive science and psychological assessment: current status and expectations. **Psychological assessment**, **14**(3), 299-303.
- Neuman, W.L. (1997). **Social research methods: qualitative and quantitative methods**. Boston: Allyn and Bacon.
- Neville, H.J. (1994). Neurobiology of cognitive and language processing: effects of early experience. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.424-448). Oxford: Blackwell.
- Newell, A. (1990). **Unified theories of cognition**. London: Harvard University Press.
- Newman, S.D. & Just, M.A. (2005). The neural basis of intelligence: a perspective based on functional neuroimaging. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.88-103). Cambridge: Cambridge University Press.
- Newton-Smith, W.H. (1981). **The rationality of science**. Boston: Routledge and Kegan Paul.
- Newton-Smith, W.H. (2001a). Hume. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.165-168). Massachusetts: Blackwell.
- Newton-Smith, W.H. (2001b). Introduction. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.1-8). Massachusetts: Blackwell.
- Newton-Smith, W.H. (Ed.). (2001c). **A companion to the philosophy of science**. Massachusetts: Blackwell.
- Neyman, J. & Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. **Philosophical transactions of the Royal Society series A**, **23**(1), 289-337.
- Niaz, M. (1998). Potential for learning as an antecedent variable in Pascual-Leone's theory of constructive operators. **Perceptual and motor skills**, **87**, 381-382.
- Niaz, M. (2001). Are neo-Piagetian and neo-Vygotskian interpretations of potential for learning complementary? **Perceptual and motor skills**, **93**, 416-418.
- Niaz, M. (2005). The quantitative imperative vs. the imperative of supposition. **Theory and psychology**, **15**(2), 247-256.
- Nichols, P.D., Chipman, S.F. & Brennan, R.L. (Eds.), (1995). **Cognitively diagnostic assessment**. New Jersey: Lawrence Erlbaum.
- Nickerson, R.S. (2000). Null hypothesis significance testing: a review of an old an continuing controversy. **Psychological methods**, **5**(2), 241-301.
- Nickles, T. (2001). Lakatos. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.207-212). Massachusetts: Blackwell,
- Nicolopoulou, A. (1993). Play, cognitive development, and the social world: Piaget, Vygotsky, and beyond. **Human development**, **36**(1), 1-23.
- Nightingale, D. & Neilands, T. (1997). Understanding and practising critical psychology. In D. Fox & I. Prilleltensky (Eds.), **Critical psychology: an introduction** (pp.68-84). London: Sage.
- Nisbett, R.E. (2003). **The geography of thought: how Asians and Westerners think differently and why**. New York: The Free Press.
- No author. (2005). 60 seconds. **New scientist**, **118**(2522), 7.
- Norman, D.A. (1988). Reflections on cognition and parallel distributed processing. In J.L. McClelland & D.E. Rumelhart (Eds.), **Parallel distributed processing** (pp.531-546). Cambridge: MIT.
- Normand, S.T. (1995). Meta-analysis software: a comparative review. **The American statistician**, **49**(3), 298-309.7
- Noskova, O.G. (1996). A social history of Russian industrial psychology in the 1920s and 1930's. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.267-284). London: Greenwood Press.
- Nowakowski, R.S. (1994). Basic concepts of CNS development. In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp 54-92). Oxford: Blackwell.
- Nuckolls, C.W. (1998). Cognitive anthropology. In W. Bechtel & G. Graham, (Eds.). **A companion to cognitive science** (pp.140-145). Massachusetts: Blackwell.
- Nunnally, J.C. (1978). **Psychometric testing**. New York: McGraw-Hill.
- Nunns, C. & Ortlepp, K. (1994). Exploring predictors of academic success in Psychology 1 At Wits university as an important component of fair student selection. **South African journal of psychology**, **24** (4), 201-207.
- Oakes, M. (1986). **Statistical inference: a commentary for the social and behavioural sciences**. John Wiley: Chichester.

- Oakland, T. (1980). An evaluation of the ABIC, pluralistic norms, and estimated learning potential. **Journal of school psychology**, **18**(1), 3-11.
- Oberhauer, K. (2005). The measurement of working memory capacity. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.393-407). London: Sage.
- Ochsner, K.N. & Kosslyn, S.M. (1999). The cognitive neuroscience approach. In B.M. Bly & D.E. Rumelhart (Eds.), **Cognitive science: handbook of perception and cognition** (pp.319-365). London: Academic Press.
- O'Connor, N. (Ed.), (1961). **Recent Soviet psychology**. Oxford: Pergamon.
- O'Connor, N. (1961). Introduction. In N. O'Connor (Ed.), **Recent Soviet psychology** (pp.7-20). Oxford: Pergamon.
- O'Connor, N. (Ed.), (1966). **Present-day Russian psychology: a symposium by seven authors**. Oxford: Pergamon.
- O'Connor, N. (1966). Introduction. In N. O'Connor (Ed.), **Present-day Russian psychology: a symposium by seven authors** (pp.vii-xiv). Oxford: Pergamon.
- Oldroyd, D. (1986). **The arch of knowledge: an introductory study of the history of the philosophy and methodology of science**. New York: Methuen.
- Oleinik, Y.N. (1996). Russian psychology at present: trends and paradoxes. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.53-59). London: Greenwood Press.
- Omnès, R. (2005). **Converging realities: toward a common philosophy of physics and mathematics**. Princeton: Princeton University Press.
- Ones, D.S., Viswesvaran, C. & Dilchert, S. (2005). Cognitive ability in selection decisions. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.431-468). London: Sage.
- Ortar, G. (1972). Some principles for adaptation of psychological tests. In L.J. Cronbach & P.J.D. Drenth (Eds.), **Mental tests and cultural adaptation** (pp.111-120). The Hague: Mouton.
- Orwin, R.G. (1994). Evaluating coding decisions. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.139-162). New York: Russell Sage Foundation.
- Osborne, H. (2000). Interpretation in science and in art. In J. McErlan (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.341-345). California: Wadsworth.
- Outhwaite, W. (1983). **Concept formation in social science**. London: Routledge.
- Outhwaite, W. (1987). **New philosophies of social science: realism, hermeneutics and critical theory**. New York: St Martin's Press.
- Owen, K. (1992). **Test-item bias: methods, findings and recommendations**. Pretoria: Human Sciences Research Council.
- Owen, K. (1998). **The role of psychological tests in education in South Africa: issues, controversies and benefits**. Pretoria: Human Sciences Research Council.
- Oyama, S. (1994). The problem of change. In M.H. Johnson, (Ed) **Brain development and cognition: a reader** (pp 19-30). Oxford: Blackwell.
- Paour, J-L. (1992). Logic structures in the mentally retarded: an assessment and intervention instrument. In H.C. Haywood & D. Tzuril (Eds.), **Interactive assessment** (pp.119-166). New York: Springer-Verlag.
- Paour, J-L. & Cèbe, S. (2002). Preschool cognitive education: why, how, with what effects? **Journal of cognitive education and psychology**, **2**(2), 141-167.
- Paour, J-L. & Soavi, G. (1992). A case study in the induction of logic structures. In H.C. Haywood & D. Tzuril (Eds.), **Interactive assessment** (pp.419-442). New York: Springer-Verlag.
- Parent, S., Normandeau, S. & Larivée, S. (2000). A quest for the holy grail in the new millennium: in search of a unified theory of cognitive development. **Child development**, **71**(4), 860-861.
- Parisi, D. (1996). Computation models of developmental mechanisms. In R. Gelman & T.K-F. Au (Eds.), **Perceptual and cognitive development** (pp.373-412). San Diego: Academic Press.
- Parsons, C. (1983). What is the iterative conception of set? In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.503-529). Cambridge: Cambridge University Press.
- Pascoe, L.C. (1992). **Teach yourself mathematics**. London: Hodder & Stoughton.
- Pascual-Leone, J. (1997). Metasubjective Processes: the missing lingua franca of cognitive science. In D.M. Johnson & C.E. Erneling (Eds.), **The future of the cognitive revolution** (pp.75-101). New York: Oxford.
- Pascual-Leone, J. & Johnson, J. (2005). A dialectical constructivist view of developmental intelligence. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.177-201). London: Sage.
- Pascual-Leone, J., Johnson, J., Baskind, S., Dworsky, S. & Severson, E. (2000). Culture-fair assessment and the processes of mental attention. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.191-214). Amsterdam: Pergamon.
- Pascual-Leone, J. & Sparkman, E. (1980). The dialectics of empiricism and rationalism: a last methodological reply to Trabasso. **Journal of experimental child psychology**, **29**, 88-101.
- Pauw, J.C. (1993). Peter Winch: the idea of a social science and its relation to philosophy. In J. Snyman (Ed), **Conceptions of social inquiry** (pp.77-85). Pretoria: Human Sciences Research Council.
- Pavitt, K. (1997). Academic research, technical change and government policy. In J. Krige & D. Pestre (Eds.), **Science in the twentieth century** (pp.143-158). Amsterdam: Harwood Academic.
- Pawlowski, T. (1980). **Concept formation in the humanities and the social sciences**. Dordrecht: D. Reidel.



- Payne, G. & Payne, J. (2004). **Key concepts in social research**. London: Sage.
- Pearl, A. (1997). Cultural and accumulated environmental deficit models. In R.R. Valencia (Ed.), **The evolution of deficit thinking: educational thought and practice** (pp.132-159). London: Falmer Press.
- Pearson, E.S. & Kendall, M.G. (Eds.). (1978). **Studies in the history of statistics and probability volume 1**. London: Charles Griffin.
- Pellegrino, J.W. (1998). Intelligence: the interaction of culture and cognitive processes. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.113-116). New Jersey: Ablex.
- Pena, E.D. (2000). Measurement of modifiability on children from culturally and linguistically diverse backgrounds. **Communication disorders quarterly**, 21(2), 87 - 97.
- Pena, E.D. & Gillam, R.B. (2000). Dynamic assessment of children referred for speech and language evaluations. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.543-575). New York: Elsevier.
- Pena, E.D., Iglesias, A. & Lidz, C.S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. **American journal of speech-language pathology**, 10, 138 - 154.
- Pennings, A.H. & Verhelst, N.D. (1993). The dynamic testing of restructuring tendency. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.195-212). Amsterdam: Swets & Zeitlinger.
- Penrose, R. (1994). **Shadows of the mind: a search for the missing science of consciousness**. Oxford: Oxford University Press.
- Penrose, R. (1999). **The emperor's new mind: concerning computers, minds and the laws of physics**. Oxford: Oxford University Press.
- Penrose, R. (2002). The rediscovery of gravity: the Einstein equation of general relativity. In G. Farmelo (Ed.) **It must be beautiful: great equations of modern science** (pp.47-79). London: Granta.
- Penrose, R. (2005). **The road to reality: a complete guide to the laws of the universe**. London: Vintage
- Percival, P.(2001). Probability. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.358-372). Massachusetts: Blackwell.
- Perline, R., Wright, B.D. & Wainer, H. (1979). The Rasch model as additive conjoint measurement. **Applied psychological measurement**, 3(2), 237-255.
- Perret-Clermont, A-N. (2000). Revisiting young Jean Piaget in Neuchâtel among his partners in learning. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.91-120). London: Routledge.
- Perret-Clermont & Bell, N. (1987). Learning processes in social and instructional interactions. In E. De Corte, H. Lodewijks, R. Parmentier & P.Span (Eds.), **Learning and instruction: European research in an international context volume 1** (pp.251-257). Oxford: Pergamon
- Pestre, D. (1997). Science, political power and the state. In J. Krige & D. Pestre (Eds.), **Science in the twentieth century** (pp.61-75). Amsterdam: Harwood Academic.
- Petersen, A. (1969). **Quantum physics and the philosophical tradition**. Massachusetts: MIT Pres.
- Petrides, M., Cadoret, G. & Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. **Nature**, 435(7046), 1235-1238.
- Petrill, S.A. (2005). Behavioral genetics and intelligence. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.165-176). London: Sage.
- Pfanzagl, J. (1994). On item parameter estimation in certain latent trait models. In G.H Fischer & D. Laming (Eds.), **Contributions to mathematical psychology, psychometrics, and methodology** (pp.249-263). NewYork: Springer-Verlag.
- Phillips, H. (2005). How life shapes the brainscape. **New scientist**, 188(2527), 12-13.
- Piaget, J. (1994). The epigenetic system and the development of cognitive functions. In M.H. Johnson, (Ed) **Brain development and cognition: a reader** (pp.31-38). Oxford: Blackwell.
- Pickover, C.A. (1999). **Strange brains and genius: the secret lives of eccentric scientists and madmen**. New York: Quill.
- Pillemer, D.B. & White, S.H. (Ed.), (2005). **Developmental psychology and social change: research, history, and policy**. Cambridge: Cambridge University Press.
- Pinker, S. (1997). **How the mind works**. London: Penguin.
- Pintner, R. (1921). Intelligence and its measurement. **Journal of educational psychology**, 12(3), 139-143.
- Plass, J.L., Chun, D.M., Mayer, R.E. & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. **Journal of educational psychology**, 90(1), 25-36.
- Plomin, R. (1989). Environment and genes. **American psychologist**, 44(2), 105-111.
- Plomin, R. (1997). Identifying genes for cognitive abilities and disabilities. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.89-104). Cambridge: Cambridge University Press.
- Plomin, R., DeFries, J.C., McClearn, G.E. & McGuffin, P.(2001). **Behavioral genetics**. New York: Worth.
- Plomin, R. & Neiderhiser, J. (1991). Quantitative genetics, molecular genetics and intelligence, **Intelligence**, 15(4), 369-387.
- Plomin, R. & Petrill, S.A. (1997). Genetics and intelligence: what's new? **Intelligence**, 24(1), 53-77.
- Plotkin, H. (1997). **Evolution in mind: an introduction to evolutionary psychology**. London: Penguin.

- Plotkin, H. (2003). **The imagined world made real: towards a natural science of culture**. London: Penguin.
- Poincaré, H. (1983). On the nature of mathematical reasoning. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.394-402). Cambridge: Cambridge University Press.
- Ponocny, I. & Ponocny-Seliger, E. (1997). Applications of the program L_{PC}M in the field of measuring change. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.209-222). London: Ablex.
- Poortinga, Y.H. (1991). Conceptual implications of item bias. In S.H. Irvine & S.E. Newstead (Eds.), **Intelligence and cognition: contemporary frames of reference** (pp.279-290). Dordrecht: Martinus Nijhoff.
- Popoff-Walker, L.E. (1982). IQ, SES, adaptive behavior, and performance on a learning potential measure. **Journal of school psychology, 20**(3), 222-231.
- Popper, K. (1965). Unity of method in the natural and social sciences. In D. Braybrooke (Ed.). **Philosophical problems of the social sciences** (pp.32-41). New York: Macmillan.
- Popper, K. (1993). Three views concerning human knowledge. In J.H. Fetzer (Ed.). **Foundations of philosophy of science: recent developments** (pp.421-440). New York: Paragon.
- Porges, S.W. (1998). Introduction of E.N. Sokolov. In J.J. McArdle & R.W. Woodcock (Eds.), **Human cognitive abilities in theory and practice** (pp.39-44). New Jersey: Lawrence Erlbaum.
- Porter, T.M. (1997). The management of society by numbers. In J. Krige & D. Pestre (Ed.). **Science in the twentieth century** (pp.97-110). Amsterdam: Harwood Academic.
- Porter, R. (2003). **Blood and guts: a short history of medicine**. London: Penguin.
- Porter, R. (2004). **Flesh in the age of reason: how the enlightenment transformed the way we see our bodies and souls**. London: Penguin.
- Portes, P.R. & Vadeboncouer, J.A. (2003). Mediation in cognitive socialization: the influence of socio-economic status. In A. Kozulin, B. Gindis, V.S.Ageyev & S.M. Miller (Eds.), **Vygotsky's educational theory in cultural context** (pp.371-449). Cambridge: Cambridge University Press.
- Posner, M.I. & DiGirolamo, G.J. (2000). Cognitive neuroscience: origins and promise. **Psychological bulletin, 126**(6), 873-889.
- Potter, J. (2000). Post-cognitive psychology. **Theory and psychology, 10**(1), 31-37.
- Preece, P.F.W. & Read, K.L.Q. (1995). A stage-theory model of cognitive development. **British journal of mathematical and statistical psychology, 48**(1), 1-7.
- Pressley, M., Van Etten, S., Yokoi, L., Freebern, G. & Van Meter, P. (1998). The metacognition of college studentship: a grounded theory approach. In D.K. Hacker, J. Dunlosky & A.C. Graesser (Eds.), **Metacognition in educational theory and practice** (pp.347-366). New Jersey: Lawrence Erlbaum.
- Preston, J. (2001). Feyerband. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.143-148). Massachusetts: Blackwell.
- Pretz, J.E. & Sternberg, R.J. (2005). Unifying the field: cognition and intelligence. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.306-318). Cambridge: Cambridge University Press.
- Pribram, K.H. (1990). From metaphors to models: the use of analogy in neuropsychology. In D.E. Leary (Ed). **Metaphors in the history of psychology** (pp.79-103) Cambridge: Cambridge University Press.
- Prilleltensky, I. & Fox, D. (1997). Introducing critical psychology: values, assumptions, and the status quo. In D. Fox & I. Prilleltensky (Eds.), **Critical psychology: an introduction** (pp.3-20). London: Sage.
- Pruzek, R.M. (1997). An introduction to Bayesian inference and its applications. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.287-318). New Jersey: Lawrence Erlbaum.
- Purves, D. (1997). Neural construction: two and a half cheers for Quartz & Sejnowski! **Behavioral and brain sciences, 20**(4), 573.
- Putnam, H. (1983a). Mathematics without foundations. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.295-311). Cambridge: Cambridge University Press.
- Putnam, H. (1983b). Models and reality. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.421-444). Cambridge: Cambridge University Press.
- Quartz, S.R. & Sejnowski, T.J. (1997a). The neural basis of cognitive development: a constructivist manifesto. **Behavioral and Brain sciences 20**(4), 537-556.
- Quartz, S.R. & Sejnowski, T.J. (1997b). Controversies and issues in developmental theories of mind: some constructive remarks. **Behavioral and Brain sciences 20**(4), 578-596.
- Quine, W.V.Q. (1983a). Truth by convention. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.329-354). Cambridge: Cambridge University Press.
- Quine, W.V.Q. (1983b). Carnap and logical truth. In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.353-376). Cambridge: Cambridge University Press.
- Quine, W.V.Q. (1993). Two dogmas of empiricism. In J.H. Fetzer (Ed.). **Foundations of philosophy of science: recent developments** (pp.23-38). New York: Paragon.
- Quine, W.V.Q. (2000). Two dogmas of empiricism. In J. McErlean (Ed.), **Philosophies of science** (pp.115-128). California: Wadsworth.
- Quinn, M.H. (1998). The role of theory in psychotherapy. **American psychologist, 53**(1), 71.

- Raab, M. & Gigerenzer, G. (2005). Intelligence as smart heuristics. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.188-207). Cambridge: Cambridge University Press.
- Radick, G. (2005). What if? **New scientist**, 187(2513), 34-47.
- Radnitzky, G. (1970). **Contemporary schools of metascience**. Göteborg: Akademiförlaget.
- Radnitzky, G. & Andersson, G. (1978). Objective criteria of scientific progress? Intuitivism, falsificationism, and relativism. In G. Radnitzky & G. Andersson (Eds.), **Progress and rationality in science volume 58** (pp.3-19). Dordrecht: D. Reidel.
- Radnitzky, G. & Andersson, G. (Eds.). (1978). **Progress and rationality in science. Boston studies in the philosophy of science volume 58**. Dordrecht: D.Reidel.
- Rahmani, L. (1966). Studies on the mental development of the child. In N. O'Connor (Ed.), **Present-day Russian psychology: a symposium by seven authors** (pp.152-177). Oxford: Pergamon.
- Ramachandran, V.S. & Blakeslee, S. (1999). **Phantoms in the brain: human nature and the architecture of the mind**. London: Fourth Estate.
- Ramey, C.T. & MacPhee, D. (1981). A new paradigm in intellectual assessment? **Contemporary psychology**, 26(7), 507-509.
- Ramul, K. (1963). Some early measurements and ratings in psychology. **American psychologist**, 18(10), 653-659.
- Rand, Y. & Kaniel, S. (1987). Group administration of the LPAD. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.196-214). New York: The Guilford Press.
- Rand, Y. & Tannenbaum, A.J. (2000). To be, to have, to do: an integration and expansion of existing concepts. A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.83-113). Amsterdam: Pergamon.
- Rasch, G. (1980). **Probabilistic models for some intelligence and attainment tests**. Chicago: The University of Chicago Press.
- Ratcliff, R. (1998). The role of mathematical psychology in experimental psychology. **Australian journal of psychology**, 50(3), 129-130.
- Ravenscroft, I. (1993). Predictive failure. **Philosophical papers**, 28(3), 143-168.
- Ray, C. (2001). Logical positivism, In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.243-251). Massachusetts: Blackwell.
- Ray, W.J. (2006). **Methods toward a science of behaviour and experience**. Australia: Thomson.
- Reader's Digest. (1988). **Universal Dictionary**. Cape Town: Reader's Digest Association.
- Reber, A.S. & Reber, E. (2001). **The Penguin dictionary of psychology**. London: Penguin.
- Ree, M.J. & Earles, J.A. (1996). Predicting occupational criteria: not much more than *g*. In I. Dennis & P.Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.151-165). New Jersey: Lawrence Erlbaum.
- Reed, E. (1997). The cognitive revolution from an ecological point of view. In D.M. Johnson, & C.E Emeling (Eds.), **The future of the cognitive revolution** (pp.261-274). New York: Oxford.
- Reed, J.G. & Baxter, P.M. (1994). Using reference databases. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp57-70). New York: Russell Sage Foundation.
- Reed, T.E. & Jensen, A. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. **Intelligence**, 16, 259-272.
- Reed, T.E., Vernon, P.A. & Johnson, A.M. (2004). Confirmation of correlation between brain nerve conduction velocity and intelligence level in normal adults. **Intelligence**, 32, 563-572.
- Reese, H.W. (1993). Developments in child psychology from the 1960's to the 1990's. **Developmental review**, 13, 503-524.
- Reeve, C.L. & Lam, H. (2005). The psychometric paradox of practice effects due to retesting: measurement invariance and stable ability estimates in the face of observed score changes. **Intelligence**, 33, 535-549.
- Reichardt, C.S. & Gollob, H.F. (1997). When confidence intervals should be used instead of statistical tests, and vice versa. In L.L. Harlow, S.A. Mulaik & J.H, Steiger (Eds.), **What if there were no significance tests?** (pp.259-284). New Jersey: Lawrence Erlbaum.
- Reichenberg, R. & Rand, Y. (2000). Reflective teaching and its relation to modes of existence in practical teaching experience. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.114-133). Amsterdam: Pergamon.
- Reschly, D.J. (1997). Diagnostic and treatment utility of intelligence tests. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.437-456). New York: Guilford.
- Resing, W.C.M (1993). Measuring inductive reasoning skills: the construction of a learning potential test. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.219-242). Amsterdam: Swets & Zeitlinger.
- Resing, W.C.M. (1997). Learning potential assessment: the alternative for measuring intelligence? **Educational and child psychology**, 14(4), 68-82.
- Resing, W.C.M. (2000). Assessing the learning potential for inductive reasoning (LIR) in young children. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.229-262). New York: Elsevier.
- Resing, W.C.M. (2001). Beyond Binet. **Issues in education**, 7(2), 225-236.
- Resing, W.C.M. & Roth-Van der werf, T.J.M. (2002). Learning potential assessment and cognitive training in inductive reasoning: emergent relationship? In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.),

- Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.175-208). Amsterdam: JAI.
- Resing, W.C.M., Ruijssenaars, A.J.J.M. & Bosma, T. (2002). Dynamic assessment: using measures for learning potential in the diagnostic process. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijssenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.29-64). Amsterdam: JAI.
- Resnick, L.B. & Nelson-Le Gall, S. (2000). Socialising intelligence. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.145-158). London: Routledge.
- Retherford, R.D. & Sewell, W.H. (1989). How intelligence affects fertility. **Intelligence**, **13**(2), 169-185.
- Revelle, W. (1987). Personality and motivation: sources of inefficiency in cognitive performance. **Journal of research in personality**, **21**(4), 436-452.
- Reynolds, C.R. (Ed). (1994). **Cognitive assessment: a multidisciplinary perspective**. New York: Plenum Press.
- Reynolds, J.H. (1987). **Computing in psychology: an introduction to programming methods and concepts**. New Jersey: Prentice-Hall.
- Reynolds, P.D. (1971). **A primer in theory construction**. New York: Bobbs-Merrill.
- Reznick, J.S. (1997). Intelligence, language, nature, and nurture in young twins. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.483-504). Cambridge: Cambridge University Press.
- Reznick, J.S. & Corley, R. (1999). What twins can tell us about the development of intelligence - a case study. In M. Anderson (Ed.), **The development of intelligence** (pp.105-136). Sussex: Psychology Press.
- Richardson, F.C. & Fowers, B.J. (1997). Critical theory, postmodernism and hermeneutics: insights for critical psychology. In D. Fox & I. Prilleltensky (Eds.), **Critical psychology: an introduction** (pp.265-283). London: Sage.
- Richter, R. (1992). **A critical evaluation of cognitive style assessment**. Pretoria: Human Sciences Research Council.
- Rigdon, E.E. (1998). Structural equation modelling. In G.A. Marcoulides (Ed.), **Modern methods for business research** (pp.251-294). New Jersey: Lawrence Erlbaum.
- Ridgway, J. (2000). Why measure development? In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond** (pp.201-208). London: Routledge.
- Riggs, F.W. (1979). The importance of concepts: some considerations on how they might be designated less ambiguously. **American sociologist**, **14**, 172-185.
- Rindskopf, D.M. (1997). Testing "small", not null, hypotheses: classical and Bayesian approaches. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.319-332). New Jersey: Lawrence Erlbaum.
- Ritzer, G. (Ed.). (1992). **Metatheorizing**. Newburg Park: Sage.
- Ritzer, G. (1996). **Sociological theory**. New York: McGraw-Hill.
- Robbins, D. (2001). **Vygotsky's psychology-philosophy. A metaphor for language theory and learning**. New York: Kluwer.
- Robbins, D. (2003). Vygotsky's non-classical dialectical metapsychology. **Journal for the theory of social behaviour**, **33**(3), 303-312.
- Roberts, D. & Adams, R.J. (1997). Bias in weighted likelihood estimation when using the Rasch model. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.279-295). London: Ablex.
- Roberts, J.M. (1995). **The Penguin history of the world**. London: Penguin.
- Roberts, R.D., Markham, P.M., Matthews, G. & Zeidner, M. (2005). Assessing intelligence: past, present, and future. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.333-360). London: Sage.
- Robertson, I.T. & Mindel, R.M. (1990). A study of trainability testing. **Journal of occupational psychology**, **53**(2), 131-138.
- Robinson, D. & Groves, J. (1999). **Introducing philosophy**. London: Icon.
- Robinson, D.N. (1985). Science, psychology, and explanation: synonyms or antonyms? In S. Koch & D.E. Leary (Eds.). **A century of psychology as science** (pp.60-74). New York: McGraw-Hill.
- Robinson, D.N. (2000). Philosophy of psychology at the turn of the century. **American psychologist**, **55**(9), 1018-1021.
- Robinson-Zanartu, C.A. & Aganza, J.S. (2000). Dynamic assessment and sociocultural context: assessing the whole child. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.443-487). New York: Elsevier.
- Rodgers, J.L. & Rowe, D.C. (1987). IQ, similarity in twins, siblings, half-siblings, cousins, and random pairs. **Intelligence**, **11**(3), 199-206.
- Rogers, E. (1992). Larry P., PASE, and social science in the courtroom: the science and politics of identifying and educating very slow learners. In H.C. Haywood & D. Tzurriel (Eds.), **Interactive assessment** (pp.470-503). New York: Springer-Verlag.
- Rogoff, B. & Chavajay, P. (1995). What's become of research on the cultural basis of cognitive development? **American psychologist**, **50**(10), 859-877.
- Rolls, E.T. & Treves, A. (1998). **Neural networks and brain function**. New York: Oxford Press.
- Rosas, R. (2004). Intelligence research in Latin America. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.391-410). **Cambridge: Cambridge University Press**.
- Rose, H. (2000). Colonising the social sciences? In H. Rose & S. Rose (Eds.). **Alas, poor Darwin: arguments against evolutionary psychology** (pp.106-128). London: Jonathan Cape.
- Rose, H. & Rose, S. (Eds.). (2000). **Alas, poor Darwin: arguments against evolutionary psychology**. London: Jonathan Cape.

- Rose, S. (1995). Minds, brains, and Rosetta stones. In J. Brockman & K. Matson (Eds.). **How things are: a science tool-kit for the mind** (pp.201-212). London: Weidenfeld and Nicolson.
- Rose, S. (Ed.). (1999). **From brains to consciousness: essays on the news sciences of the mind**. London: Penguin.
- Rosenberg, A. (1985). The place of psychology in a vacuum of theories. In K.B. Madsen & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 3** (pp.95-102). New York: Plenum Press.
- Rosenberg, A. (2001). Philosophy of social science. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.451-460). Massachusetts: Blackwell.
- Rosenberg, M.S., Adams, D.C. & Gurevith, J. (2005). **MetaWin**. Retrieved April 13 2005 from www.metawinsoft.com
- Rosengren, K.E. (1981). Advances in Scandinavian content analysis: an introduction. In K.E. Rosengren (Ed.), **Advances in content analysis (pp.9-19)**. Beverly Hills: Sage.
- Rosengren, K.E. (Ed.) (1981). **Advances in content analysis**. Beverly Hills: Sage.
- Rosenthal, M.C. (1994). The fugitive literature. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp85-94). New York: Russell Sage Foundation.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. **Psychological bulletin**, **86**(3), 638-641.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp231-244). New York: Russell Sage Foundation
- Rosenthal, R. (1995). Writing meta-analytic reviews. **Psychological bulletin**, **118**(2), 183-192.
- Rosenzweig, M.R. (1998). Historical perspectives on the development of the biology of learning and memory. In J.L. Martinez & R.P.Kesner (Eds.), **Neurobiology of learning and memory** (pp.1-53). San Diego: Academic Press.
- Ross, P.E. (2004). Draining the language out of colour. **Scientific American**, **290**(4), 46-47.
- Ross, P.E. (2005). Math without words. **Scientific American**, **292**(6), 28-30.
- Ross, S. (1964). **Logical foundations of psychological measurement: a study in the philosophy of science**. Copenhagen: Munksgaard.
- Rossi, J.S. (1997). A case study in the failure of psychology as a cumulative science: the spontaneous recovery of verbal learning. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.175-197). New Jersey: Lawrence Erlbaum.
- Rothman, H.R. & Semmel, M.I. (1990). Dynamic assessment: a comprehensive review of the literature. In R. Gaylord-Ross (Ed.), **Issues and research in special education volume 1** (pp.355-386). New York: Teachers College.
- Royce, J.R. (1959). The search for meaning. **American scientist**, **47**, 515-535.
- Royce, J.R. (1964). A note on the development of the psycho-epistemological profile (PEP). **Psychological reports**, **14**, 297-298.
- Royce, J.R. (1970a). The present situation in theoretical psychology. In J.R. Royce (Ed.). **Toward unification in psychology: the first Banff conference on theoretical psychology** (pp.10-37). Toronto: University of Toronto Press.
- Royce, J.R. (1970b). (Ed.). **Toward unification in psychology: the first Banff conference on theoretical psychology**. Toronto: University of Toronto Press.
- Royce, J.R. (1973). The present situation in theoretical psychology. In B. Wolman (Ed). **Handbook of general psychology** (pp8-21). New Jersey: Prentice-Hall.
- Royce, J.R. (1975). How can metatheory advance substantive theory? Review of "Modern theories of motivation: a comparative metascientific study" by K.B. Madsen. **Contemporary psychology**, **20**(10), 817-818.
- Royce, J.R. (1978). How we can best advance the construction of theory in psychology. **Canadian psychological review**, **19**(4), 259-276.
- Royce, J.R. (1987). A strategy for developing unifying theory in psychology. In A.W. Staats & L.P.Mos (Eds.), **Annals of theoretical psychology, volume 5** (pp275-285). New York: Plenum Press.
- Royce, J.R., Coward, H., Egan, E., Kessel, F. & Mos, L. (1978). Psychological epistemology: a critical review of the empirical literature and the theoretical issues. **Genetic psychology monographs**, **97**, 265-353.
- Royce, J.R. & Mos, L.P.(Eds.) (1984). **Annals of theoretical psychology. Volume 2**. New York: Plenum Press.
- Rozeboom, W.W. (1970). The art of metascience, or, what should a psychological theory be? In J.R. Royce (Ed.). **Toward unification in psychology: the first Banff conference on theoretical psychology** (pp.54-160). Toronto: University of Toronto Press.
- Rozeboom, W.W. (1995). Toward more effective metatheory: a response to Kukla. **New ideas in psychology**, **13**(3), 237-245.
- Rozeboom, W.W. (1997). Good science is abductive, not hypothetico-deductive. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.335-391). New Jersey: Lawrence Erlbaum.
- Rubinštejn, S.L. (1987). Problems of psychology in the works of Karl Marx. **Studies in soviet thought**, **33**(2), 111-130.
- Rucker, R. (1986). **The fourth dimension and how to get there**. New York: Penguin.
- Rucker, R. (1987). **Mind tools: the five levels of mathematical reality**. Boston: Houghton Mifflin.
- Ruijsenaars, A.J.J.M., Castelijns, J.H.M. & Hamers, J.H.M. (1993). The validity of learning potential tests. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.69-82). Amsterdam: Swets & Zeitlinger.
- Rumbaugh, D.M. (2002). Emergents and rational behaviourism. **Journal of cognitive education and psychology**, **2**(2), 168-176.
- Ruml, B. (1921). Intelligence and its measurement. **Journal of educational psychology**, **12**(3), 143-144.

- Runyan, W.M. (2005). Toward a better story of psychology. Sheldon White's contributions to the history of psychology, a personal perspective. In D.B. Pillemer & S.H. White (Ed.), **Developmental psychology and social change: research, history, and policy** (pp.59-80). Cambridge: Cambridge University Press.
- Rupp, A.A., Dey, D.K. & Zumbo, B.D. (2004). To Bayes or not to Bayes, from whether to when: applications of Bayesian methodology to modelling. **Structural equation modelling**, 11(3), 424-451.
- Russell, B. (1983). Selections from "Introduction to mathematical philosophy". In P.Benacerraf & H. Putnam (Eds.). **Philosophy of mathematics: selected readings** (pp.160-182). Cambridge: Cambridge University Press.
- Rust, J. & Golombok, S. (1992). **Modern psychometrics: the science of psychological assessment**. London: Routledge.
- Ryan, A. (1970). **The philosophy of the social sciences**. London: Macmillan.
- Ryan, G.W. & Bernard, H.R. (2000). Data management and analysis methods. In N.K. Denzin & Y.S. Lincoln (Eds.), **Handbook of qualitative research** (pp.769-802). London: Sage.
- Ryan, J.J. & Lopez, S.J. (2001). Wechsler adult intelligence scale-III. In W.I. Dorfman & M. Hersen (Eds.), **Understanding psychological assessment** (pp.19-42). New York: Kluwer.
- Ryan, J.J., Lopez, S.J. & Sumerall, S.W. (2001). Understanding test construction. In W.I. Dorfman & M. Hersen (Eds.), **Understanding psychological assessment** (pp.1-15). New York: Kluwer.
- Rzhetsky, A., Lossifov, I., Koike, T., Krauthammer, M., Kra, P., Morris, M., Yu, H., Duboué, P.A., Weng, W., Wilbur, W.J., Hatzivassiloglou, V. & Friedman, C. (2004). GeneWays: a system for extracting, analysing, visualizing, and integrating molecular pathway data. **Journal of biomedical informatics**, 37(1), 43-53.
- Sacks, O. (1995). **An anthropologist on Mars: seven paradoxical tales**. New York: Picador.
- Sagan, C. (1997). **The demon-haunted world; science as a candle in the dark**. London: Headline.
- Sahakian, W.S. (1975). **History and systems of psychology**. New York: Halsted Press.
- Sahakian, W.S. (1981). **History of psychology: a source book in systematic psychology**. Illinois: F.E. Peacock.
- Salmon, W.C. (2001). Logical empiricism. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.233-242). Massachusetts: Blackwell.
- Sampson, E. (2000). Of rainbows and differences. In T. Sloan (Ed.), **Critical psychology: voices for change** (pp.1-5). London: Macmillan.
- Samuels, M.T. (2000). Assessment of post-secondary students with learning difficulties: using dynamic assessment in a problem solving process. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.521-542). New York: Elsevier.
- Samuels, M.T., Killip, S.M., MacKenzie, H. & Fagan, J. (1992). Evaluating preschool programs: the role of dynamic assessment. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.251-274). New York: Springer-Verlag.
- Samuels, M.T., Lamb, C.H. & Oberholtzer, L. (1992). Dynamic assessment of adults with learning difficulties. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.275-299). New York: Springer-Verlag.
- Sanders, C. & Rappard, H.V. (1985). Psychology and philosophy of science. In K.B. Madsen & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 3** (pp.219-267). New York: Plenum Press.
- Santiago-Delefosse, M.J. & Delefosse, J-M. (2002). Spielrein, Piaget and Vygotsky: three positions on child thought and language. **Theory and psychology**, 12(6), 723-747.
- Sarason, S.B. (1976). The unfortunate fate of Alfred Binet and school psychology. **Teachers college record**, 77(4), 579-592.
- Sardar, Z. Ravetz, J. & Van Loon, B. (2000). **Introducing mathematics**. Cambridge: Icon.
- Sato, T., Namiki, H., Ando, J. & Hatano, G. (2004). Japanese conception of and research on human intelligence. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.302-324). Cambridge: Cambridge University Press.
- Savage, C.W. & Ehrlich, P.(Eds.), (1992). **Philosophical and foundational issues in measurement theory**. New Jersey: Lawrence Erlbaum.
- Savage, C.W. & Ehrlich, P.(1992). A brief introduction to measurement theory and to the essays. In C.W. Savage & P.Ehrlich (Eds.), **Philosophical and foundational issues in measurement theory** (pp.1-14). New Jersey: Lawrence Erlbaum.
- Scarr, S. (1997). Behavior-genetic and socialization theories of intelligence: truce and reconciliation. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.3-41). Cambridge: Cambridge University Press.
- Scarr, S. (1998a). How do families affect intelligence? Social environmental and behavior genetic predictions. J.J. McArdle & R.W. Woodcock (Eds.), **Human cognitive abilities in theory and practice** (pp.113-136). New Jersey: Lawrence Erlbaum.
- Scarr, S. (1998b). Intelligence: revisited. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.117-120). New Jersey: Ablex.
- Schaeffer, D.L. (1984). ...But discretion were discretion the better part of valor. In J.R. Royce & L.P.Mos (Eds.), **Annals of theoretical psychology. Volume 2** (pp.225-233). New York: Plenum Press.
- Schafer, E.W.P.(1982). Neural adaptability: a biological determinant of behavioral intelligence. **International journal of neuroscience**, 17, 183-191.
- Schaffer, S. (1997). What is science? In J. Krige & D. Pestre (Ed.). **Science in the twentieth century** (pp.27-41). Amsterdam: Harwood Academic.
- Schank, R.C. (1998). Explaining intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.121-131). New Jersey: Ablex.

- Schardt, A., Whitten, C. & Gamlin, P. (1996). A walk on our Moccasins: applying dynamic approaches in a remote native community. In M. Luther, E. Cole & P. Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.316-325). Ontario: Captus Press.
- Scharnhorst, U. & Büchel, F.P. (1995). Teaching and evaluation of inductive reasoning in nonretarded and mildly retarded students. In J.S. Carlson (Ed.), **Advances in cognition and educational practice vol 3: European contributions to dynamic assessment** (pp.209-240). London: JAI Press.
- Scheibel, A.B. (1997). The right way, the wrong way, and the army way: a dendritic parable. **Behavioral and brain sciences**, **20**(4), 575.
- Scheuneman, J.D. (1991). Item bias and individual differences. In S.H. Irvine & S.E. Newstead (Eds.), **Intelligence and cognition: contemporary frames of reference** (pp.259-278). Dordrecht: Martinus Nijhoff.
- Schlatter, C. & Büchel, F.P. (2000). Detecting reasoning abilities of persons with moderate mental retardation: the analogical reasoning test (ARLT). In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.155-186). New York: Elsevier.
- Schlesinger, I.M. (1995). What is a theory of cognitive development a theory of? In G. Ben-Shakhar & A. Lieblich (Eds.), **Studies in psychology in honour of Solomon Kugelmass** (pp.261-273). Jerusalem: Magnes Press.
- Schmidt, F.L. & Hunter, J.E. (1997). Eight common but false objections to discontinuation of significance testing in the analysis of research data. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.37-64). New Jersey: Lawrence Erlbaum.
- Schmiedek, F. (2005). Item response theory and the measurement of cognitive processes. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.265-277). London: Sage.
- Schmiedeskamp, M. (2000). Preventing good brains from going bad. **Scientific American**, **11**(2), 72-79.
- Schneider, E. & Ganschow, L. (2000). Dynamic assessment and instructional strategies for learners who struggle to learn a foreign language. **Dyslexia**, **6**, 72-82.
- Schönemann, P.H. (1994). Measurement: the reasonable ineffectiveness of mathematics in the social sciences. In I. Borg & P. Mohler (Eds.), **Trends and perspectives in empirical social research** (pp.149-160). Berlin: Walter De Gruyter.
- Schöttke, H., Bartram, M & Wiedl, K.H. (1993). Psychometric implications of learning potential assessment: a typological approach. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.153-173). Amsterdam: Swets & Zeitlinger.
- Schouten, M.K.D. (2003). Levels and labyrinths: learning about LTP. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.378-386). Ontario: Captus University Publications.
- Schur, Y., Skuy, M., Zietsman, A. & Fridjhon, P. (2002). *A thinking journey* based on constructivism and mediated learning experience as a vehicle for teaching science to low functioning students and enhancing their cognitive skills. **School psychology international**, **23**(1), 36-67.
- Schraw, G. & Olafson, L. (2003). Teachers' epistemological world views and educational practices. **Journal of cognitive education and psychology**, **3**(1), 178-235.
- Schulze, R. (2005). Modeling structures of intelligence. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.241-263). London: Sage.
- Schwartz, G. & Bishop, P.W. (Eds.), (1959a). Moments of discovery: the origins of science. Vol 1. New York: Basic books.
- Schwartz, G. & Bishop, P.W. (Eds.), (1959b). Moments of discovery: the development of modern science. Vol 2. New York: Basic books.
- Schwarzer, R. (1989). Meta-analysis programs. Retrieved July 25, 2004, from <http://www.RalfSchwarzer.De>.
- Schwebel, M. (1992). The social role in cognitive education and development. In J.S. Carlson (Ed.), **Advances in cognition and educational practice: a research annual volume 1A** (pp.147-168). Connecticut: JAI Press.
- Scientific American Editorial. (2005). Brief points. **Scientific American**, **293**(3), 36.
- Scott, B. & Louis, C. (2005). Generation of tissue-specific transgenic birds with lentiviral vectors. **Proceedings of the national academy of sciences**, **102**(45), 16443-16447.
- Scruton, R. (2004). **Modern philosophy: an introduction and survey**. London: Pimlico.
- Seager, W. (2001). Physicalism. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.340-342). Massachusetts: Blackwell.
- Secord, P.F. (1986). Explanation in the social sciences and in life situations. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.197-221). Chicago: University of Chicago Press.
- Segerstråle, U. (2001). **Defenders of the truth: the socio-biology debate**. Oxford: Oxford University Press.
- Segré, E. (1980). **From x-rays to quarks: modern physicists and their discoveries**. San Francisco: W.H. Freeman.
- Sehlapelo, M & Terre Blanche, M. (1996). Psychometric testing in South Africa: views from above and below. **Psychology in society**, **21**, 49-59.
- Serlin, R.C. & Lapsley, D.K. (1990). Meehl on theory appraisal. **Psychological inquiry**, **1**(2), 169-172.
- Sewell, T.E. (1987). Dynamic assessment as a non-discriminatory procedure. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.426-443). New York: The Guilford Press.

- Shadish, W.R. (1989). The perception and evaluation of quality in science. In B. Gholson, W.R. Shadish, R.A. Neimeyer, & A.C. Houts, (Eds.). **Psychology of science: contributions to metascience (pp.383-426)**. Cambridge: Cambridge University Press.
- Shadish, W.R. & Haddock, C.K. (1994). Combining estimates of effect size. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.261-281). New York: Russell Sage Foundation.
- Shadish, W.R., Houts, A.C., Gholson, B. & Neimeyer, R.A. (1989). The psychology of science: an introduction. In B. Gholson, W.R. Shadish, R.A. Neimeyer, & A.C. Houts, (Eds.). **Psychology of science: contributions to metascience (pp.1-16)**. Cambridge: Cambridge University Press.
- Shadish, W.R., Robinson, L. & Wu, C. (2005). 'ES' an effect size computational programme. Retrieved April 13 2005 from <http://faculty.ucmerced.edu/wshadish/es.htm>
- Shakespeare, T. & Erickson, M. (2000). Different strokes: beyond biological determinism and social constructionism. In H. Rose & S. Rose (Eds.). **Alas, poor Darwin: arguments against evolutionary psychology** (pp.190-205). London: Jonathan Cape.
- Shames, M.L. (1987a). Lagging behind the Papacy: whither psychology's aggiornamento? Stam, H.J., Rogers, T.B. & Gergen, K.J. (Eds.). (1987). **The analysis of psychological theory: metapsychological perspectives** (pp.25-40). Cambridge: Hemisphere.
- Shames, M.L. (1987b). Methodocentricity, theoretical sterility, and the socio-behavioural sciences. In W.J. Baker., M.E. Hyland., H. Van Rappard. & A.W. Staats (Eds.), **Current issues in theoretical psychology: selected/edited proceedings of the founding conference of the international society for theoretical psychology held in Plymouth, U.K., 30 August - 2 September, 1985** (pp.261-282). Amsterdam: North-Holland.
- Shamir, A. & Tzuriel, D. (2002). Peer mediation: a novel model for development of mediation skills and cognitive modifiability of young children. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.365-383). Amsterdam: JAI.
- Shao-Horn, Y., Croguennec, L., Delmas, C., Nelson, E.C. & O'Keefe, M. (2003). Atomic resolution of lithium ions in LiCoO₂. **Nature Materials**, **2**, 464-467.
- Sharratt, P.A. & Van den Heuvel, E. (1995). Metamemorial knowledge in a group of black South African school children. **South African journal of psychology**, **25**(2), 59-73.
- Shayer, M. (2000). Piaget and Vygotsky: a necessary marriage for effective educational intervention. In L. Smith, J. Dockrell & P. Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.36-59). London: Routledge.
- Sheng, Y.H. (1996). A Chinese perspective on Leontiev's theory of personality. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.207-212). London: Greenwood Press.
- Shepherd, M.J. (2001). History lessons. In A.S. Kaufman & N.L. Kaufman (Eds.), **Specific learning disabilities and difficulties in children and adolescents: psychological assessment and evaluation** (pp.3-28). Cambridge: Cambridge University Press.
- Shermer, M. (2001). **The borderlands of science: where sense meets nonsense**. Oxford: Oxford University Press.
- Shermer, M. (2005). **Skeptic**. Scientific American, 292(5), 37.
- Shi, J. (2004). Diligence makes people smart. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.325-343). Cambridge: Cambridge University Press.
- Shochet, I.M. (1986). **Manifest and potential performance in advantaged and disadvantaged students**. Unpublished D.Phil. dissertation, University of the Witwatersrand.
- Shochet, I.M. (1992). A dynamic assessment for undergraduate admission: the inverse relationship between modifiability and predictability. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.332-355). New York: Springer-Verlag.
- Shochet, I.M. (1994). The moderator effect of cognitive modifiability on a traditional undergraduate admissions test for disadvantaged black students in South Africa. **South African journal of psychology**, **24** (4), 208-215.
- Short, E.J. & Weissberg-Benchell, J.A. (1989). The triple alliance for learning: cognition, metacognition, and motivation. In C.B. McCormick, G.E. Miller, & M. Pressley (Eds.), **Cognitive strategy research: from basic research to educational implications** (pp.33-63). London: Springer-Verlag.
- Shotter, J. (1989). Vygotsky's psychology: joint activity in a developmental zone. **New ideas in psychology**, **7**(2), 185-204.
- Shotter, J. (1993). Bakhtin and Vygotsky: internalisation as a boundary phenomenon. **New ideas in psychology**, **11**(3), 379-390.
- Shotter, J. (1997). Cognition as a social practice: from computer power to word power. In D.M. Johnson, & C.E. Erneling (Eds.), **The future of the cognitive revolution** (pp.317-334). New York: Oxford.
- Shrout, P.E. (1997). Should significance tests be banned? Introduction to a special section exploring the pros and cons. **Psychological science**, **8**(1), 1-2.
- Shweder, R.A. (1986). Divergent rationalities. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.163-196). Chicago: University of Chicago Press.
- Shweder, R.A. & Fiske, D.W. (1986). Introduction: Uneasy social science. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.1 - 19). Chicago: University of Chicago Press.



- Sijtsma, K. (1993a). Classical and modern test theory with an eye toward learning potential testing. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.117-133). Amsterdam: Swets & Zeitlinger.
- Sijtsma, K. (1993b). Psychometric issues in learning potential assessment. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.175-193). Amsterdam: Swets & Zeitlinger.
- Silva, A.J. & Giese, K.P. (1998). Gene targeting: a novel window into the biology of learning and memory. In J.L. Martinez & R.P. Kesner (Eds.), **Neurobiology of learning and memory** (pp.89-142). San Diego: Academic Press.
- Silverman, D. (2000). Analyzing talk and text. In N.K. Denzin & Y.S. Lincoln (Eds.), **Handbook of qualitative research** (pp.821-834). London: Sage.
- Silverman, D. (2005). **Interpreting qualitative data: methods for analysing talk, text and interaction**. London: Sage.
- Silverman, D. & Waksman, M. (1992). Assessing the learning potential of penitentiary inmates: an application of Feuerstein's learning potential assessment device. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.94-115). New York: Springer-Verlag.
- Silvestri, H. & Kose, G. (2003). Rationality and the practices of science: a re-reading of Kant. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.70-77). Ontario: Captus University Publications.
- Simonton, D.K. (2003). Francis Galton's hereditary genius: its place in the history and psychology of science. In R.J. Sternberg (Ed.), **The anatomy of impact: what makes the great works of psychology great** (pp.3-18). Washington: American psychological association.
- Singer, W. (1999). Consciousness from a neurobiological perspective. In S. Rose (Ed.) **From brains to consciousness? Essays on the new sciences of mind** (pp.228-245). London: Penguin.
- Skinner, Q. (1986). Introduction: the return of grand theory. In Q. Skinner (Ed.). **The return of grand theory in the human sciences** (pp.1-20). Cambridge: Cambridge University Press.
- Skinner, Q. (Ed.). (1986). **The return of grand theory in the human sciences**. Cambridge: Cambridge University Press.
- Sklar, L. (2001). Role of convention. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.56-64). Massachusetts: Blackwell.
- Skuy, M. (1989). **Dynamic assessment of South African children and adolescents**. Unpublished document. Johannesburg: University of the Witwatersrand.
- Skuy, M. (1996). Cross-cultural and inter-dimensional implications of Feuerstein's construct of mediated learning experience. In P.Engelbrecht, S. Kriegler & M. Booysen (Eds.), **Perspectives on learning difficulties: international concerns and South African realities** (pp.185-198). Pretoria: Van Schaik.
- Skuy, M. (2002). The development of MLE-based intervention programmes in the context of education in South Africa. **School psychology international**, **23**(1), 85-111.
- Skuy, M.S., Archer, M. & Roth, I. (1987). Use of the learning potential assessment device in assessment and remediation of a learning problem. **South African journal of education**, **7**(1), 53-58.
- Skuy, M.S., Gewer, A., Osrin, Y., Khunou, D., Fridjon, P., & Rushton, J. P.(2002). Effects of mediated learning experience on Raven's Matrices scores of African and non-African university students in South Africa. **Intelligence**, **30**, 221-232.
- Skuy, M. & Mentis, M. (1990). Applications and adaptations of Feuerstein's instrumental enrichment programme in South Africa. In H. Van Niekerk (Ed.), **Cognitive development in the Southern African context: papers of the seminar on cognitive development, 1 November 1990** (pp.132-152). Pretoria: Human Sciences Research Council.
- Skuy, M.S. & Mentis, M. (1992). Applications and adaptations of Feuerstein's instrumental enrichment program in South Africa. In J.S. Carlson (Ed.), **Advances in cognition and educational practice: a research annual volume 1B** (pp.105-127). Connecticut: JAI Press.
- Skuy, M.S., Zolezzi, S., Mentis, M., Fridjon, P. & Cockroft, K. (1996). Selection of advantaged and disadvantaged South African students for university admission. **South African journal of higher education**, **10** (1), 110-118.
- Skyrms, B. (1993). The Goodman paradox and the new riddle of induction. In J.H. Fetzer (Ed.). **Foundations of philosophy of science: recent developments** (pp.59-72). New York: Paragon.
- Slenders, A.P.A.C. & Resing, W.C.M. (1997). Leerpotentieelonderzoek bij allochtone leerlingen: een verkenning met De leerpotentieeltest voor inductief redeneren. **Pedagogische studiën**, **74**, 183-196.
- Slife, B.D. & Williams, R.N. (1997). Toward a theoretical psychology: should a subdiscipline be formally recognized? **American psychologist**, **52**(2), 117-129.
- Sloan, T. (Ed.). (2000). **Critical psychology: voices for change**. London: Macmillan.
- Smedslund, G. (2000). A pragmatic basis for judging models and theories in health psychology: the axiomatic method. **Journal of health psychology**, **5**(2),n 133-149.
- Smith, E.V. & Smith, R.M. (2004). (Eds.). **Introduction to Rasch measurement: theories, models and applications**. Minnesota: JAM.
- Smit, G.J. (1996). **Psychometrics: aspects of measurement**. Pretoria: Kagiso Tertiary.
- Smith, B.D. (1988). Personality: multivariate systems theory and research. In J.R. Nesselroade & R.B. Cattell (Eds.), **Handbook of multivariate experimental psychology** (pp.687-736). New York: Plenum.



- Smith, L. (1996). The social construction of rational understanding. In A. Tryphon & J. Vonèche (Eds.), **Piaget-Vygotsky: the social genesis of thought** (pp.107-123). Sussex: Psychology Press.
- Smith, L., Dockrell, J. & Tomlinson, P. (Eds.), (2000). **Piaget, Vygotsky and beyond**. London: Routledge.
- Smith, L.B. & Katz, D.B. (1996). Activity-dependent processes in perceptual and cognitive development. In R. Gelman & T.K-F. Au (Eds.), **Perceptual and cognitive development** (pp.413-445). San Diego: Academic Press.
- Smith, R.M. (1997). Pre/post comparisons in Rasch measurement. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.297-312). London: Ablex.
- Smullyan, R.M. (1981). An epistemological nightmare. In D.R. Hofstadter & D.C. Dennett (Eds.), **The mind's I: fantasies and reflections on self and soul (pp.427-429)**. London: Penguin
- Snow, C.P. (1981). **The physicists**. London: Macmillan.
- Snow, R.E. (1990). Progress and propaganda in learning assessment. **Contemporary psychology**, **35**(12), 1134-1135.
- Snow, R.E. (1998a). Abilities as aptitudes and achievements in learning situations. J.J. McArdle & R.W. Woodcock (Eds.), **Human cognitive abilities in theory and practice** (pp.93-112). New Jersey: Lawrence Erlbaum.
- Snow, R.E. (1998b). On intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.133-139). New Jersey: Ablex.
- Snow, R.E. & Lohman, D.F. (1984). Toward a theory of cognitive aptitude for learning from instruction. **Journal of educational psychology**, **76**(3), 347-376.
- Snyman, J. (1993). Social science according to the Frankfurt school. In J. Snyman (Ed.), **Conceptions of social inquiry** (pp.159-179). Pretoria: Human Sciences Research Council.
- Snyman, J. (Ed.). (1993). **Conceptions of social inquiry**. Pretoria: Human Sciences Research Council.
- Sober, E. (2001). Simplicity. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.433-441). Massachusetts: Blackwell.
- Social Neuroscience. (2005). Social Neuroscience - a new academic journal. Retrieved November 7 2005 from <http://www.social-neuroscience.com/>
- Solso, R.L. (1996). Russian psychology and the "new capitalism". In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.144-158). London: Greenwood Press.
- Solzhenitskyn, A. (1974). **The Gulag archipelago 1918-1956**. Glasgow: Collins.
- Spearman, C.E. (1981). 'General intelligence' objectively determined and measured. In W.S. Sahakian (Ed.), **History of psychology: a source book in systematic psychology** (pp.249-252). Illinois: F.E. Peacock. Originally published in 1904.
- Spence, K.W. & Spence, J.T. (Eds.), (1968). **The psychology of learning and motivation: advances in research and theory (vol 2)**. New York: Academic Press.
- Spiker, C.C. (1989). Cognitive psychology: mentalistic or behavioristic? **Advances in child development**, **21**, 73-90.
- Spitz, H.H. (1999). Attempts to raise intelligence. In M. Anderson (Ed.), **The development of intelligence** (pp.275-293). Sussex: Psychology Press.
- Sporns, O. (1997). Deconstructing neural constructivism. **Behavioral and brain sciences**, **20**(4), 576-577.
- Squire, L.R. (1986). Memory and the brain. In S.L. Friedman, K.A. Klivington & R.W. Peterson (Eds.), **The brain, cognition, and education** (pp.171-202). Orlando: Academic Press.
- Stam, H.J. (1987). The psychology of control: a textual critique. In H.J. Stam, T.B. Rogers, & K.J. Gergen. (Eds.). **The analysis of psychological theory: metapsychological perspectives** (pp.131-156). Cambridge: Hemisphere.
- Stam, H.J. (1998). Theory and disciplinarity. **American psychologist**, **53**(1), 70-71.
- Stam, H. J. (2001). Introduction: social constructionism and its critics. **Theory and psychology**, **11**(3), 291-296.
- Stam, H.J., Rogers, T.B. & Gergen, K.J. (Eds.). (1987). **The analysis of psychological theory: metapsychological perspectives**. Cambridge: Hemisphere.
- Stankov, L. (2004). Similar thoughts under different stars: conceptions of intelligence in Australia. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.344-363). **Cambridge: Cambridge University Press**.
- Stankov, L. (2005a). Reductionism versus charting: ways of examining the role of lower-order cognitive processes in intelligence. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.51-67). Cambridge: Cambridge University Press.
- Stankov, L. (2005b). g factor: issues of design and interpretation. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.279-293). London: Sage.
- Stanley, N.V., Siegel, J., Cooper, L. & Marshall, K. (1995). Identification of gifted with the dynamic assessment procedure (DAP). **Gifted educational international**, **10**, 85-87.
- Stegmüller, W. (1976). **The structure and dynamics of theories**. New York: Springer-Verlag.
- Steiger, J.H. & Fouladi, R.T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), **What if there were no significance tests?** (pp.221-257). New Jersey: Lawrence Erlbaum.
- Stephenson, N., Radtke, H.L., Jorna, R & Stam, H.J. (Eds.), (2003). **Theoretical psychology: critical contributions**. Ontario: Captus University Publications.
- Sternberg, R.J. (1979). Six authors in search of a character: a play about intelligence tests in the year 2000. **Intelligence**, **3**, 283-293.

- Sternberg, R.J. (Ed.), (1982). **Advances in the psychology of human intelligence volume 1**. New Jersey: Lawrence Erlbaum.
- Sternberg, R.J. (1988). A framework for understanding conceptions of intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.3-15). New Jersey: Ablex.
- Sternberg, R.J. (1991). Death, taxes, and bad intelligence tests. **Intelligence**, **15**, 257-269.
- Sternberg, R.J. (Ed.), (1994a). **Encyclopedia of human intelligence vol 1**. New York: Macmillan.
- Sternberg, R.J. (Ed.), (1994b). **Encyclopedia of human intelligence vol 2**. New York: Macmillan.
- Sternberg, R.J. (1994c). Human intelligence: its nature, use, and interaction with context. In D.K. Detterman (Ed.), **Current topics in human intelligence volume 4: theories of intelligence** (pp.361-407). New Jersey: Ablex.
- Sternberg, R.J. (1996a). Matching abilities, instruction, and assessment: reawakening the sleeping giant of ATI. In I. Dennis & P. Tapsfield (Eds.), **Human abilities: their nature and measurement** (pp.167-181). New Jersey: Lawrence Erlbaum.
- Sternberg, R.J. (1996b). IQ counts, but what really counts is successful intelligence. **Journal for middle level and high school leaders**, **80**(583), 18-23.
- Sternberg, R.J. (1997a). Educating intelligence: infusing the triarchic theory into school instruction. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.343-362). Cambridge: Cambridge University Press.
- Sternberg, R.J. (1997b). The triarchic theory of intelligence. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.92-104). New York: Guilford.
- Sternberg, R.J. (1997c). The concept of intelligence and its role in lifelong learning and success. **American psychologist**, **52**(10), 1030-1037.
- Sternberg, R.J. (1998). Intelligence is mental self-government. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.141-148). New Jersey: Ablex.
- Sternberg, R.J. (2000a). Prologue: dynamic assessment: prevailing models and applications. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.xiii-xv). New York: Elsevier.
- Sternberg, R.J. (2000b). Group and individual differences in intelligence: what can and should we do about them? A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.55-82). Amsterdam: Pergamon.
- Sternberg, R.J. (2003a). Issues in the theory and measurement of successful intelligence: a reply to Brody. **Intelligence**, **31**, 331-337.
- Sternberg, R.J. (Ed.). (2003b). **The anatomy of impact: what makes the great works of psychology great**. Washington: American psychological association.
- Sternberg, R.J. (Ed.). (2004a). **Handbook of intelligence**. Cambridge: Cambridge University Press.
- Sternberg, R.J. (Ed.). (2004b). **International handbook of intelligence**. Cambridge: Cambridge University Press.
- Sternberg, R.J. (2004c). North American approaches towards intelligence. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.411-444). Cambridge: Cambridge University Press.
- Sternberg, R.J. (2004d). Preface. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.ix-xi). Cambridge: Cambridge University Press.
- Sternberg, R.J. (2004e). The concept of intelligence. In R.J. Sternberg (Ed.), **Handbook of intelligence** (pp.3-15). Cambridge: Cambridge University Press.
- Sternberg, R.J. & Berg, C. A. (1998). Quantitative integration: definitions of intelligence: a comparison of the 1921 and 1986 symposia. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.155-162). New Jersey: Ablex.
- Sternberg, R.J. & Detterman, D.K. (Eds.). (1988). **What is intelligence? Contemporary viewpoints on its nature and definition**. New Jersey: Ablex.
- Sternberg, R.J., Ferrari, M., Clinkenbeard, P. & Grigorenko, E.L. (1996). Identification, instruction, and assessment of gifted children: a construct validation of a triarchic model. **Gifted child quarterly**, **40**(3), 129-137.
- Sternberg, R.J. & Grigorenko, E.L. (Eds.). (1997). **Intelligence, heredity and environment**. Cambridge: Cambridge University Press
- Sternberg, R.J. & Grigorenko, E.L. (2001a). All testing is dynamic. **Issues in Education**, **7**(2), 138-172.
- Sternberg, R.J. & Grigorenko, E.L. (2001b). The truth can now be told. **Issues in Education**, **7**(2), 251-261.
- Sternberg, R.J. & Grigorenko, E.L. (2001c). Ability testing across cultures. In L.A. Suzuki, J.G. Ponterotto & P.J. Meller (Eds.), **Handbook of multicultural assessment: clinical, psychological, and educational applications** (pp.335-358). San Francisco: Jossey-Bass.
- Sternberg, R.J. & Grigorenko, E.L. (2002). **Dynamic Testing: the nature and measurement of learning potential**. Cambridge: Cambridge University Press.
- Sternberg, R.J., Grigorenko, E.L., & Kidd, K.K. (2005). Intelligence, race, and genetics. **American psychologist**, **60**(1), 46-59.
- Sternberg, R.J., Grigorenko, E.L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C., Jukes, M. & Bundy, D.A. (2002). Assessing intellectual potential in rural Tanzanian school children. **Intelligence**, **30**, 141-162.
- Sternberg, R.J. & Jarvin, L. (2003). Alfred Binet's contributions as a paradigm for impact in psychology. R.J. Sternberg (Ed.), **The anatomy of impact: what makes the great works of psychology great** (pp.89-107). Washington: American psychological association.

- Sternberg, R.J. & Pretz, J.E. (Eds.), (2005). **Cognition and intelligence: identifying the mechanisms of the mind**. Cambridge: Cambridge University Press.
- Sternberg, R.J. & Wagner, R.K. (Eds.), (1986). **Practical intelligence: nature and origins of competence in the everyday world**. Cambridge: Cambridge University Press.
- Sternberg, R.J. & Williams, W.M. (Eds.), (1998). **Intelligence, instruction, and assessment: theory into practice**. New Jersey: Lawrence Erlbaum.
- Sternheim, M.M. & Kane, J.W. (1991). **General physics**. New York: John Wiley.
- Stevens, S.S. (1946). On the theory of scales of measurement. **Science**, **103**(2684), 677-680.
- Stevens, S.S. (Ed.). (1951a). **Handbook of experimental psychology**. New York: John Wiley.
- Stevens, S.S. (1951b). Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.). **Handbook of experimental psychology** (pp.1-49). New York: John Wiley.
- Stevenson, L. (2000). Is science value-neutral? In J. McErlean (Ed.), **Philosophies of science** (pp.211-216). California: Wadsworth.
- Stewart, I. (1990). **Does God play dice? The new mathematics of chaos**. London: Penguin.
- Stewart, I. & Golubitsky, M. (1993). **Fearful symmetry: is God a geometer?** London: Penguin.
- Stewart, I. (1995). **Nature's numbers: discovering order and pattern in the universe**. London: Weidenfeld & Nicolson.
- Stewart, I. (1998). Chaos and order. In J. Gribbin (Ed.). **A brief history of science** (pp.17-49). London: Weidenfeld & Nicolson.
- Stich, S.P. & T.A. Warfield, T.A. (Eds.) (2003). **The Blackwell guide to philosophy of mind**. Massachusetts: Blackwell.
- Stigler, S.M. (1999). **Statistics on the table: the history of statistical concepts and methods**. London: Harvard University Press.
- Stillings, N.A., Weisler, S.E., Cahse, C.H., Feinstein, M.H., Garfield, J.L. & Rissland, E.L. (1995). **Cognitive science: an introduction**. Massachusetts: Bradford.
- Stix, G. (2005). Molecular treasure hunt. **Scientific American**, **292**(5), 88-91.
- Stock, W.A. (1994). Systematic coding for research synthesis. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp.125-138). New York: Russell Sage Foundation.
- Stott, D.H. (1987). Relating learning styles to learning capability. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.215-236). New York: Guilford Press.
- Strauman, T.J. (2001). Paradigms, promises, and the potential of clinical psychology. **Journal of clinical psychology**, **57**(9), 1125-1132.
- Strauss, A. & Corbin, J. (1998). **Basics of qualitative research: techniques and procedures for developing grounded theory**. London: Sage.
- Strauss, S. (2005). Teaching as a natural cognitive ability: implications for classroom practice and teacher education. In D.B. Pillemer & S.H. White (Ed.), **Developmental psychology and social change: research, history, and policy** (pp.368-388). Cambridge: Cambridge University Press.
- Strube, G. (2000). Generative theories in cognitive psychology. **Theory & psychology**, **10**(1), 117-125.
- Strube, M.J. (1985). Combining and comparing significance levels from nonindependent hypothesis tests. **Psychological bulletin**, **97**(2), 334-341.
- Strube, M.J. & Hartmann, D.P.(1983). Meta-analytic techniques, applications, and functions. **Journal of consulting and clinical psychology**, **51**(1), 14-27.
- Styles, I. (1999). The study of intelligence - the interplay between theory and measurement. In M. Anderson (Ed.), **The development of intelligence** (pp.19-42). Sussex: Psychology Press.
- Süß, H-M. & Beauducel, A. (2005). Faceted models of intelligence. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.313-332). London: Sage.
- Suen, H.K. (1990). **Principles of test theories**. New Jersey: Lawrence Erlbaum.
- Sufficiency statistics. (2006). Retrieved 6 March 2006 from http://en.wikipedia.org/wiki/Sufficiency_statistics
- Suizzo, M. (2000). The social-emotional and cultural contexts of cognitive development: Neo-Piagetian perspectives. **Child development**, **71**(4), 846-849.
- Super, C.M. (2005). The globalization of developmental psychology. In D.B. Pillemer & S.H. White (Ed.), **Developmental psychology and social change: research, history, and policy** (pp.11-33). Cambridge: Cambridge University Press.
- Suppe, F. (1993). What's wrong with the received view on the structure of scientific theories? In J.H. Fetzer (Ed.). **Foundations of philosophy of science: recent developments (pp.111-126)**. New York: Paragon.
- Suppe, F. (2001a). Axiomatization. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.9-11). Massachusetts: Blackwell.
- Suppe, F. (2001b). Craig's theorem. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.65-67). Massachusetts: Blackwell.
- Suppe, F. (2001c). Ramsey sentences. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.390-392). Massachusetts: Blackwell.
- Suppes, P.(2000). The plurality of science. In J. McErlean (Ed.), **Philosophies of science: from foundations to cotemporary issues** (pp.476-485). California: Wadsworth.

- Sutton, A. (1988). L.S. Vygotskii: the cultural-historical theory, national minorities and the zone of next development. In R.M. Gupta & P.Coxhead (Eds.), **Cultural diversity and learning efficiency: recent developments in assessment** (pp.89-117). New York: St Martin's Press.
- Suzuki, L.A., Ponterotto, J.G. & Meller, P.J. (2001). Multicultural assessment: trends and directions revisited. In L.A. Suzuki, J.G. Ponterotto & P.J. Meller (Eds.), **Handbook of multicultural assessment: clinical, psychological, and educational applications** (pp.569-574). San Francisco: Jossey-Bass.
- Suzuki, L.A., Ponterotto, J.G. & Meller, P.J. (Eds.), (2001). **Handbook of multicultural assessment: clinical, psychological, and educational applications**. San Francisco: Jossey-Bass.
- Suzuki, L.A. & Valencia, R.R. (1997). Race-ethnicity and measured intelligence. **American psychologist**, **52**(10), 1103-1114.
- Svinicki, M.D. (1998). A theoretical foundation for discovery learning. **Advances in physiology education**, **20**(1), s4-s7.
- Swanson, H.L. (1988). Toward a metatheory of learning disabilities. **Journal of learning disabilities**, **21**(4), 196 - 209.
- Swanson, H.L. (1995). Using the cognitive processing test to assess ability: development of a dynamic assessment measure. **School psychology review**, **24**(4), 672-693.
- Swanson, H.L. (2000). Swanson-cognitive processing test: review and applications. In C.S. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.71-107). New York: Elsevier.
- Swanson, H.L. (2005). Working memory, intelligence, and learning disabilities. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.409-429). London: Sage.
- Sylva, K. (2000). Psychological theory that "works" in the classroom. In L. Smith, J. Dockrell & P.Tomlinson (Eds.), **Piaget, Vygotsky and beyond: future issues for developmental psychology and education** (pp.60-64). London: Routledge.
- Szentágothai, J. (1989). The "brain-mind" relation: a pseudoproblem? In C. Blakemore & S. Greenfield (Eds.), **Mindwaves: thoughts on intelligence, identity and consciousness** (pp.323-336). Oxford: Basil Blackwell.
- Tayler, J. (1996). **Assessment of the predictive validity learning ability battery**. Unpublished M.A. dissertation, University of the Witwatersrand.
- Taylor, R.L. & Richards, S.B. (1990). Validity of the estimated learning potential and other measures of learning potential. **Perceptual and motor skills**, **71**(1), 225-229.
- Taylor, T.R. (1987). **The future of cognitive assessment**. Pretoria: Human Sciences Research Council.
- Taylor, T.R. (1994). A review of three approaches to cognitive assessment, and a proposed integrated approach based on a unifying theoretical framework. **South African journal of psychology**, **24**(4), 184-193.
- Tegmark, M. (2003). Parallel universes: not just a staple of science fiction, other universes are a direct implication of cosmological observations. **Scientific American**, **288**(5), 40-51.
- Tellegen, P.J. & Laros, J.A. (1993). The Snijders-Oomen nonverbal intelligence tests: general intelligence tests of tests for learning potential? In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.267-283). Amsterdam: Swets & Zeitlinger.
- Templer, D.I. & Arikawa, H. (2006). Temperature, skin color, per capita income, and IQ: An international perspective. **Intelligence**, **34**(2), 121-139.
- te Nijenhuis, J., Van Vianen, A. & Van der Flier, H. (2005). Score Gains on *g*-Loaded Tests: No *g*. Manuscript submitted for publication.
- Teplov, B.M. (1961). Typological properties of the nervous system and their psychological manifestations. In N. O'Conner (Ed.), **Recent Soviet psychology** (pp.21-51). Oxford: Pergamon.
- Terman, L.M. (1921). Intelligence and its measurement. **Journal of educational psychology**, **12**(3), 127-133.
- Tharp, R. & Gallimore, R. (1988). **Rousing minds to life**. Cambridge: Cambridge University Press.
- Thelen, E. (1994). Self-organization in developmental processes: can systems approaches work? In M.H. Johnson, (Ed). **Brain development and cognition: a reader** (pp.555-591). Oxford: Blackwell.
- Thorndike, E. L. (1921). Intelligence and its measurement. **Journal of educational psychology**, **12**(3), 124-127.
- Thorndike, E.L. (1981). Laws of effect and exercise. In W.S. Sahakian (Ed), **History of psychology: a source book in systematic psychology** (pp.193-194). Illinois: F.E. Peacock.
- Thorndike, R.M. (1997). The early history of intelligence testing. In D.P.Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp.3-16). New York: Guilford.
- Thorne, K. (1995). **Black holes and time warps: Einstein's outrageous legacy**. London: Papermac.
- Tiles, M. (2001). Philosophy of technology. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.483-491). Massachusetts: Blackwell.
- Tipler, F.J. (1996). **The physics of immortality: modern cosmology, god and the resurrection of the dead**. New York: Pan.
- Tissink, J., Hamers, J.H.M. & Van Luit, J.E.H. (1993). Learning potential tests with domain-general and domain-specific tasks. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.223-266). Amsterdam: Swets & Zeitlinger.
- Tobach, E. (1996). Is there a comparative psychology in the future of the former member nations of the USSR? In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.297-311). London: Greenwood Press.
- Todes, D. (1984). Biological psychology and the tsarist censor: the dilemma of scientific development. **Bulletin of the history of medicine**, **58**, 529-544.

- Tolman, C. (1987). Dialectical materialism as psychological metatheory. In H.J. Stam, T.B. Rogers, & K.J. Gergen. (Eds.). **The analysis of psychological theory: metapsychological perspectives** (pp.211-229). Cambridge: Hemisphere.
- Tomasello, M. (2001). **The cultural origins of human cognition**. Massachusetts: Harvard University Press.
- Toomela, A. (2000). Activity theory is a dead end for cultural-historical psychology. **Culture and psychology**, 6(3), 353-364.
- Torff, B. & Gardner, H. (1999). The vertical mind - the case for multiple intelligences. In M. Anderson (Ed.), **The development of intelligence** (pp.139-159). Sussex: Psychology Press.
- Toren, C. (1993). Making history: the significance of childhood cognition for a comparative anthropology of mind. **Man**, 28(3), 461-477.
- Toulmin, S. & Leary, D.E. (1985). The cult of empiricism in psychology, and beyond. In S. Koch & D.E. Leary (Eds.). **A century of psychology as science** (pp.594-617). New York: McGraw-Hill.
- Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: surprising insights from Bayes's theorem. **Psychological review**, 110(3), 526-535.
- Trout, J.D. (2001a). Measurement. In W.H. Newton-Smith (Ed.). **A companion to the philosophy of science** (pp.265-276). Massachusetts: Blackwell.
- Trout, J.D. (2001b). Paradoxes of confirmation. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.53-55). Massachusetts: Blackwell.
- Tryphon, A. & Vonèche, J. (Eds.), (1996). **Piaget-Vygotsky: the social genesis of thought**. Sussex: Psychology Press.
- Tryphon, A. & Vonèche, J. (1996b). Introduction. In A. Tryphon & J. Vonèche (Eds.), **Piaget-Vygotsky: the social genesis of thought** (pp.1-10). Sussex: Psychology Press.
- Tugaybayeva, B.N. (1996). Comments on the history of psychoanalysis in Russia. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.253-266). London: Greenwood Press.
- Tuomela, R. (1973). **Theoretical concepts**. New York: Springer-Verlag.
- Turnbull, W.W. (1979). Intelligence testing in the year 2000. **Intelligence**, 3, 275-282.
- Turner, J.H. (1982). **The structure of sociological theory**. Illinois: The Dorsey Press.
- Turner, J.H. (Ed.). (1989). **Theory building in sociology: assessing theoretical cumulation**. Newbury Park: Sage.
- Turner, J.H. (1993). A "pure theory" research program on microdynamics. In J. Berger & M. Zelditch (Eds.). **Theoretical research programs: studies in the growth of theory** (pp.104-135). California: Stanford University Press
- Turner, M. (1967). **Psychology and the philosophy of science**. New York: Appleton-Century-Crofts.
- Turner, M. (2001). Toward the founding of cognitive social science. **Chronicle of higher education**, 48(6), B11-B12.
- Turner, S.P. & Roth, P.A. (2003). Introduction. Ghosts and the machine: issues of agency, rationality, and scientific methodology in contemporary philosophy of social science. In S.P. Turner & P.A. Roth (Eds.). **The Blackwell guide to the philosophy of the social sciences** (pp.1-17). Oxford: Blackwell.
- Turner, S.P. & Roth, P.A. (Eds.). (2003). **The Blackwell guide to the philosophy of the social sciences**. Oxford: Blackwell.
- Tweney, R.D. (1989). A framework for the cognitive psychology of science. In B. Gholson, W.R. Shadish, R.A. Neimeyer, & A.C. Houts, (Eds.). **Psychology of science: contributions to metascience** (pp.342-366). Cambridge: Cambridge University Press.
- Tyron, W.W. (1996). Measurement units and theory construction. **The journal of mind and behavior**, 17(3), 213-228.
- Tzuriel, D. (1996). Inferential thinking modifiability in young socially disadvantaged and advantaged children. In M. Luther, E. Cole & P. Gamlin (Eds.), **Dynamic assessment for instruction: from theory to application** (pp.146-157). Ontario: Captus Press.
- Tzuriel, D. (2000a). The Seria-Think instrument; development of a dynamic test for young children. **School psychology international**, 21(2), 177-194.
- Tzuriel, D. (2000b). Developmental perspectives of mediated learning experience theory. In A. Kozulin & Y. Rand (Eds.), **Experience of mediated learning: an impact of Feuerstein's theory in education and psychology** (pp.217-239). Amsterdam: Pergamon.
- Tzuriel, D. (2000c). The cognitive modifiability battery (CMB): assessment and intervention. In C. Lidz & J.G. Elliott (Eds.), **Dynamic assessment: prevailing models and applications** (pp.375-406). New York: Elsevier.
- Tzuriel, D. (2000d). Dynamic assessment of young children: educational and intervention perspectives. **Educational psychology review**, 12(4), 385-435.
- Tzuriel, D. (2001). **Dynamic assessment of young children**. New York: Kluwer.
- Tzuriel, D. (2002). Cognitive education: the menace and the hope. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.355-363). Amsterdam: JAI.
- Tzuriel, D. & Feuerstein, R. (1992). Dynamic group assessment for prescriptive teaching: differential effects of treatments. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.187-206). New York: Springer-Verlag.
- Tzuriel, D. & Haywood, H.C. (1992). The development of interactive-dynamic approaches to assessment of learning potential. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.3-37). New York: Springer-Verlag.
- Tzuriel, D. & Klein, P.S. (1987). Assessing the young child: children's analogical thinking modifiability. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.268-287). New York: Guilford Press.

- Tzuriel, D., Samuels, M.T. & Feuerstein, R. (1988). Non-intellective factors in dynamic assessment. In R.M. Gupta & P.Coxhead (Eds.), **Cultural diversity and learning efficiency: recent developments in assessment** (pp.141-163). New York: St Martin's Press.
- Uglov, J. (2003). *The lunar men: the friends who made the future*. London: Faber and Faber.
- Ukrainetz, T.W., Harpell, S., Walsh, C. & Colye, C. (2000). A preliminary investigation of dynamic assessment with native American kindergartners. **Language, speech and hearing services in schools**, **31**(2), 142-154.
- Undheim, J.O. (1987). Diagnostic and instructional implications of some recent cognitive ability research within the psychometric tradition. In E. De Corte, H. Lodewijks, R. Parmentier & P.Span (Eds.), **Learning and instruction: European research in an international context volume 1** (pp.89-97). Oxford: Pergamon
- Ushakova, T.N. (1997). Pavlov's theory and Russian psychology. **European psychologist**, **2**(2), 97-101.
- Utley, C.A., Haywood, H.C. & Masters, J.C.(1992). Policy implications of psychological assessment of minority children. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.445-469). New York: Springer-Verlag.
- Valencia, R.R. (Ed.), (1997). **The evolution of deficit thinking: educational thought and practice**. London: Falmer Press.
- Valencia, R.R. (1997a). Introduction. In R.R. Valencia (Ed.), **The evolution of deficit thinking: educational thought and practice** (pp.ix-vii). London: Falmer Press.
- Valencia, R.R. (1997b). Conceptualizing the notion of deficit thinking. In R.R. Valencia (Ed.), **The evolution of deficit thinking: educational thought and practice** (pp.1-12). London: Falmer Press.
- Valencia, R.R., Henderson, R.W. & Rankin, R.J. (1985). Family status, family constellation, and home environmental variables as predictors of cognitive performance of Mexican American children. **Journal of educational psychology**, **77**(3), 323-331.
- Valencia, R.R. & Suzuki, L.A. (2001). **Intelligence testing and minority students. Foundations, performance factors, and assessment issues**. California: Sage.
- Valsiner, J. (1988). **Developmental psychology in the Soviet Union**. Sussex: Harvester Press.
- Valsiner, J. (1996). Social utopias and knowledge construction in psychology. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.70-84). London: Greenwood Press.
- Van Aswegen, M. (1997). **The standardisation of a learning potential battery for the selection of poorly qualified employees**. Unpublished M.Com. dissertation, University of Pretoria.
- Van de Vijver, F.J.R. (1991). Group differences in structuring tests. In S.H. Irvine & S.E. Newstead (Eds.), **Intelligence and cognition: contemporary frames of reference** (pp.397-417). Dordrecht: Martinus Nijhoff.
- Van de Vijver, F.J.R. (1993). Learning potential assessment from a cross-cultural perspective. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp.313-340). Amsterdam: Swets & Zeitlinger.
- Van de Vijver, F.J.R. & Jongmans, M.J. (2002). Measuring learning potential in a cross-cultural perspective: problems and prospects. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.211-226). Amsterdam: JAI.
- Van der Aalsvoort, G.M. & Lidz, C.S. (2002). Reciprocity in dynamic assessment in classrooms: taking contextual influences on individual learning into account. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.11-144). Amsterdam: JAI.
- Van der Aalsvoort, G.M., Resing, W.C.M. & Ruijsenaars, A.J.J.M. (2002). **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology. Advances in cognition and educational practice volume 7**. Amsterdam: JAI-Elsevier Science.
- Van der Aalsvoort, G.M., Resing, W.C.M. & Ruijsenaars, A.J.J.M. (2002). New developments in learning potential assessment and cognitive training. In G.M. Van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), **Learning potential assessment and cognitive training: actual research and perspectives in theory building and methodology volume 7** (pp.1-8). Amsterdam: JAI.
- Van der Linden, W.J. (1994). Optimum design in item response theory: test assembly and item calibration. In G.H Fischer & D. Laming (Eds.), **Contributions to mathematical psychology, psychometrics, and methodology** (pp.305-318). NewYork: Springer-Verlag.
- Van der Maas, H.L.J. & Molenaar, P.C.M. (1992). Stagemwise cognitive development: an application of catastrophe theory. **Psychological review**, **99**(3), 395-471.
- Van der Veer, R. (1996). Structure and development. Reflections by Vygotsky. In A. Tryphon & J. Vonèche (Eds.), **Piaget-Vygotsky: the social genesis of thought** (pp.45-56). Sussex: Psychology Press.
- Van der Veer, R. (2000). Back to classic Russian psychology. **Contemporary psychology**, **45**(4), 418-421.
- Van der Veer, R. & Valsiner, J. (1991). **Understanding Vygotsky: a quest for synthesis**. Oxford: Blackwell.
- Van der Veer, R. & Valsiner, J. (1994). Introduction. Reading Vygotsky: from fascination to construction. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.1-9). Oxford: Blackwell.
- Van der Veer, R. & Valsiner, J. (Eds.), (1994). **The Vygotsky reader**. Oxford: Blackwell.



- Van Ede, D.M. (1995). Adapting the metamemory in adulthood (MIA) questionnaire for cross-cultural application in South Africa. **South Africa journal of psychology**, 25(2), 74-80.
- Van Eeden, R., De Beer, M. & Coetzee, C.H. (2001). Cognitive ability, learning potential, and personality traits as predictors of academic achievement by engineering and other science and technology students. **South African journal of higher education**, 15 (1), 171-179.
- Van Geert, P.(2000). The dynamics of general developmental mechanisms: from Piaget and Vygotsky to dynamic systems models. **Current directions in psychological science** (9)2, 64 - 68.
- Van Hezewijk, R. (2003). Psychology and evolutionary biology: causal analysis, evidence, and nomothetic laws. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.405-415). Ontario: Captus University Publications.
- Van Lil, J.B. (1987). **'n Kritiese analise van die wetenskapsfilosofie van die sielkunde**. Unpublished DPhil thesis, University of South Africa.
- Van Niekerk, H. (Ed.), (1990). **Cognitive development in the Southern African context: papers of the seminar on cognitive development, 1 November 1990**. Pretoria: Human Sciences Research Council.
- Van Schaik, C. (2006). Why are some animals so smart? **Scientific American**, 294(4), 64-71.
- Van Schuur, W.H. (1997). Nonparametric IRT models for dominance and proximity data. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.313-331). London: Ablex.
- Van Strien, P.J. (1987). Model disciplines, research traditions, and the theoretical unification of psychology. In W.J. Baker., M.E. Hyland., H. Van Rappard. & A.W. Staats (Eds.), **Current issues in theoretical psychology: selected/edited proceedings of the founding conference of the international society for theoretical psychology held in Plymouth, U.K., 30 August - 2 September, 1985** (pp.333-344). Amsterdam: North-Holland.
- Vasilev, D.D. (1996). The crisis in psychology and its Russian dimensions. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.213-228). London: Greenwood Press.
- Velleman, P.& Wilkinson, L. (1994). Nominal, ordinal, interval, and ratio typologies are misleading. In I. Borg & P.Mohler (Eds.), **Trends and perspectives in empirical social research** (pp.161-177). Berlin: Walter De Gruyter.
- Vernon, P.A. (1986). Relationships between speed-of-processing, personality, and intelligence. In C. Bagley & G.K. Verma (Eds.), **Personality, cognition and values** (pp.132-167). London: Macmillan.
- Vogler, G.P.& Fulker, D.W. (1988). Human behaviour genetics. In J.R. Nesselroade & R.B. Cattell (Eds.), **Handbook of multivariate experimental psychology** (pp.475-503). New York: Plenum.
- Von Bertalanffy, L. (1970). Comments. In J.R. Royce (Ed.). **Toward unification in psychology: the first Banff conference on theoretical psychology** (pp.39-41). Toronto: University of Toronto Press.
- Von der Malsburg, C. (1987). Synaptic plasticity as basis of brain organization. In J-P.Changeux & M. Konishi (Eds.), **The neural and molecular bases of learning** (pp.411-432). Chichester: John Wiley.
- Von Foerster, H. (1985). Apropos epistemologies. **Family process**, 24(4), 517-521.
- Von Hirschfeld, S. & Downs, S. (1992). Testing for trainability. **People dynamics**, 10(9), 24-26.
- Von Neumann, J. (1983). The formalist foundations of mathematics. In P.Benacerraf & H. Putnam (Eds.). (1983). **Philosophy of mathematics: selected readings** (pp.61-65). Cambridge: Cambridge University Press.
- Voracek, M. (2006). Phlogiston, fluid intelligence, and the Lynn-Flynn effect. **Behavioral and brain sciences**, 29, 142-143.
- Vosniadou, S. (1996a). Towards a revised cognitive psychology for new advances in learning and instruction. **Learning and instruction**, 6(2), 95-109.
- Vosniadou, S. (1996b). Learning environments for representational growth and cognitive flexibility. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), **International perspectives on the design of technology-supported learning environments** (pp.13-23). New Jersey: Lawrence Erlbaum.
- Vosniadou, S. De Corte, E. Glaser, R. & Mandl, H. (Eds.), (1996). **International perspectives on the design of technology-supported learning environments**. New Jersey: Lawrence Erlbaum.
- Vye, N.J., Burns, M.S., Delclos, V.R. & Bransford, J.D. (1987). A comprehensive approach to assessing intellectually handicapped children. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp.237-359). New York: Guilford Press.
- Vygotsky, L. S. (1978). Mind in society: the development of higher psychological processes. In M. Cole, V. John-Steiner, S. Scribner & E. Soubberman (Eds.), Vygotsky's **"Mind in society: the development of higher psychological processes"** (pp.17-119). Cambridge: Harvard University Press.
- Vygotsky, L. S. (1981a). The instrumental method in psychology. In J.V. Wertsch (Ed.), **The concept of activity in Soviet psychology** (pp.136-143). New York: M.E. Sharpe.
- Vygotsky, L.S. (1981b). The genesis of higher mental functions. In J.V. Wertsch (Ed.), **The concept of activity in Soviet psychology** (pp.147-188). New York: M.E. Sharpe.
- Vygotsky, L.S. (1981c). The development of higher forms of attention in childhood. In J.V. Wertsch (Ed.), **The concept of activity in Soviet psychology** (pp.189-240). New York: M.E. Sharpe.
- Vygotsky, L.S. (1994a). Principles of social education for deaf and dumb children in Russia. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.19-26). Oxford: Blackwell.



- Vygotsky, L.S. (1994b). The methods of reflexological and psychological investigation. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.27-45). Oxford: Blackwell.
- Vygotsky, L.S. (1994c). The problem of the cultural development of the child. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.27-45). Oxford: Blackwell.
- Vygotsky, L.S. (1994d). The socialist alteration of man. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.175-184). Oxford: Blackwell.
- Vygotsky, L.S. (1994e). Thinking and concept formation in adolescence. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.175-184). Oxford: Blackwell.
- Vygotsky, L.S. (1994f). The problem of the environment. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.338-354). Oxford: Blackwell.
- Vygotsky, L.S. (1994g). The development of academic concepts in school aged children. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.355-370). Oxford: Blackwell.
- Vygotsky, L.S. & Luria, A. (1994). Tool and symbol in child development. In R. Van der Veer & J. Valsiner (Eds.), **The Vygotsky reader** (pp.99-174). Oxford: Blackwell.
- Wagner, C. (2003). **Placing psychology: a critical exploration of research methodology curricula in the social sciences**. Unpublished DPhil thesis, University of Pretoria.
- Wahlsten, D. & Gottlieb, G. (1997). The invalid separation of effects of nature and nurture: lessons from animal experimentation. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.163-192). Cambridge: Cambridge University Press.
- Wainwright, Mark.A., Wright,M.J., Luciano, M., Geffen,G.M. & Martin,N.G. (2005). Multivariate genetic analysis of academic skills of the Queensland core skills test and IQ highlight the importance of genetic *g*. **Twin research and human genetics**, 8(6), 602-608.
- Wakefield, J. (2001). A mind for consciousness. **Scientific American**, 285(1), 26-27.
- Waldman, I.D. (1997). Unresolved questions and future directions in behaviour-genetic studies of intelligence. In R.J. Sternberg & E.L. Grigorenko (Eds.), **Intelligence, heredity and environment** (pp.552-570). Cambridge: Cambridge University Press.
- Wallis, T. (2004). Psychological tests do not measure what they claim to measure: a re-evaluation of the concept of construct validity. **South African journal of psychology**, 34(1), 101-112.
- Walliser, B. & Zwirn, D. (2002). Can Bayes' rule be justified by cognitive rationality principles? **Theory and decision**, 53(2), 95-135.
- Walters, J.M. & Gardner, H. (1986). The theory of multiple intelligences: some issues and answers. In R.J. Sternberg & R.K. Wagner (Eds.), **Practical intelligence: nature and origins of competence in the everyday world** (pp.163-182). Cambridge: Cambridge University Press.
- Wang, H. (1974). **From mathematics to philosophy**. London: Routledge and Kegan Paul.
- Wang, H. (1983). The concept of set. In P. Benacerraf & H. Putnam (Eds.), **Philosophy of mathematics: selected readings** (pp.530-570). Cambridge: Cambridge University Press.
- Wang, H. (1995). Can minds do more than brains? In J. Brockman & K. Matson (Eds.), **How things are: a science tool-kit for the mind** (pp.145-149). London: Weidenfeld and Nicolson.
- Wang, W-C. & Wilson, M. (2005). The Rasch testlet model. **Applied psychological measurement**, 29(2), 126-149.
- Wang, W-C., Wilson, M. & Adams, R.J. (1997). Rasch models for multidimensionality between items and within items. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.139-155). London: Ablex.
- Wang, Y. (2002). On cognitive informatics. **Proceedings of the First IEEE international conference on cognitive informatics (ICCI '02)**, 34-42.
- Wang, Y. (2003a). Cognitive informatics: a new transdisciplinary research field. **Brain and Mind**, 4, 115-127.
- Wang, Y. (2003b). On cognitive informatics. **Brain and Mind**, 4, 151-167.
- Wang, Y. & Liu, D. (2003). On information and knowledge representation in the brain. **Proceedings of the Second IEEE international conference on cognitive informatics (ICCI '03)**, 26-31.
- Wang, Y. & Wang, Y. (2002). Cognitive models of the brain. **Proceedings of the First IEEE international conference on cognitive informatics (ICCI '02)**, 259-269.
- Wang, Y., Patel, S., Patel, D. & Wang, Y. (2003). A layered reference model of the brain. **Proceedings of the Second IEEE international conference on cognitive informatics (ICCI '03)**, 7-17.
- Warwick, K. (1998). **In the mind of the machine: the breakthrough in artificial intelligence**. London: Arrow.
- Watanabe, T. (2003). Significance and fundamental problems of the philosophy of psychology with some of its difficulties in Japan. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.124-133). Ontario: Captus University Publications.
- Watkins, M.L. & Mauer, K.F. (1994). The performance values of white and black managers in South Africa. **South African journal of psychology**, 24(2), 78-85.
- Watson, P. (2005). Not written in stone. **New Scientist**, 187(2514), 42-44.

- Weaver, I. C. G., Cervoni, N., Champagne, F.A., D'Alessio, A. C., Sharma, S., Seckl, J. R., Dymov, S., Szyf, M. & Meaney, M.J. (2004). Epigenetic programming by maternal behavior. **Nature Neuroscience**, **7** (8), 847-854.
- Weber, R.L. (1980). **Pioneers of science: Nobel prize winners in physics**. London: The institute of physics.
- Weber, R.P.(1986). **Basic content analysis**. Beverly Hills: Sage.
- Weems, C.F. (1998). Theoretical psychology: subdiscipline or metadiscipline? **American psychologist**, **53**(1), 68-69.
- Weinberg, R.A. (1989). Intelligence and IQ: landmark issues and great debates. **American psychologist**, **44**(2), 98-104.
- Weinert, F.E. (1987). Developmental processes and instruction. In E. De Corte, H. Lodewijks, R. Parmentier & P.Span (Eds.), **Learning and instruction: European research in an international context volume 1** (pp.1-17). Oxford: Pergamon.
- Weinstein, D.A. & Weinstein, M.A. (1992). The postmodern discourse of metatheory. In G. Ritzer (Ed.). **Metatheorizing** (pp.135-150). Newbury Park: Sage.
- Wellman, H.M. & Gelman, S.A. (1992). Cognitive development: foundational theories of core domains. **Annual review of psychology**, **43**, 337-375.
- Wenke, D., Frensch, P.A. & Funke, J. (2005). Complex problem solving and intelligence: empirical relation and causal direction. In R.J. Sternberg & J.E. Pretz (Eds.), **Cognition and intelligence: identifying the mechanisms of the mind** (pp.160-187). Cambridge: Cambridge University Press.
- Wertsch, J.V. (1981). The concept of activity in Soviet psychology: an introduction. In J.V. Wertsch (Ed.), **The concept of activity in Soviet psychology** (pp.3-36). New York: M.E. Sharpe.
- Wertsch, J.V. (Ed.), (1981). **The concept of activity in Soviet psychology**. New York: M.E. Sharpe.
- Wertsch, J.V. (1996). The role of abstract rationality in Vygotsky's image of mind. In A. Tryphon & J. Vonèche (Eds.), **Piaget-Vygotsky: the social genesis of thought** (pp.25-44). Sussex: Psychology Press.
- Wertsch, J.V. (1998). Mediated action. In W. Bechtel & G. Graham, (Eds.). **A companion to cognitive science** (pp.518-525). Massachusetts: Blackwell.
- Wertsch, J.V. (2000). Gal'perin's elaboration of Vygotsky. **Human development**, **43**, 103-106.
- Wertsch, J.V. & Tulviste, P. (1992). L.S. Vygotsky and contemporary developmental psychology. **Developmental psychology**, **28**(4), 548-557.
- Weyant, R.G. (1987). Artificial intelligence: what if anything, can machines teach us about human psychology? In H.J. Stam, T.B. Rogers, & K.J. Gergen. (Eds.). **The analysis of psychological theory: metapsychological perspectives** (pp.281-296). Cambridge: Hemisphere.
- White, H.D. (1994). Scientific communication and literature retrieval. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp41-55). New York: Russell Sage Foundation.
- White, S.H. & Pillemer, D.B. (2005). Introduction: what kind of science is developmental psychology? In D.B. Pillemer & S.H. White (Ed.), **Developmental psychology and social change: research, history, and policy** (pp.1-7). Cambridge: Cambridge University Press.
- Whitely, S.E. (1980). Latent trait models in the study of intelligence. **Intelligence**, **4**, 97-132.
- Whitley, B.E. (2002). **Principles of research in behavioural science**. Boston: McGraw Hill.
- Whitt, L.A. (1988). Conceptual dimensions of theory appraisal. **Studies in history and philosophy of science**, **19**(4), 517-529.
- Wiedl, K.H. (May, 2002). **Dynamic testing - a comprehensive model and current fields of application**. Paper presented at the European regional conference, International Association for cognitive education (IACE) in collaboration with Associazione Connessioni, Universita di Rimini, Italy.
- Wiedl, K.H. (2003). Dynamic testing: a comprehensive model and current fields of application. **Journal of cognitive education and psychology**, **3**(1), 93-119.
- Wiedl, K.H., Guthke, J. & Wingenfeld, S. (1995). Dynamic assessment in Europe: historical perspectives. In J.S. Carlson (Ed.), **Advances in cognition and educational practice vol 3: European contributions to dynamic assessment** (pp.185-208). London: JAI Press.
- Wiedl, K.H. & Schlittke, H. (1995). Dynamic assessment of selective attention in schizophrenic subjects: the analysis of intraindividual variability of performance. In J.S. Carlson (Ed.), **Advances in cognition and educational practice vol 3: European contributions to dynamic assessment** (pp.145-183). London: JAI Press.
- Wilczek, F. (2002). A piece of magic: the Dirac equation. In G. Farmelo (Ed.). **It must be beautiful: great equations of modern science** (pp.102-130). London: Granta.
- Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.373-392). London: Sage.
- Wilhelm, O. & Engle, R.W. (2005). Intelligence: a diva and a workhorse. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.1-9). London: Sage.
- Wilhelm, O. & Engle, R.W. (Eds.), **Handbook of understanding and measuring intelligence**. London: Sage.
- Wilkinson, L. & the Task Force on Statistical Inference: APA Board of Scientific Affairs. (1999). Statistical methods in psychology journals: guidelines and expectations. **American psychologist**, **54**(8), 594-604.
- Wille, U. (1994). An axiomatisation for bilinear models. In G.H Fischer & D. Laming (Eds.), **Contributions to mathematical psychology, psychometrics, and methodology** (pp.239-247). NewYork: Springer-Verlag.
- Willer, D. & Markovsky, B. (1993). Elementary theory: its development and research program. In J. Berger & M. Zelditch (Eds.). **Theoretical research programs: studies in the growth of theory** (pp.323-363). California: Stanford University Press

- Willerman, L., Schultz, R., Rutledge, J.N. & Bigler, E.D. (1994). Brain structure and cognitive function. In C.R. Reynolds (Ed). **Cognitive assessment: a multidisciplinary perspective** (pp.35-55). New York: Plenum Press.
- Williams, M.R. (1997). **A history of computing technology**. California: IEEE Computer Society Press.
- Williams, R.H., Zimmerman, D.W., Zumbo, B.D. & Ross, D. (2003). Charles Spearman: British behavioral scientist. **Human nature review**, **3**, 114-118.
- Williams, R.N. (2003). Psychology in the breach: the importance of being earnest. In N. Stephenson, H.L. Radtke, R. Jorna & H.J. Stam (Eds.), **Theoretical psychology: critical contributions** (pp.2-10). Ontario: Captus University Publications.
- Wilson, E.A. (1996). Projects for a scientific psychology: Freud, Derrida, and connectionist theories of cognition. **Differences: a journal of feminist cultural studies**, **8**(3), 21-52.
- Wilson, E.O. (1999). **Consilience: the unity of knowledge**. London: Abacus.
- Wilson, M. (1985). What can theory tell us about observation? In P.M. Churchland & C.A. Hooker (Eds.), **Images of science: essays on realism and empiricism, with a reply from Bas C. Van Fraassen** (pp.222-242). Chicago: The university of Chicago Press.
- Wilson, M. (1989). Saltus: a psychometric model of discontinuity in cognitive development. **Psychological bulletin**, **105**(2), 276-289.
- Wilson, M., Engelhard, G. & Draney, K. (Eds.), **Objective measurement: theory into practice volume 4**. London: Ablex.
- Wilson, M. & Draney, K. (1997). Partial credit in a developmental context: the case for adopting a moisture model approach. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.333-350). London: Ablex.
- Wilson, M. & Wang, W-C. (1995). Complex composites: issues that arise in combining different modes of assessment. **Applied psychological measurement**, **19**(1), 51-71.
- Wilson, R.S. & Matheny, A.P. (1983). Mental development: family environment and genetic influences. **Intelligence**, **7**, 195-215.
- Wimsatt, W.C. (1986). Heuristics and the study of human behavior. In D.W. Fiske & R.A. Shweder (Eds.), **Metatheory in social science: pluralisms and subjectivities** (pp.293-315). Chicago: University of Chicago Press.
- Winch, P.(1970). **The idea of a social science and its relation to philosophy**. London: Routledge and Kegan Paul.
- Wittmann, W.W. (2005). Group differences in intelligence and related measures. In O. Wilhelm & R.W. Engle (Eds.), **Handbook of understanding and measuring intelligence** (pp.223-239). London: Sage.
- Wolf, F.M. (1987). **Meta-analysis: quantitative methods for research synthesis**. Beverly Hills: Sage.
- Wolman, B. (Ed). (1973). **Handbook of general psychology**. New Jersey: Prentice-Hall.
- Woodcock, R.W. (1998). Extending GfGc theory into practice. J.J. McArdle & R.W. Woodcock (Eds.), **Human cognitive abilities in theory and practice** (pp.137-156). New Jersey: Lawrence Erlbaum.
- Woodcock, R.W. (1999). What can Rasch-based scores convey about a person's test performance? In S.E. Embretson & S.L. Hershberger (Eds.), **The new rules of measurement: what every psychologist and educator should know** (pp.105-127). New Jersey: Lawrence Erlbaum.
- Woodfield, A. (2001). Teleological explanation. In W.H. Newton-Smith (Ed.), **A companion to the philosophy of science** (pp.492-494). Massachusetts: Blackwell.
- Woods, J. (2003). **Paradox and paraconsistency: conflict resolution in the abstract sciences**. Cambridge: Cambridge University Press.
- Woodward, J. (2002). Explanation. In P.Machamer & M. Silberstein (Eds.), **The Blackwell guide to the philosophy of science** (pp.37-54). Massachusetts: Blackwell.
- Woodward, W.R. & Clark, S.C. (1996). The reflection of Soviet psychology in East German psychological practice. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.236-250). London: Greenwood Press.
- Woolf, H. (Ed.), (1961). **Quantification: a history of the meaning of measurement in the natural and social sciences**. New York: Bobbs-Merrill.
- Woolley, B. (1993). **Virtual worlds: a journey in hype and hyperreality**. London: Penguin.
- Worrall, J. (2002). Philosophy of science: classic debates, standard problems, future prospects. In P.Machamer & M. Silberstein (Eds.), **The Blackwell guide to the philosophy of science** (pp.18-136). Massachusetts: Blackwell.
- Wortis, J. (1987). The history of intelligence. **Biological psychiatry**, **22**, 1425-1428.
- Wortman, P.M. (1994). Judging research quality. In H. Cooper & L.V. Hedges (Eds.), **The handbook of research synthesis** (pp97-109). New York: Russell Sage Foundation.
- Wozniak, R.H. (1996). Qu'est-ce que l'intelligence? Piaget, Vygotsky, and the 1920s crisis in psychology. In A. Tryphon & J. Vonèche (Eds.), **Piaget-Vygotsky: the social genesis of thought** (pp.11-24). Sussex: Psychology Press.
- Wright, B.D. (1980). Foreward. In G. Rasch. **Probabilistic models for some intelligence and attainment tests** (pp.ix - xix). Chicago: The University of Chicago Press
- Wright, B.D. (1997a). Rasch factor analysis. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective measurement: theory into practice volume 4** (pp.113-137). London: Ablex.
- Wright, B.D. (1997b). A history of social science measurement. **Educational measurement, issues and practice**, **16**(4), 33-45, 52.



- Wright, B.D. (1999). Fundamental measurement for psychology. In S.E. Embretson & S.L. Hershberger (Eds.), **The new rules of measurement: what every psychologist and educator should know** (pp.65-104). New Jersey: Lawrence Erlbaum.
- Wright, B.D. & Stone, M.H. (1979). **Best test design**. Chicago: Mesa Press.
- Xiang, C., Andersen, H. & Barker, P.(1998). Kuhn's theory of scientific revolutions and cognitive psychology. **Philosophical psychology**, **11**(1), 5-28.
- Yanai, H., Okada, A., Shigemasu, K., Kano, Y. & Meulman, J.J. (Eds.), (2003). **New developments in psychometrics. Proceedings of the International Meeting of the Psychometric Society 2001**. Tokyo: Springer.
- Yao, Y.Y. (2004). Concept formation and learning: a cognitive informatics perspective. **Proceedings of the Third IEEE international conference on cognitive informatics (ICCI'04)**, 42-51.
- Yaroshevsky, M.G. (1996). Marxism in Soviet psychology: the social role of Russian science. In V.A. Koltsova, Y.N. Oleinik, A.R. Gilgen & C.K. Gilgen (Eds.), **Post-Soviet Perspectives on Russian Psychology** (pp.3161-186). London: Greenwood Press.
- Youngson, R. (1998). **Scientific blunders: a brief history of how wrong scientists can sometimes be**. London: Robinson.
- Yussen, S.R. (1985). The role of metacognition in contemporary theories of cognitive development. In D.L. Forrest-Pressley, G.E. MacKinnon & T.G. Waller (Eds.), **Metacognition, cognition and human performance: theoretical perspectives volume 1** (pp.253-283). Orlando: Academic Press.
- Zeidner, M., Matthews, G. & Roberts, R.D. (2004). Intelligence theory, assessment, and research: the Israeli experience. In R.J. Sternberg (Ed.), **International handbook of intelligence** (pp.212-247). Cambridge: Cambridge University Press.
- Zhao, S. (1991). Metatheory, metamethod, meta-data-analysis: what, why, and how? **Sociological perspectives**, **34**(3), 377-390.
- Zigler, E. (1998). Intelligence: a developmental approach. In R.J. Sternberg & D.K. Detterman (Eds.), **What is intelligence? Contemporary viewpoints on its nature and definition** (pp.149-152). New Jersey: Ablex.
- Zimmer, C. (2003). **Parasite rex: inside the bizarre world on nature's most dangerous creatures**. London: Arrow.
- Zimmer, C. (2005a). The neurobiology of the self. **Scientific American**, **293**(5), 92-101.
- Zimmer, C. (2005b). **Soul made flesh: how the secrets of the brain were uncovered in seventeenth-century England**. London: Arrow.
- Zinchenko, V. (1982). Preface. In K. Levitin (Ed.), **One is not born a personality: profiles of Soviet education psychologists** (pp.7-12). Moscow: Progress.
- Zinchenko, V. & Gordon, V.M. (1981). Methodological problems in the psychological analysis of activity. In J.V. Wertsch (Ed.), **The concept of activity in Soviet psychology** (pp.73-133). New York: M.E. Sharpe.
- Zolezzi, S.A. (1992). **Alternative selection measures for university undergraduate admissions**. Unpublished M.Ed. dissertation, University of the Witwatersrand.
- Zolezzi, S.A. (1995). **The effectiveness of dynamic assessment as an alternative aptitude testing strategy**. Unpublished D.Phil. dissertation, University of South Africa.

Appendix 1 Meta-analysis of South African dynamic assessment research

Section A Meta-analysis

1. Introduction

This meta-analysis has two aims: the first aim is to determine the efficacy of dynamic assessment as intervention strategy to improve on pretest scores as evidenced in South Africa research and secondly to analyse and compare two separate meta-analytic software programmes in terms of their robustness and utility value. The need for such a study is expanded upon below, followed by a brief tour of meta-analysis as technique. The method of data gathering is discussed next and includes a discussion of the reason given for the two-fold purpose of the study. Two software programmes and their subsequent structures are explored, compared and critiqued. The results section tackles the criteria used for inclusion of studies, followed by results from two data runs. A discussion then focuses on the two programmes utilised for the study; a search for potential moderators is included and limitations of the study are viewed. Implications of the results for dynamic assessment research in South Africa are discussed along with recommendations and is then followed by a conclusion which forms the final section. Lastly, an aside to the meta-analysis is included and deals with a potentially rich method of information extraction which is considered useful for assessment in psychology. During 2001-2002 a study was undertaken to detail the then current research situation into dynamic assessment in South Africa (Murphy, 2002; Murphy & Maree, 2006). This study was a narrative exposition on the status on this field of enquiry and did not seek to quantify the results but merely to survey the area.

A meta-analysis was conducted on the studies surveyed during 2001-2002¹ in an attempt to emulate and bring to the field of South African dynamic assessment results similar to those offered by Lussier and Swanson (2002). The aforementioned authors' study yielded more effect sizes than did the present study, due entirely to the paucity of primary studies in this field in South Africa in comparison to the field surveyed by Lussier et al. which included other studies than those conducted in the United States of America (although being limited to English studies). Considering a potential database of 303 articles, Lussier et al. refined their criteria to a point which allowed for only 30 articles to be included in their final analysis. Lussier and Swanson (2002) investigated the degree to which effect sizes, as a function of dynamic assessment as opposed to static assessment, were statistically comparable between ability groups. They also investigated the question of whether the effect size was related to dynamic assessment intervention purely as a methodological artifact or if the effect size was due to the type of research design, intensity of treatment and nature of instructions given. Lussier and Swanson (2002) sourced PSYCINFO for their database of dynamic assessment intervention studies.

Secondly, this South African meta-analysis was deemed timely due to the sufficient number of studies available for such an analysis to be conducted and more importantly the study was warranted based on the unknown summarized significance of the quantitative effect sizes (based on posttest score results). Assessing the efficacy of dynamic assessment interventions within single case studies may not reflect the true cumulative efficacy of such interventions. Meta-analysis is a quantitative statistical review technique, which summarizes the empirical results of any number of studies (Lussier & Swanson, 2002; Wolf, 1987). Isolated studies can never solve any one particular problem and the foundation of scientific progress can be regarded as the accumulation of knowledge gathered from the results of many studies resulting in a quantitative synthesis of research (Hunter, Schmidt & Jackson, 1982; Wolf, 1987). Meta-analysis is an independent specialty within statistics and is specifically suited to the calculation of the various effect sizes emanating from individual studies and determining the significance of the combined effect size (i.e. the cumulated effect size for all samples) (Cooper & Hedges, 1994a). In cumulating each study's effect size the significance of the overall effect size across all studies can be computed. As meta-analysis is a quantitative statistical technique it cannot comment on the qualitative value that dynamic assessment as an intervention strategy offers individuals.

Thirdly, South Africa is unique in terms of the need to assess many prospective tertiary education students, the majority of whom can be considered as previously disadvantaged. Of all learners enrolled in tertiary education institutions in South Africa, 60% are

¹ It was decided to use 2002 as a cut-off point for inclusion of studies into the meta-analysis. Studies included herein range from those conducted in 1961 - 2001. However, it must be noted that only two studies could be found that had since been added to the field of dynamic assessment in South Africa which included the article by Skuy, Gewer, Osrin, Khunou, Fridjon & Rushton (2002) which would have been a welcome addition to this meta-analysis as the empirical study would have added another independent sample finding. These findings were ascertained by searching South African databases, such as the NRF database, SABINET databases including among others, completed masters and doctoral research dissertation and theses, electronic databases, research conducted at Technikons and Universities as well as South African article searches (SAE publications). SABINET is linked to all South African research institutions and only publications that are indexed in individual library collections are included. Thus, this present search did not locate sources which were not indexed. However, any potential null "file drawer" results are factored into the meta-analytic results (as originally highlighted by Rosenthal, 1979).

previously disadvantaged students (Department of Education, 2003).² In other countries these students are almost always considered the minority in terms of number. There is thus an urgent need to fill the assessment gap for potential tertiary education students and the focus in many of the studies surveyed here have as their sample such individuals.

Murphy (2002) as well as Murphy and Maree (2006) reviewed 29 studies dealing with dynamic assessment in South Africa. Six were purely qualitative studies and one was a validation study. These studies were thus not amenable to a quantitative synthesis. This left 22 studies that were considered for possible inclusion in a meta-analysis. Of the 22 studies considered for inclusion only 7 studies complied with the necessary requirements for a meta-analysis to be conducted using the two software packages. This resulted in the exclusion of 15 studies. The format for the data necessitated by the software packages led to the inclusion of only between-groups research designs. All 15 excluded studies contained data from within-groups research designs and could thus not be included. The small number of studies eventually included in the study may militate against conducting such a study yet this leaves one of two options open to the prospective researcher; either wait until more studies avail themselves or conduct a study now and conduct another one at a later date. Nevertheless, not only is the original pool of studies small but the further delimitation of only seven as final amount included warrants due caution for any and all conclusions that are reached in this study. Readers are warned at the outset that the results of this study are to be considered tentative. Further details pertaining to all the studies can be found in Section B of this appendix, including all primary study characteristics and their statistical results as well as more specific results. This has been included for those readers who wish to replicate the analysis using the data in section B, for which purpose, all primary data has already been extracted from the original texts. The question to be answered by this meta-analysis is as follows:

does dynamic assessment intervention make a significant difference as opposed to no (static) intervention across separate studies? In other words, regardless of the significance of the original primary findings, what does the cumulative finding result in?

This study also investigates the usefulness of two meta-analytic software programmes freely available via the internet. The programmes are compared in terms of their ease of use, documentary user-support and final analysis that is outputted as results.

1.1 The need for a meta-analysis of South African dynamic assessment

As at 2001, 29 empirical studies utilising dynamic assessment in South Africa had been conducted yielding results mostly in favour of the efficacy of dynamic assessment interventions when compared to static or no interventions (Murphy, 2002; Murphy & Maree, 2006). In order to determine whether the cumulative effect of dynamic assessment was in keeping with the individual case study results a meta-analysis of these studies was deemed necessary. Secondly, there was no empirical study, which had as yet, investigated the efficacy of dynamic assessment across studies. Thirdly, due to the unique nature and challenges facing South African higher education, where 60% of higher education students are considered previously disadvantaged, the case for the utilisation of dynamic assessment as potential entrance assessment tool becomes an even more urgent one (Department of Education, 2003).

Dynamic assessment as a method of testing is uniquely placed in South Africa as the majority of learners in this country have suffered moderate to severe educational handicaps due to past segregationist policies, the results of which are still prevalent (Skuy, Gewer, Osrin, Khunou, Fridjon, & Rushton, 2002). As such, dynamic assessment is considered a method less biased towards the socially disenfranchised (Elliott, 2000) and hence more suitable as a viable alternative to current psychometric tests (Hessels & Hamers, 1993; Sewell, 1987; Van de Vijver, 1993). Gains in scores between pretest and posttest South African dynamic assessment interventions has evidenced that, in general, dynamic assessment has proved efficacious as a method of helping individuals improve on tasks requiring skills in varying test batteries (Murphy, 2002; Murphy & Maree, 2006).

The current educational crisis within South Africa and the assessment of previously disadvantaged learners and their entrance to tertiary educational institutions is of concern. Dynamic assessment may well prove a viable option as choice of assessment instrument, if as evidenced from South African studies (Murphy, 2002; Murphy & Maree, 2006), dynamic assessment does in fact have a significant and sizeable effect. In order to determine this, a meta-analysis is conducted on the current research in South Africa, the results of which will help to inform future policy governing the assessment of previously disadvantaged individuals and their entrance to institutions of higher education. Dynamic assessment is not only a method of assessment but serves in the capacity of mediatory tool which can result in effects other than those obtained in pretest-posttest studies. The importance of the qualitative relationship between the assessor and the testee is one such aspect (Lidz & Elliott, 2000b) that cannot, for instance, be measured by only studying posttest scores.

² These figures are based on 2001 statistics reported in the Department of Education's 2003 report.

However, the individual studies did not assess for this relationship and likewise neither did the meta-analysis. In order to determine the full efficacy of the mediatory aspect inherent in dynamic assessment, means, other than those utilised by cumulating effect sizes across studies is necessitated. There is thus a limitation to which such a meta-analysis can proceed. Effect size results will not necessarily highlight the total effect of dynamic assessment intervention on posttest scores. They will also not necessarily inform the process as to the overall effect that such an intervention will have on individuals undergoing such intervention strategies. For instance, looking at only posttest scores after sessions of dynamic assessment interventions can in no appreciable way be informative regarding any potential long-term effects of cognitive mediation. More qualitative and long-term research investigations are necessary to determine the fuller impact that dynamic assessment may or may not evidence. The effect size results in this study are thus to be interpreted as evidencing either a cumulative effect or lack of such effect across studies, but the results do not in any manner reflect the value and nature of the full scope that dynamic assessment interventions have to offer. The original intention within each of the individual studies was to determine the significance of dynamic assessment interventions. The meta-analysis merely reviews this endeavour by synthesizing the effect of dynamic assessment interventions. If the original studies were able to conclude that dynamic assessment did or did not have a significant effect, then by extension this too can be applied to the meta-analysis.

Since 1961 a number of studies in South Africa have used dynamic assessment interventions as instruments of mediation in order to have as a result increased scores on pretest-posttest research designs. The results of each study when taken in isolation from other studies yields results evidencing the efficacy of these dynamic assessment interventions. However, in order to empirically investigate whether cumulative efficacy is apparent across all studies, a meta-analysis needs to be conducted. Murphy (2002) as well as Murphy and Maree (2006) reviewed South Africa dynamic assessment research (1961-2001) and used the primary studies highlighted in the research for purposes of this meta-analysis. Making use of vote counting, Murphy (2002) as well as Murphy and Maree (2006) concluded that of 29 primary empirical studies, 21 revealed that dynamic assessment interventions indeed had a significant effect as intervention strategy to improve scores on pre-tests. Two primary empirical studies yielded non-significant effects evidencing lack of support for the notion that dynamic assessment interventions significantly improve post-test scores. Six studies yielded confounding results, evidencing both significant and non-significant results (i.e. in these studies the same sample would be utilised for more than one experiment). The study concluded, that, based on these findings dynamic assessment in South Africa was efficacious in bringing about significant change in pretest scores. As is at times the case, a meta-analysis will either reveal significant effects across cumulated studies or non significant effects (contrary to those effects evidenced within individual studies). In order to determine whether dynamic assessment was efficacious an empirical meta-analysis was conducted to either further support the original conclusions in Murphy (2002) as well as Murphy and Maree (2006) or to caution against possible inferences made from the conclusion.

1.2 A brief tour of meta-analysis

In essence meta-analysis seeks to cumulate findings across primary studies, analyse the combined findings and derive conclusions from the total number of studies. In so doing, it may happen that results counter findings in the primary studies, in other words, what may have seemed to be an effectual experimental intervention in a study may not in fact contribute much proportionally on a larger scale thus nullifying the original results. This is not necessarily the case in all studies but such findings are not outside the norm. It may happen that experimental results are indeed robust in terms of results when compared to control groups for many studies and when cumulated result in even greater yield in terms of effectiveness. Meta-analysis typically finds its niche in studies designed to test the differences between experimental and control groups but is not limited to such designs (Chambers, 2004; Kulik & Kulik, 1989). The d family of effect sizes is used in this study including Hedge's g , Glass' delta and Cohen's d (Rosenthal, 1994) which necessitates both control and experimental groups (Schwarzer, 1989; Strube & Hartmann, 1983). This fact along with the fact that the two software programmes utilised in this study also made use of this family of effects size statistic led to the preference in this meta-analysis to locate studies with between-groups designs only, i.e. studies with comparisons between experimental and control groups.

The name for this technique was first introduced by Gene Glass in 1976 (Chambers, 2004) and as such is quite recent in terms of statistical methodology, however Pearson had already in 1904 taken the average correlation results of medical studies and utilised them in research, with similar techniques being used throughout the early half of the twentieth century (Bangert-Drowns, 1992; Cooper & Hedges, 1994a). The need to allocate an effect size for each study in terms of its overall contribution to the final result was an outgrowth of behavioural scientists' need to summarise large databases of literature in as systematic a fashion as possible (Rosenthal, 1979, 1995). The quantitative generalisation of such a systematic investigation into the results of many primary studies would seem to offer more value; the "strengthening" of methods (Hall, Tickle-Degnen, Rosenthal & Mosteller, 1994) in terms of strategic recommendations based on such findings (Cooper & Lemke, 1991; Arthur, Bennett & Huffcutt, 1994; Hunter, Schmidt & Jackson, 1982; Strube, 1985; Wolf, 1987), notwithstanding the usefulness of narrative reviews of studies (Strube & Hartmann, 1983). The above-mentioned 2001-2002 narrative study could be crudely construed as a vote-counting

method of sorts³ (Bushman, 1994; Kline, 2004; Schwarzer, 1989) and hence arose the need to test and quantify similar hypotheses results across independent studies⁴ (Kalaian & Raudenbush, 1996). This reflects a normal advance in any area of research interest, signifying the entry of the particular area into mainstream research territory and alerting the reader to a new body of research (Myers, 1991). As mentioned, Section B contains further information pertaining to each individual study and can be construed as a coding scheme for this analysis' purposes (Orwin, 1994; Schwarzer, 1989; Stock, 1994).

Synthesising research can never replace the need for reading original sources, and this statement is made apart from the fact that it is merely good practice to do so, but is sated more as a result of the number of discretionary steps taken by meta-analysts when deciding on what to include and what to leave out in their final analysis (Arthur et al., 1994). For instance, although fail-safe computation has partially addressed the problem of excluding null findings in the meta-analysis (a result partly due to publication bias favouring significant findings⁵ for example; Begg, 1994; Hunter, 1997; Nester, 1996; see chapter 4's discussion on the statistical issue surrounding significance testing), the onus rests with the researcher to locate any and all information pertaining to the area of study (Glass et al, 1981; Strube & Hartmann, 1983). It is not the opinion of this author to obviate the need to study previous research results regardless of the findings of the overall result as has been alluded to by David Hilbert albeit in a somewhat different context (Glass, McGraw and Smith, 1981). The researcher has to decide on the criteria for inclusion of studies, the model assumptions to be used, the use or lack thereof of programmes individually tailored to the needs of the particular meta-analytic study, the necessity of inter-rater reliability when coding large numbers of studies and much else besides (Dickersin, 1994; Hunter et al., 1982; Reed & Baxter, 1994; Rosenthal, 1994; White, 1994; Wortman, 1994).

Meta-analysis approaches are often found to emanate from one of two major groupings, namely the combination of significance levels and the combination of effect sizes; the latter being used in this particular study (Strube & Hartmann, 1983). Combining statistical significance levels indicates the degree to which chance plays a role in the findings whereas the combination of effect sizes examines the magnitude of effect across studies (Becker, 1994; Shadish & Haddock, 1994; Wolf, 1987). Two "families" of effect sizes are available to the meta-analyst and include the r family (which also includes Zr , Fisher's transformation or r); and the d family which includes Hedge's g , Glass' delta and Cohen's d (Rosenthal, 1994). The latter family of statistics are used within this analysis.

2. Method

2.1 Two-fold purpose of the meta-analysis

The overriding reason for running the meta-analysis is to impart to the field of dynamic assessment pertinent information regarding the scope of dynamic assessment in South African research. Moreover, it was thought prudent to run the analysis on two separate computer software programmes in order to highlight advantages and disadvantages of these programmes so as to offer the reader a choice of application should further analysis be undertaken. Statistical packages such as SAS and SPSS run standard statistical techniques used in the behavioural sciences and are thus fairly widespread, as such, details of statistical runs are not often discussed in research reports.

However, the same cannot be said of meta-analytic packages which do not run "as a programme" or subroutine within SPSS nor SAS, although macros and more programmes are becoming available.⁶ Initially most meta-analytic software was available only for mainframes and not microcomputers, however this has now changed (Arthur et al., 1994). Standardised packages are however not yet the norm as each package assumes various models, theoretical and conceptual underpinnings (Arthur et al., 1994). Is it for this reason that these packages are compared and detailed in their functioning. This section thus serves two purposes: to assess the primary study results cumulatively across studies and to evaluate the usefulness of two software

³ As Bushman rightly adds, "when effect sizes are medium to small [as is the case with this study], the conventional vote-counting procedure frequently fails to detect the differences" (1994, p.195).

⁴ Combining independent studies sharing the same or at the very least similar hypotheses are the necessary requirements when conducting a meta-analysis. Combining studies at will without due consideration of various hypotheses is blatant nonsense. Cf. Eysenck (1995) for an attack launched at unthinking use of meta-analysis in just such a scenario.

⁵ Hopefully something of the past and if the social sciences are to follow suit, something similar to an initiative to publish negative findings from clinical trials will result, especially in the psychological discipline (Editorial, *New Scientist*, 2005).

⁶ For the interested reader the following are a few programmes currently available. Commercial programmes include: '**Comprehensive meta-analysis**' by Borenstein, M & Rothstein, H. available at www.MetaAnalysis.com; '**DSTAT**' by Johnson, B.T. available at www.erlbaum.com; '**ES**' an effect size computational programme by Shadish, W.R., Robinson, L & Wu, C.; '**FAST*PRO**' by Eddy, D.M. & Hasselblad, V.; '**MetaWin**' by Rosenberg, M.S., Adams, D.C. & Gurevith, J. available at www.metawinsoft.com; '**Metaxis**' available at www.update-software.com/metaxis/metaxis-frame.html. As availability of demonstration programmes changes from month to month on the internet it is possible that some of these programmes might be available at no cost for a specified length of time. There are a host of programmes freely available as freeware as well as shareware. The above-mentioned information as well as information pertaining to the freeware as well as SAS and SPSS macros are available at the following comprehensive website as at April 2005: www.um.es/facpsi/metaanalysis/software.php. Some of these programmes are reviewed in the professional literature (cf. Arthur, et al., 1994; Normand, 1995). Of interest is the number of meta-analytical models "on the market" each with its own advantages and disadvantages (Bangert-Drowns, 1992).



programmes. The latter will be addressed in 2.2 and 2.3 below. Comparison of meta-analytic programmes has in the past offered researchers the opportunity to make informed decisions when deciding on the use of one programme over another (Arthur et al., 1994; Normand, 1995). The choice of these two programmes were made due to their availability and cost (they are both freeware products) and their accompanying recommendations made by peers within the field. Both programmes were fairly small to download (91 and 212 kilobytes for both the Kenny and Schwarzer programmes respectively); run hassle-free within the windows environment and have fairly good to good manuals which accompany the software. The author of the first programme (Kenny) is also available for questions about his programme. No information was sought from Schwarzer.

2.2. "META – Easy to answer" version III by D. A. Kenny

This software programme was developed by David A. Kenny at the University of Connecticut, United States of America and is a compiled version of a QuickBasic programme with a DOS-like appearance which runs in the windows environment.⁷ This is a shareware version offered free of charge and can be downloaded over the internet at the following address: <http://davidakenny.net/meta.htm>. Kenny (2003) cautions the user however as to its as yet demonstration status and the user is advised to check computational output. This cautionary note further propelled the need for a second programme's analysis and a double-check of sorts was conducted to compare output. This programme computes effect sizes for each study, pools the results and calculates the degree to which the result differs from zero and also tests for homogeneity of effect sizes across studies. It allows for the weighting of studies based on sample size, variance or user-inputted values.

2.2.1 Programme structure

The programme encompasses three stages; the first stage seeks overall study information, and prompts for the following user information:

Effect size option
 Cohen's d: D
 Correlation: R
 Difference between proportions: P
 None (only combine p's): N

If **Cohen's D** is chosen as an option, the programme seeks the following information:

Do you want the Hedge's correction to d? Yes or No?
 Do you want to assume equal group sizes? Yes or No?
 Do you have an input data file? Yes or No?
 Do you want to weight studies? Yes or No?
 If Yes, weight by:
 Degrees of freedom: D
 Sample size: N
 Study variance (inverse): V
 Other: O
 Take the square root for weight? Yes or No?
 Does sample size equal degrees of freedom plus two? Yes or No?
 Completed stage 1

If **correlation** is chosen as an option, the programme seeks the following information:

Do you want to use the Fisher's Z transformation? Yes or No?
 Do you have an input data file? Yes or No?
 Do you want to weight studies? Yes or No?
 If Yes, weight by:
 Degrees of freedom: D
 Sample size: N
 Study variance (inverse): V
 Other: O
 Take the square root for weight? Yes or No?
 Does sample size equal degrees of freedom plus two? Yes or No?
 Complete stage 1

If **difference between proportions** is chosen as an option the programme seeks the following information:

Transformation options:
 Difference in proportions: P
 Probit: N
 Logit: L
 Arcsin: A

⁷ A larger windows environment programme is also available but caution was attached to the use of this programme (Kenny, personal communication, 23 March, 2005).



Do you want to assume equal group sizes? Yes or No?
 Do you have an input data file? Yes or No?
 Do you want to weight studies? Yes or No?
 If Yes, weight by:
 Degrees of freedom: D
 Sample size: N
 Study variance (inverse): V
 Other: O
 Take the square root for weight? Yes or No?
 Does sample size equal degrees of freedom plus two? Yes or No?
 Complete stage 1

If **none (only combine p's)** is chosen as an option the programme seeks the following information:

Do you want to assume equal group sizes? Yes or No?
 Do you have an input data file? Yes or No?
 Do you want to weight studies? Yes or No?
 If Yes, weight by:
 Degrees of freedom: D
 Sample size: N
 Study variance (inverse): V
 Other: O
 Take the square root for weight? Yes or No?
 Does sample size equal degrees of freedom plus two? Yes or No?
 Complete stage 1

Only Cohen's D was made use of for this meta-analysis. No further information will be provided for the programme regarding the remaining options. The second stage in the programme prompts the user for the following information:

Type of test statistic:

T - t test	M (means and variances)
F - f test	P (p values)
R - correlations	S (proportions)
X - Chi square	Z (z test)
E - (no more studies)	

If the **t test** results are used the programme seeks the following information:

Enter the t statistic for study X
 Group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1
 Enter the N (at least 1) for group 2

If no, enter the N for study X (*Kenny's programme will then divide this N by two, hence not allowing the user to make use of within-groups studies*)

Is the result in the expected direction? Yes or No?

OUTPUT screen includes the following:

N (number)	df (degrees of freedom)
d (effect size)	r (effect size r)
z (z statistic)	t (t statistic for variance)
Variance	BESD (binomial effect size distribution)
Effect size	
Weight	

Do you want to include these results in the meta-analysis? Yes or No? (*The user has to write out per hand the results of each study, should they find this necessary, as this demonstration edition offers no saving or printing options*)

If the **f test** results are used the programme seeks the following information:

Enter the F statistic for study X
 Group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1
 Enter the N (at least 1) for group 2

If no, enter the N for study X

Is the result in the expected direction? Yes or No?

OUTPUT screen includes the same information as given in the "t test" option

If the **correlation** results are used the programme seeks the following information:

Enter the correlations <-1> and <1> for study X
 Enter the N for study X
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option

If the **Chi square** results are used the programme seeks the following information:

Enter the chi square for study X
 Group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1



- Enter the N (at least 1) for group 2
 If no, enter the N for study X
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option
- If the **means and variances** results are used the programme seeks the following information:
 Enter the mean for group 1
 Enter the mean for group 2
 Group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1
 Enter the N (at least 1) for group 2
 If no, enter the N for study X
 Are you entering standard deviation (s) or variances (v)?
 If (s), are they pooled (p) or unpooled (u)
 If (p), enter the pooled variability measure
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option
 If (u), enter the variability measure for group 1 (*Kenny refers to this standard deviation as 'variability' which may be confusing if one is not really working with variance*)
 Enter the variability measure for group 2
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option
 If (v), the same procedure as above is followed but variance and not standard deviation is used
- If the **p value** results are used the programme seeks the following information:
 Enter the two-tailed p value for study X
 Is the p value from t or F (T) or chi square (Z)
 If (T), Group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1
 Enter the N (at least 1) for group 2
 If no, enter the N for study X
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option
 If (Z), Group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1
 Enter the N (at least 1) for group 2
 If no, enter the N for study X
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option
- If the **proportion** results are used the programme seeks the following information:
 Enter the proportion for the first group for study X
 Enter the proportion for the second group for study X
 Is there a chi square? Yes or No?
 If yes, enter the chi square for study X
 Group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1
 Enter the N (at least 1) for group 2
 If no, enter the N for study X
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option
 If no, group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1
 Enter the N (at least 1) for group 2
 If no, enter the N for study X
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option
- If the **Z statistic** results are used the programme seeks the following information:
 Enter the Z statistic for study X
 Group sizes equal? Yes or No?
 If yes, enter the N (at least 1) for group 1
 Enter the N (at least 1) for group 2
 If no, enter the N for study X
 Is the result in the expected direction? Yes or No?
 OUTPUT screen includes the same information as given in the "t test" option
- The third stage integrates the input from stages 1 and 2. The results of the third stage are in fact the meta-analytic results of the combined studies and includes the following output:
 Study number
 Subject n
 Average effect size



Effect size standard deviation
 T test of effect size
 Average d
 Average r
 BESD
 Homogeneity of effect sizes
 Chi square
 Average Z
 Fail-safe N (*The user has to write out per hand the results of the final analysis, as this demonstration edition offers no saving or printing options*)

The data that is entered into META is placed into a data file which is then used by META to compute the final analysis. This data file consists of numeric characters but occasionally alpha-numerics are used. The data file is easy and flexible to manage and changes can be made directly in the data file, although this is not recommended unless the user understands the derivation of all the computations. However, studies can be deleted and weightings changed as desired by going through the data within the META programme itself as opposed to changing the data within the data file itself. Table 9 shows the second data file run in META. Both individual study results as well as the final output has to be written out by hand as there is no print and save option available.

2.3 Meta-analysis programme version 5.3 by R. Schwarzer

This programme was developed by R. Schwarzer at the Freie Universität Berlin, Germany and was written in Turbo Pascal 5.0. The program is not public domain but is distributed under the User Supported Software concept (Schwarzer, 1989) and can be downloaded from <http://www.RalfSchwarzer.De>. Schwarzer's programme allows for the computation of probabilities, effect sizes d , and effect sizes r (correlations). Depending on the data available to the researcher, any of these three can be selected. Also available is a data editor and a number of utilities which provide transformed data results. The main menu consists of the following:

Appendix 1 Table 1 Schwarzer programme menu options (Schwarzer, 1989)

GENERAL	p VALUES	d VALUES	r VALUES	UTILITIES
Editor	Meta-analysis	Meta-analysis	Meta-analysis	Conversion r
Directory	r-File	Cluster		
Effect Size d				
Change Dir	r-File	StemLeaf	Signif. Of corr	
Info		Weighted M , V , C		
Calculator		t-Tests		
Quit				

2.3.1 Programme structure

The general menu allows the user to access the data file (editor) which has to be entered according to a specific format (depending on which data are entered, i.e. correlations, proportions or d values). The programme reads the data file and computes the final result which can be saved and/or printed. The individual results, however, have to be hand written. For this study, the "utilities" menu was used to compute the individual study effect sizes based on group mean and standard deviation. The data file can then be assembled and is shown in the tables that follow.

Depending on which test statistic is available for computation into an effect size and also depending on how the final analysis will be run, Schwarzer's programme requires that the data be in certain formats. Effect size computation using d facilitates up to 10 groups. For instance, when computing effect sizes using d values, the data file needs to include the study number, sample size for group 1, sample size for group 2, effect size and a reliability coefficient. As reliability coefficients were not available in the primary studies, unity was maintained throughout the studies by inputting 1.0 as suggested by Schwarzer (1989). For probabilities, study number, sample size and p -values are needed; for effect sizes r , the study number, sample size as well as correlations and the variables' reliabilities are needed. In essence, after having computed the effect sizes for each study using the utilities menu and compiling a data file, this programme merely runs the file according to the chosen statistic. Schwarzer (1989) maintains that meta-analysis of effect sizes are superior to those using only combinations of probabilities. Table 2 illustrates the utilities menu, "conversion to r ", in which Schwarzer offers the following options:

Appendix 1 Table 2 Utilities menu in Schwarzer's programme

Select a coefficient to be transformed	
Point biserial correlation coefficient	p)oint
t-value for 2 independent samples	t)value
F-value for 2 or more independent samples	F)value
Chi square value for contingency tables	x)square
*Four cells frequencies	c)ells
*U-value (Mann-Whitney)	U)value
Exact one-tailed probability p	e)xact p
Effect size g (standardized mean difference)	g)value
r to Fishers z transformation:	r) to z
Back transformation	z) to r
Normal distribution Z to probability p	N)ormal
Quit Transformation Program	Q)uit

The "effect sizes d' " option seeks the number of groups to be compared and whether standard deviation or variance is available. The "significance of correlation" option seeks correlation values for the chosen number of groups. Means, variances and correlations can also be input and weighted. Lastly t values can be computed for the chosen number of groups. Schwarzer is cognisant of the broad variety of available primary statistics to the meta analytic researcher and as such, this programme offers a variety of statistical manipulations for the chosen transformation statistic and is flexible in terms of the data available from the primary research.

2.4 Comparison of the two programmes

Both programmes output similar results in terms of individual studies, although Schwarzer's programme is more comprehensive and offers more variation in types of output. Kenny's programme offers less variation in output results. During the input stage, Kenny's programme requires more information per study, integrates the data file with the output and allows for a more comprehensive data file. Schwarzer's programme is not integrated in similar fashion and does not require as much detail during input. However, Schwarzer's programme allows for very comprehensive transformation utilities which can then be used within the data editor for the meta-analysis of choice (either analysis based on d' values, r values as well as cluster analysis output and stem-and-leaf displays).⁸

Unlike Kenny's programme where the data input and calculation take place in seemingly one step, Schwarzer's programme requires the user to first compute effect sizes using the "utilities" option. These results are then recorded by hand and typed into the data file. The final analysis simply runs the data file. Thus, two separate steps are necessitated. For input, Kenny makes use of an effect size which is the equivalent of the Schwarzer's " g " which is the effect size based on pooled variance. Kenny refers to Schwarzer's " g " as " d' ", which can lead to some confusion, this being acknowledged by Schwarzer (1989). Kenny makes use

* Schwarzer's programme makes available non-parametric transformation formulae that Kenny's does not.

* Although derivation of effect size computation (such as delta) can be computed from available t statistics, conversions from non-parametric statistics is more difficult than originally thought. As Glass et al. state, "no simple transformation of U into delta is possible since the U test...[does] not test simple hypotheses about population means" (1981, p.130). Nevertheless this option is available in Schwarzer's programme allowing meta-analysts to conduct runs on older non-parametric data.

⁸ Stem-and-leaf displays are not used for this data as there are too few effect sizes for a reasonable display; moreover, these displays are better suited to correlation values.

of Hedge's unbiased estimator d in his meta computation. There is a slight difference in the numerical value of the statistic that both programmes use.

Individual study results are however exactly the same. Kenny's programme in essence works on a random model principle (Kenny, personal communication, 11 April 2005) and Schwarzer's programme presents both fixed and random-effects model results. Fixed-effects models assume that any differences between samples are strictly due to sampling error with an average effect size simply being an unbiased estimate or simple average of a population effect whereas random-effects models assume that differences may also be attributed to aspects other than sampling error with the assumption that the sample has been drawn from a population. Therefore there is not only one population effect size but a distribution of populations effects size resulting in sample characteristics that are not only dissimilar but which also reflect true underlying population differences (Cooper & Hedges, 1994b; Cooper, 1994; Normand, 1995; Schwarzer, 1989; Shadish & Haddock, 1994). Residual variation indicated among others include results from the chi square analysis and tests of homogeneity (Chambers, 2004) the significance of which will prompt further investigation into random model usage. Kenny's programme, based on a random-effects model, assumes that the study is used as the sampling unit (Kenny, personal communication, 11 April 2005; Kenny, 2003) unlike fixed-effects models which use as their sampling unit individuals within the studies (Rosenthal, 1995).

The manuals that accompany both programmes offer the necessary and requisite information in order for correct data input; knowledge of how the programme functions within the operating environment; both allow printing options on only some menus and are consistent across manuals in terms of current research into the statistical area of meta-analysis. Both manuals enable the user to perform the necessary computation in order to obtain output. Schwarzer's manual is however more comprehensive, serving as an introduction to and brief overview of the field of meta-analysis. It also elucidates the statistical formulae used within the computations themselves which Kenny's manual does not offer. Schwarzer's programme also offers more variety in terms of output, such as cluster analysis for both d and r values and visual display of effect sizes which Kenny's does not. Cluster analysis allows the user to search for potential moderating factors which present themselves in terms of how the effect sizes are clustered. Kenny's manual states that the researcher look for moderator variables but does not allow a similar option. The manuals are available for downloading at the same above-mentioned web addresses that are accessed to download the programmes. It is advisable that the researcher study both manuals before attempting to use either programme.

2.5 Limitations of the programmes and violation of assumptions

Neither Kenny's nor Schwarzer's programme can handle within-groups studies nor repeated-measures designs and it is for this reason that only between-groups studies were included in the analyses and this criterion results in the further delimitation of the number of studies eventually included in the study and can be considered as a type of selection bias.⁹ The fact that some meta-analytic studies are not based on repeated-measures designs (Normand, 1995) is noted, however, multiple end-point studies do complicate the methodology involved in synthesising such data. It is reasoned that more commercially available software would better cater for such studies and is perhaps something to think about in terms of re-conducting this study so as to include those studies left out of this one. For the seven primary studies used for this analysis, 22 effect sizes were generated. This was possible due to the input of more than one dependent samples per study. However, two primary studies generated three and four independent results respectively thus averaging 10 independent effect sizes. An option to average out the effect sizes per study was considered but rejected as too few effect sizes would have made this endeavour superfluous. This study has thus violated an assumption inherent in both programmes, that of independent samples. When interpreting the results it is prudent to keep in mind this violation.

3. Results

3.1 Criteria used for inclusion of studies

Of the original 22 studies considered for inclusion, Murphy (2002) as well as Murphy and Maree (2006), surveyed 29 studies. However 7 studies were purely qualitative studies and thus not amenable to a meta-analytic study, 7 made use of a within-groups design and 14 made use of between-groups design. The two programmes chosen to conduct the meta-analysis do not allow for the analysis of within-group studies, as the main aim of the analysis is to determine the effect of dynamic assessment on an experimental group versus the effect of no dynamic assessment on the control group. The between-groups assumption inherent in both these programmes therefore cannot run analyses on repeated measures designs on one-group scenarios. This cannot be said to be a major flaw in the programmes because most meta-analyses are utilised for the express purpose of determining effects on experimental groups as opposed to the lack thereof on control groups. It is perhaps telling then, that local research designs veer more towards within-group analyses and not between-group analyses.

⁹ It is a pity that their programmes are unable to contend with such study characteristics as the Lussier and Swanson study (2002) was able to compute for repeated-measures design.

Of the original 22 studies, eleven used means and standard deviations; five used correlations; three used *t* tests and the remaining two used multiple regression as well as discriminant analyses. Of the final seven chosen for inclusion (i.e. the between-groups studies) six used means and standard deviations and one used a combination of *t* test results as well as means and standard deviations. For the sake of complete comparison between the two programmes used, only means and standard deviations were utilised (including the means and standard deviations for one study which also included *t* test results). Two analyses were conducted as the first analysis yielded effect sizes that were very large, too large in fact when compared to normal effect sizes as emanating from meta-analysis research literature. Due to the unusually large effect size results obtained from both programmes (1.26 and 1.72 respectively), the data was studied and three outliers were identified. These outliers emanated from two primary studies, namely Lloyd and Pidgeon (1961) (study numbers 20 and 21 samples 1 and 2) and Gewer (1988) (study 19 sample number 4). Effect sizes greater than two to three standard deviations of the mean may be construed as outliers (Chambers, 2004).

Individual effect sizes for these studies ranged from 2.2 - 13.8. Analysing the primary research yielded the reasons for these effect sizes: the Lloyd et al. study presented with very small standard deviations for both the control and experimental groups (0.83, 0.85, 1.08) which was exacerbated by the already large differences between the means of both groups. As the calculation of effect size is very dependent on standard deviation and mean, it stands to reason that this would be the case. Lloyd et al. state the following: "it is not thought that the low variance can be attributed to unrepresentativeness but rather to the greater homogeneity of the Natal children when compared to the English children" (p.150). It is for this reason that these two samples were eliminated from the data during the second analyses (yielding effect sizes of 6.416 and 13.806 respectively). The sample of Gewer presents with a large difference between both groups (yielding an effect size of 2.211). Thus this particular sample was also eliminated during the second analyses. Table 3 shows study numbers referring to the following primary study authors and the results of the initial data run which yielded 25 effect sizes. Asterisked studies are those with outliers identified as above and were eliminated during the second analysis, thus lowering the total number of effect sizes down to 22.¹⁰ It is this data (22 effect sizes) with which this study concerns itself. A brief descriptive outline of the seven studies included for the analysis is presented in Table 4.

Appendix 1 Table 3 Study numbers and sample classification

Study 1	Andrews (1996)	sample 1	dependent sample
Study 2	Andrews (1996)	sample 2	dependent sample
Study 3	Andrews (1996)	sample 3	dependent sample
Study 4	Andrews (1996)	sample 4	dependent sample
Study 5	Boeyens (1989)	sample 1	dependent sample
Study 6	Boeyens (1989)	sample 2	dependent sample
Study 7	Murray (1988)	sample 1	dependent sample
Study 8	Murray (1988)	sample 2	dependent sample
Study 9	Murray (1988)	sample 3	dependent sample
Study 10	Murray (1988)	sample 4	dependent sample
Study 11	Murray (1988)	sample 5	dependent sample
Study 12	De Villiers (1999)	sample 1	independent sample
Study 13	De Villiers (1999)	sample 2	independent sample
Study 14	De Villiers (1999)	sample 3	independent sample
Study 15	De Villiers (1999)	sample 4	independent sample
Study 16	Gewer (1988)	sample 1	dependent sample
Study 17	Gewer (1988)	sample 2	dependent sample
Study 18	Gewer (1988)	sample 3	dependent sample
Study 19*	Gewer (1988)	sample 4	dependent sample
Study 20*	Lloyd and Pidgeon (1961)	sample 1	independent sample
Study 21*	Lloyd and Pidgeon (1961)	sample 2	independent sample
Study 22	Lloyd and Pidgeon (1961)	sample 3	independent sample
Study 23	Hoffenberg (1988)	sample 1	dependent sample
Study 24	Hoffenberg (1988)	sample 2	dependent sample
Study 25	Hoffenberg (1988)	sample 3	dependent sample

¹⁰ For the following studies; 12, 13, 16 and 17 the authors divided both experimental and control groups into two groups each, thus yielding four sub-groupings. In order to benefit from an increased sample size for the purposes of the meta-analysis the four sub-groupings were "collapsed" into two groups. In order to do so the following formulae were used to calculate means and standard deviations respectively: Average mean: $M = [(M_{Exp} \times n_1) + (M_{Con} \times n_2)] / (n_1 + n_2)$; Average standard deviation: $SD = [(n_1 - 1)SD_1 + (n_2 - 1)SD_2] / (n_1 + n_2 - 2)$.

Appendix 1 Table 4 Brief description of the seven studies included in the meta-analysis

Study	Brief description
Andrews (1996)	To determine if cognition is modified in a group administration of the LPAD; and to detect differences in the degree of modifiability
Boeyens (1989)	To evaluate the performance of a learning potential instrument and to investigate the relationship between academic performance and learning potential
De Villiers (1999)	To investigate the practical application of Vygotsky's construct of the zone of proximal development to the selection of disadvantaged students in higher education and to determine alternative predictors of academic performance other than the traditional matriculation examination results used in South Africa
Gewer (1988)	The study investigated the application of dynamic assessment to a sample of black children within a South African township clinic setting
Hoffenberg (1988)	The study aimed to assess the effectiveness of dynamic assessment among a group of academically superior individuals from a disadvantaged black community in South Africa.
Lloyd and Pidgeon (1961)	To compare the performance of children from different cultural groups on non-verbal tests, half the children were coached and the other half were not
Murray (1998)	To test the effectiveness of a dynamic assessment approach (LPAD) among groups of socio-politically and educationally disadvantaged Indian and Coloured adolescents

3.2 First run with outliers included (25 effect sizes)

3.2.1 Results using META by Kenny

Table 5 shows the results for the meta-analysis

Appendix 1 Table 5 Meta analysis results using Kenny's programme

Study Number	25
Subject n	2619
Average effect size	1.7288*
Effect size standard deviation	5.09
T test of effect size	1.6982 p 0.1024 df 24
Average d	1.7384
Average r	0.2816
BESD	0.3592 - 0.6408
Homogeneity of effect sizes Chi Square	1032.7149* df 24
Average Z	24.4208*
Fail-safe N	3857

Note. * Figures are accurate till four decimal places.

* $p < .0001$

The data file is a normal ASCII file which was saved in notepad. For the sake of clarity the following key is provided for interpretation of the data file:

Appendix 1 Table 6 META data file using Kenny's programme

Study number	Degrees of freedom	Total N per study	Effect size	d	r	z	t	p2 (confidence interval)	p1 (confidence interval)	n1 (experimental group 1)	n2 (control group 2)	weight
1	19	21	1.1221	1.1689	0.5230	2.4319	2.6752	0.7615	0.2384	10	11	4.5825
2	19	21	0.6204	0.6463	0.3213	1.4201	1.4792	0.6606	0.3393	10	11	4.5825
3	19	21	0.7252	0.7554	0.3687	1.6441	1.7289	0.6843	0.3156	10	11	4.5825
4	19	21	0.1232	0.1283	6.7253	0.2895	0.2938	0.53362	0.4663	10	11	4.5825
5	181	183	-8.1786	-8.2127	-4.1288	-5.5472	-5.5549	0.49793	0.5020	91	92	13.5277
6	181	183	0.8667	0.8704	0.4008	5.6236	5.8874	0.70044	0.2995	91	92	13.5277
7	106	108	0	0	0	0	0	0.5	0.5	54	54	10.3923
8	106	108	0.2482	0.25	0.1251	1.2908	1.2990	0.5625	0.4374	54	54	10.3923
9	106	108	0.9929	1	0.4505	4.8927	5.1961	0.7252	0.2747	54	54	10.3923
10	106	108	8.2456	8.3045	4.1875	0.4303	0.4315	0.5209	0.4790	54	54	10.3923
11	106	108	0.1985	0.2	0.1004	1.0341	1.0392	0.5502	0.4497	54	54	10.3923
12	60	62	0.3057	0.3096	0.1541	1.1965	1.2088	0.5770	0.4229	35	27	7.87400
13	61	63	0.5386	0.5453	0.2670	2.1154	2.1640	0.6335	0.3664	32	31	7.93725
14	100	102	0.9937	1.0012	0.3892	4.0434	4.2260	0.6946	0.3053	79	23	10.0995
15	67	69	0.4101	0.4148	0.2059	1.6977	1.7227	0.6029	0.3970	34	35	8.30662
16	70	72	0.5013	0.5068	0.2354	1.9912	2.0272	0.6177	0.3822	48	24	8.48528
17	70	72	0.7047	0.7123	0.3223	2.7617	2.8495	0.6611	0.3388	48	24	8.48528
18	34	36	1.0931	1.1180	0.4767	2.9388	3.1622	0.7383	0.2616	24	12	6
19	34	36	2.1627	2.2119	0.7315	5.0657	6.2562	0.8657	0.1342	24	12	6
20	274	276	6.3990	6.4166	0.9549	25.7780	53.2658	0.9774	2.2524	143	133	16.6132
21	273	275	13.7687	13.8067	0.9897	32.5656	114.4726	0.9948	5.1283	136	139	16.5831
22	264	266	-0.5047	-0.5061	-0.2462	-4.0592	-4.1274	0.3768	0.6231	133	133	16.3095
23	98	100	-0.1290	-0.1300	-6.5503	-0.6474	-0.6498	0.4672	0.5327	52	48	10
24	98	100	-8.8529	-8.9214	-4.4978	-0.4443	-0.4457	0.4775	0.5224	52	48	10
25	98	100	0.1547	0.1559	7.8443	0.7757	0.77894	0.5392	0.4607	52	48	10

3.2.2 Results using meta-analysis by Schwarzer

Table 7 shows the results for the meta-analysis

Appendix 1 Table 7 Meta-analysis results using Schwarzer's programme

Statistic	Result	Statistic	Result
Number of effect sizes	25	Total sample size	2619
Unweighted mean of effect sizes g	1.2614	SE	0.5845
Observed variance of effect sizes g	8.5416	SD	2.9226
Unweighted mean of adjusted effect sizes d	1.2500	SE	0.5829
Observed variance of adj. effect sizes d	8.4967	SD	2.9149
"Weighted Integration Method"			
Mean effect size d+	0.4675	SE	0.0400
Significance Z	10.4424*		
Variance	0.0016	SD	0.0400
95% Confidence interval	0.3798 to 0.5553	Homogeneity Q	1032.6967 df = 24
"Random Effects Model"			
Mean effect size DELTA	1.2313	SE	0.5829
95% Confidence interval	0.0887 to 2.3739**	Significance Z	2.1122
Observed variance	8.49671	Error variance	0.0957
Population variance	8.4009	Homogeneity Q	1013.7210 df = 24
Amount of variance explained by sampling error	1.13		
Kraemer (1983) method			
Mean effect size d	1.2621	95% Confidence interval	1.1712 to 1.355
Population effect size Rho	0.5336	Variance of rho	0.0003
95% Confidence interval	0.5053 to 0.5609	Homogeneity Chi-square	2138.8561
Orwin's Fail-safe n based on "random effects model" DELTA			
Fail-safe for critical d of .20	128.9151	Fail-safe for critical d of .50	36.5660
Fail-safe for critical d of .80	13.4787		

* $p < .0001$. ** $p < .005$.

The data file is a normal ASCII file which was saved in notepad. Table 8 delineates the data file from Schwarzer's programme.

Appendix 1 Table 8 Data file from Schwarzer's programme

Study number	Sample size group 1 (experimental group)	Sample size group 2 (control group)	Effect size	Reliability coefficient
1	10	11	1.1689	1
2	10	11	0.6205	1
3	10	11	0.7554	1
4	10	11	0.1232	1
5	91	92	-0.0082	1
6	91	92	0.87041	1
7	54	54	0	1
8	54	54	0.25	1
9	54	54	1	1
10	54	54	0.083	1
11	54	54	0.2	1
12	35	27	0.3096	1
13	32	31	0.5454	1
14	79	23	1.0013	1
15	34	35	0.4148	1
16	48	24	0.506	1
17	48	24	0.7124	1
18	24	12	1.118	1
19	24	12	2.2119	1
20	143	133	6.4167	1
21	136	139	13.8067	1
22	133	133	-0.5061	1
23	52	48	-0.1301	1
24	52	48	-0.0892	1
25	52	48	0.1559	1

3.3 Second run with outliers excluded (22 effect sizes)

3.3.1 Results using META by Kenny

Table 9 shows the results for the meta-analysis using the programme by Kenny and Table 10 includes the data file created by the programme.

Appendix 1 Table 9 Meta-analysis results using Kenny's programme

Study Number	22
Subject n	2032
Average effect size	0.3354
Effect size standard deviation	0.4787
T test of effect size	3.2863 * <i>df</i> 21
Average d	0.3408
Average r	0.1540
BESD	0.4230 - 0.5770
Homogeneity of effect sizes Chi Square	103.3336 ** <i>df</i> 21
Average Z	5.5697 **
Fail-safe N	156

* $p < .005$. ** $p < .0001$

Appendix 1 Table 10 META Data file using Kenny's programme

Study Number	Degrees of freedom	Total N per study	Effect size	d	r	z	t	p2 (confidence interval)	p1 (confidence interval)	n1 (experimental group 1)	n2 (control group 2)	weight
1	19	21	1.1221	1.1689	0.5230	2.4319	2.6752	0.7615	0.2384	10	11	4.5825
2	19	21	0.6204	0.6463	0.3213	1.4201	1.4792	0.6606	0.3393	10	11	4.5825
3	19	21	0.7252	0.7554	0.3687	1.6441	1.7289	0.6843	0.3156	10	11	4.5825
4	19	21	0.1232	0.1283	6.7253	0.2895	0.2938	0.5336	0.4663	10	11	4.5825
5	181	183	-8.1786	-8.2127	-4.1288	-5.5472	-5.5549	0.4979	0.5020	91	92	13.527
6	181	183	0.8667	0.8704	0.4008	5.6237	5.8872	0.7004	0.2995	91	92	13.527
7	106	108	0	0	0	0	0	0.5	0.5	54	54	10.392
8	106	108	0.2482	0.25	0.1251	1.2908	1.2990	0.5625	0.4374	54	54	10.392
9	106	108	0.9929	1	0.4505	4.8927	5.1961	0.7252	0.2747	54	54	10.392
10	106	108	8.2456	8.3045	4.1875	0.4303	0.4315	0.5209	0.4790	54	54	10.392
11	106	108	0.1985	0.2	0.1004	1.0341	1.0392	0.5502	0.4497	54	54	10.392
12	60	62	0.3057	0.3096	0.1541	1.1965	1.2088	0.5770	0.4229	35	27	7.8740
13	61	63	0.5386	0.5453	0.2670	2.1154	2.1640	0.6335	0.3664	32	31	7.9372
14	100	102	0.9937	1.0012	0.3892	4.0434	4.2260	0.6946	0.3053	79	23	10.099
15	67	69	0.4101	0.4148	0.2059	1.6977	1.7227	0.6029	0.3970	34	35	8.3066
16	70	72	0.5013	0.5068	0.2354	1.9912	2.0272	0.6177	0.3822	48	24	8.4852
17	70	72	0.7047	0.7123	0.3223	2.7617	2.8495	0.6611	0.3388	48	24	8.4852
18	34	36	1.0931	1.1180	0.4767	2.9388	3.1622	0.7383	0.2616	24	12	6
19	264	266	-0.5047	-0.5061	-0.2462	-4.0592	-4.1274	0.3768	0.6231	133	133	16.309
20	98	100	-0.1290	-0.1300	-6.5503	-0.6474	-0.6498	0.4672	0.5327	52	48	10
21	98	100	-8.8529	-8.9214	-4.4978	-0.4443	-0.4457	0.4775	0.5224	52	48	10
22	98	100	0.1547	0.1559	7.8443	0.7757	0.7789	0.5392	0.4607	52	48	10

3.3.2 Results using meta-analysis by Schwarzer

Table 11 shows the results for the meta-analysis using the programme by Schwarzer and Table 12 includes the data file created by the programme.

Appendix 1 Table 11 Meta-analysis results using Schwarzer's programme

Statistic	Result	Statistic	Result
Number of effect sizes	22	Total sample size	2032
Unweighted mean of effect sizes g	0.41369	SE	0.09676
Observed variance of effect sizes g	0.20600	SD	0.45387
Unweighted mean of adjusted effect sizes d	0.40551	SE	0.09496
Observed variance of adj. effect sizes d	0.19839	SD	0.44541
"Weighted Integration Method"			
Mean effect size d+	0.23703	SE	0.04511
Significance Z	5.1901*		
Variance	0.00203	SD	0.04511
95% Confidence interval	0.1475 to 0.3265	Homogeneity Q	1032.2411 df= 21*
"Random Effects Model"			
Mean effect size DELTA	0.3481	SE	0.0910
95% Confidence interval	0.1697 to 0.5266	Significance Z	3.8237*
Observed variance	0.1983	Error variance	0.0788
Population variance	0.1195	Homogeneity Q	102.6985 df= 21*
Amount of variance explained by sampling error	39.73 %		
Kraemer (1983) method			
Mean effect size d	0.2587	95% Confidence interval	0.1698 to 0.3481
Population effect size Rho	0.12828	Variance of rho	0.0005
95% Confidence interval	0.0846 to 0.1715	Homogeneity Chi-square	109.9362
Orwin's Fail-safe n based on "random effects model" DELTA			
Fail-safe for critical d of .20	16.2940	Fail-safe for critical d of .50	-6.6824
Fail-safe for critical d of .80	-12.4265		

Note. * p < .000

Appendix 1 Table 12 Data file using Schwarzer's programme

Study number	Sample size group 1 (experimental group)	Sample Size group 2 (control group)	Effect size	Reliability coefficient
1	10	11	1.1689	1
2	10	11	0.6205	1
3	10	11	0.7554	1
4	10	11	0.1232	1
5	91	92	-0.0082	1
6	91	92	0.87041	1
¹¹ 7	54	54	0	1
8	54	54	0.25	1
9	54	54	1	1
10	54	54	0.083	1
11	54	54	0.2	1
12	35	27	0.3096	1
13	32	31	0.5454	1
14	79	23	1.0013	1
15	34	35	0.4148	1
16	48	24	0.506	1
17	48	24	0.7124	1
18	24	12	1.118	1
19	133	133	-0.5061	1
20	52	48	-0.1301	1
21	52	48	-0.0892	1
22	52	48	0.1559	1

¹¹ Note that during data input it was necessary to swap some studies within Kenny's programme resulting in both programme's data input being in slightly different order



4. Discussion

As the first data run included outliers, these results will not be discussed. The results of the second analyses will however now be discussed.

4.1 Kenny's programme

It must be noted at the outset that scrutiny of normal data should not pose much of a problem seeing as meta-analysis cumulates findings resulting in large enough sample sizes to rest on the assumption of normality through reference to the central limit theorem (Krishnamurty, Kasovia-Schmitt & Ostroff, 1995; Normand, 1995). Twenty-two effect sizes with a sample of 2032 yielded an average effect size of 0.3354 which is significant when the *t* test results (two-tailed) is studied; $3.2863 p < 0.005 df = 21$. Thus the effect size differed significantly from zero resulting in the conclusion that dynamic assessment did in fact have an effect on posttest scores when studies are cumulated. The *t* test treats *study* as unit of analysis, but *z* treats *person* as unit which also happens to be significant in this instance; average $z = 5.5697 p < 0.001$. This answers affirmatively the question presented earlier of whether dynamic assessment intervention makes a significant difference as opposed to no (static) intervention across separate studies on posttest scores. This result, as highlighted earlier, has no bearing on the value of dynamic assessment as method of mediatory tool and caution is attached to the interpretation of this significant value. In other words, regardless of the significance of the original primary findings, what does the cumulative finding mean?

The BESD (binomial effect size) measures the estimated difference between the experimental and control groups in terms of proportions. The test of homogeneity relies on the chosen statistic used to compute the effect size which in this case was *d* and thus Hedge's test of homogeneity was employed by the programme. The chi square statistic is used to compute the test for homogeneity which in this instance is highly significant thus indicating that the studies are not homogenous and hence effect sizes differ due to factors other than sampling error. The fail-safe number generated by Kenny's programme yields 156 null studies which would have to be generated for this test to be not significant i.e. 156 similar studies will need to be uncovered for this result to be nullified (Strube, 1985; Strube & Hartmann, 1983). No studies were transformed by any means other than by making use of Hedge's transformation¹² which as Kenny correctly points out is not truly a new weighting but merely a sample size correction factor (Kenny, 2003), thus, no untransformed estimate of average effect size is produced. Determination of effect size used by Cohen ranges from 0.2 (small effect), 0.5 (medium effect) and 0.8 (large effect) (Schwarzer, 1989). Thus the effect size of 0.334 can be considered as halfway between a small and medium effect size.

4.2 Schwarzer's programme

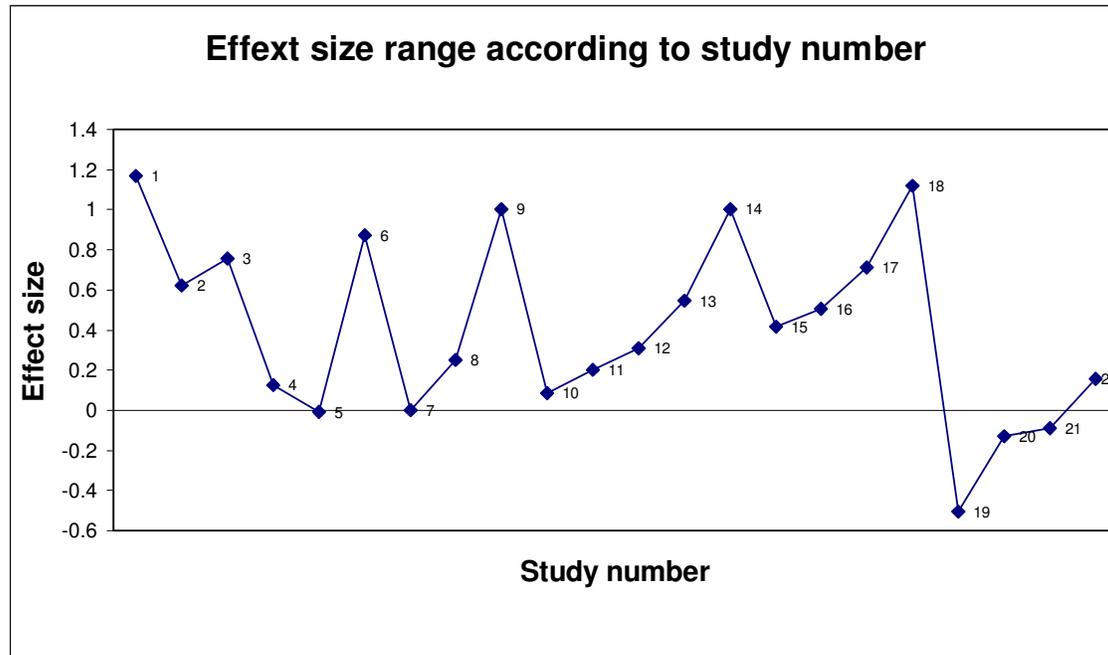
Twenty-two effect sizes with a sample of 2032 yielded an average effect size of 0.23703 which although significant was calculated using the "weighted integration method" and the subsequent chi square statistics yielded a highly significant result thus prompting the user to investigate the "random effect model" instead (as Schwarzer's programme offers three types of output). The mean effect size delta is 0.34813 and is significant in the random effects model. An aspect which is worrying is that 39.73% of the variance explained is due to sampling error. This underlies the original hesitation of running a meta-analysis with data from master's and doctoral studies in which original samples chosen for each study was not randomly chosen according to this meta-analysis. Nevertheless the effect size answers affirmatively the question presented earlier of whether dynamic assessment intervention makes a significant difference as opposed to no (static) intervention across separate studies on posttest scores. Once again, this result has no bearing on the value of dynamic assessment as such. In other words, regardless of the significance of the original primary findings, what does the cumulative finding mean?

The fail-safe number generated by Schwarzer's programme is interpreted differently from that of Kenny's fail-safe number. Schwarzer's number is the amount of studies needed for critical effect sizes of 0.2, 0.5 and 0.5 respectively. As the fail-safe for both the 0.5 and 0.8 delta levels exceed that of 0.34813 these two fail-safe numbers are meaningless. Although Kenny does not specifically mention how the fail-safe number is computed, it is assumed that Rosenthal's formula has been used. Schwarzer however makes use of Orwin's fail-safe number computation which is an adapted version of the original Rosenthal formula. Kenny's effect size of 0.3354 and Schwarzer's effect size of 0.34813 (delta, random effects model) differs by 0.01273 and is slight. Also, values from the various confidence intervals (the 95% confidence intervals described in the weighted integration, random effects and the Kraemer (1983) models used by Schwarzer, 1989) do not contain zeros, further supporting a significant effect; a zero in the interval could possibly indicate that there is no effect (Chambers, 2004). As has been highlighted (Arthur et al., 1994) any differences between packages usually results in fourth and higher decimal place differences which is acceptable.

¹² "A correction which improves the sample estimate of the standardized mean difference between the two groups" (Kalaian & Raudenbush, 1996, p.229).

The similarity of output further minimises any judgement calls used when making the decision to run the analyses on two software programmes. Figure 1 illustrates the range of effect sizes in study number order ranging from study 1 to study 22.

Appendix 1 Figure 1 Range of effect size across study number



4.3 A search for moderators

Due to the heterogeneity of effect sizes (the chi-square distribution indicating that homogeneity was rejected) an effect size cluster analysis was conducted in order to determine the heterogeneity of the data set. The resultant classification of two clusters is evident yet the second cluster comprises only one study and is thus considered an outlier in terms of heterogeneity of effect size when compared to the other 21 studies. A robust search for potential moderators was not carried out due to the small sample size of effect sizes, however this is strongly advised for larger data sets (Chambers, 2004; Eagly & Wood, 1994; Kenny, 2003; Rosenthal, 1995 Schwarzer, 1989). Cluster analysis decomposes the number of effect sizes into smaller sets and effect sizes are rank ordered according to their similarity and although this procedure is suited to smaller effect size samples (as is the case here), it works less well for unequal sample sizes (Schwarzer, 1989). A cursory glance of the output of effect size cluster analysis (see table 13) computed on the data file using Schwarzer's programme evidenced two clusters with only one study located in the second cluster (considered more of an outlier but this terminology is used by Schwarzer) and ranged from across the 1%, 5% and 10% levels of significance (the study by Lloyd et al; study number 19 in the second data run). According to Schwarzer (1989), "the formulae for the critical values and the computer algorithm for the disjoint cluster analysis have been taken from Mullen and Rosenthal, (1985)" (p.33). Additional information is available at the end of table 13 and yields information on average sample size and standard deviation. A larger database would have allowed more probing investigative analysis in terms of moderator effects such as the potential moderating effects of gender, age, level of education and cultural grouping. The quality of the primary studies regarding greater elucidation of sample characteristics and also the small number of studies eventually included did not warrant such an investigation at this stage.

Appendix 1 Table 13 Effect size cluster analysis

CLUSTERS AT 1 % 5% and 10% LEVELS OF SIGNIFICANCE	
CLUSTER 1	
StudyID 1	Effect Size= 1.1689
StudyID 18	Effect Size= 1.1180
StudyID 14	Effect Size= 1.0013
StudyID 9	Effect Size= 1.0000
StudyID 6	Effect Size= 0.8704
StudyID 3	Effect Size= 0.7554
StudyID 17	Effect Size= 0.7124
StudyID 2	Effect Size= 0.6205
StudyID 13	Effect Size= 0.5454
StudyID 16	Effect Size= 0.5060
StudyID 15	Effect Size= 0.4148
StudyID 12	Effect Size= 0.3096
StudyID 8	Effect Size= 0.2500
StudyID 11	Effect Size= 0.2000
StudyID 22	Effect Size= 0.1559
StudyID 4	Effect Size= 0.1232
StudyID 10	Effect Size= 0.0830
StudyID 7	Effect Size= 0.0000
StudyID 5	Effect Size= -0.0082
StudyID 21	Effect Size= -0.0892
StudyID 20	Effect Size= -0.1301
CLUSTER 2	
StudyID= 19	Effect Size= -0.5061
Additional Information	
Average Sample Size = 92.3636	
Sample Size Std. Dev= 58.3006	
Correlation between Sample and Effect Sizes = -0.5098	

4.4 Limitations of the meta-analysis

Firstly only 7 studies were included (out of a potential number of 22) for this meta-analysis primarily due to the unavailability of the two software programs to proceed with the accumulation of within-groups studies and repeated-measures designs. Secondly of these 7 studies, 22 effect sizes were generated and of these 22 effect sizes, 12 were dependent samples and 10 were independent samples thus violating the inherent assumption of independence. The reason for violating this assumption was if only 10 independent samples were to be used the exercise of a meta-analysis would then be rendered null and void. The results should thus be interpreted with due caution. Although there is a significant small to medium effect size evident in the South African research literature pertaining to dynamic assessment as an effective intervention strategy in terms of effecting posttest score results, this effect is nevertheless resultant on the quality of studies included for assessment (mostly master's and doctoral studies); the small number of studies finally included (seven from twenty two studies) and the fact that fifteen of the original studies assessed within-groups results and not between-groups results. This is of itself an important finding which can be used to highlight the differences between the South African and overseas research literature in this area. Had the original pool of studies been larger and more varied in terms of quality, i.e. peer reviewed published results and had as their research designs between-groups as opposed to within-groups designs the results may have been even more significant in terms veracity, applicability and generalisability.

4.5 Implications of the findings for dynamic assessment research in South Africa

The effect size of 0.3354 obtained using Kenny's programme yielded a significant result with a similar effect size of 0.3481 obtained using Schwarzer's programme. In answer to the question of whether dynamic assessment interventions across cumulated South African research indeed had any cumulative significant effect, it can be stated that there is a small to medium effect size across studies thus supporting the utilisation of dynamic assessment in South Africa but only as it pertains to the improvement of posttest score results. The study cannot comment on the value of dynamic assessment as mediatory tool. This serves to add credence to an approach which seeks to assess individuals in as unbiased a manner as possible and which has



as a core philosophy the understanding of individual change through the learning process. These findings do, however, have to be tempered with the fact that these results are based on only 7 studies.

4.6 Recommendations

Meta-analysis is dependent on the accuracy and robustness of primary research data and can never replace the need for primary research. Most of the studies included in this meta-analysis were Master's and Doctoral studies and the data was not, in most instances, in the correct format for a meta-analysis to be conducted. It is recommended that in future primary empirical dynamic assessment studies be conducted with future meta-analyses in sight and in so doing prepare the design and statistical analysis in such a way that the data becomes more amenable to meta-analytic data analysis. Much worthwhile data was not included in this analysis as much of the necessary data information was not included in the primary texts. As more primary research data is added to the field of dynamic assessment in South Africa it will become increasingly important for cumulative studies to be conducted in order to determine the cumulative efficacy of all the research and in order to do just this, the correct format of data needs to be included in the original studies. Although meta-analyses can be conducted on smaller sample sets it is also necessary that more primary empirical research be added to the field of dynamic assessment in order to make even more robust any future final meta-analytic results. It will be of great interest to conduct another such study once the primary pool has increased somewhat.

Regarding the utility of the two meta-analytic software programmes, both Kenny and Schwarzer's programmes yield similar effect sizes, both programmes differ in some ways and are alike in others. The programmes were freely available over the internet and as such functionality was not fully operational (such as the lack of saving and printing facilities) but considering that these programmes run at no cost this cannot be considered a criticism. Although more resources would allow for easier usage of purchased programmes these two programmes are considered worth the use if one is limited in resources. Dynamic assessment is more than a tool utilised for increasing scores within pretest-posttest research studies. Assessing the synthesized effect sizes of qualitative mediatory interventions through meta-analysis may also prove fruitful. However, most quantitative primary studies in South Africa do not study this aspect of dynamic assessment and have as their focus pretest-posttest research designs. It is recommended that primary studies could perhaps quantify such mediatory qualitative studies thus allowing later meta-analysts an opportunity to cumulate effect sizes across studies.

5. Conclusion

The aim of this study was two-fold: to determine the significance of the efficacy of dynamic assessment as a viable assessment strategy in South African studies in terms of increasing scores on posttest test results and to compare and analyse two meta-analytic software programmes. Meta-analysis is a powerful technique which can aid in the determination of how effectual cumulated studies in fact are and may evidence results contrary to individual study results. Twenty two studies were originally coded for inclusion into the meta-analysis but due to the nature of the original data only seven of these studies were included in the final analysis. Effect sizes of 0.3354 and 0.3481 respectively were calculated utilising two meta-analysis software programmes. Due to this small sample size, however, only limited conclusions can be drawn. Nevertheless, it was evidenced that the findings were indeed significant and had as a result the affirmation of the question posed as to whether dynamic assessment as an assessment tool is efficacious within South African research. Limitations surrounding the use of the two packages and violations inherent in the analysis of the data was addressed as further cautions against generalisation of results. It is suggested that any future research designs conducted within the field of dynamic assessment be set out in such a way as to accommodate future meta-analysis as this technique is dependent on data that can be successfully utilised within such an analysis. The recommendation concerning the utility of the two software programmes was based on their ease of use, technical features, similarity of output and cost-effectiveness.

6. Excursion into a potentially rich field of investigation – an aside to this meta-analysis

6.1 From GeneWays to MemeWays¹³

A concept, which has since 1996 come to fruition in the bioinformatics field¹⁴ (in this instance genes), may be of potentially great significance within the behavioural sciences (for example the study of memes) and involves the amalgamation of computer

¹³ Casting such a broad analogy might at first glance seem ill-conceived, however it is the author's contention that if a system such as GeneWays can culminate in such fruitful research within bioinformatics, then it can surely be conceived to bring to fruition a similar functioning system (albeit with severe modifications) to the social sciences and memes are hereby called to mind, serving at once a behavioural science counterpart (Blackmore, 2000) and an equally enticing homonym.

¹⁴ The National Bioinformatics Network is just such an initiative which has been launched in South Africa (at present working with the University of Pretoria as a node within the network among others) and more information can be found at www.nbn.co.za. To see an endeavour of a similar nature burgeon within the social sciences in the field of cognitive informatics would further spur research and discovery in this area.



science, mathematics and information science. Although not intended as a review of this method of data extraction it is considered useful at least at this preliminary stage to consider the advantages that such a data mining technique might hold for fields as diverse and complex as intelligence research.

In essence, "GeneWays" (Rzhetsky, Lossifov, Koike, Krauthammer, Kra, Morris, Yu, Duboué, Weng, Wilbur, Hatzivassiloglou & Friedman, 2004) is an automated search tool that combs through bibliographical research literature databases, a task that is becoming increasingly difficult to conduct due to the sheer number of articles published currently in academia and industry (Stix, 2005). The mining of potential information and new links that can be found within and between a variety of research topics and domains remains within the grasp of behavioural as well as natural sciences. Making use of the speed and efficiency of such a search tool impacts significantly on the number of hidden or hitherto unknown variable linkages that occur within large databases, links that may not necessarily be found in the traditional literature search. Effectively this tool can partially replace the need for researchers to mine information from texts manually. Intelligence research of which dynamic assessment has of late become more prominent, may offer the researcher the opportunity thus far afforded by meta-analytic techniques¹⁵ (see Appendix 1, Section A) the further opportunity of describing links within various data pools by using a technique similar to that of GeneWays. Briefly, the search tool functions as follows:

Online full-text articles are scanned for various key phrases pertinent to the topic at hand and are then downloaded to a computer, after which the articles are stripped of html and any other coding. Processing of the information includes the refining and defining of certain key words and the links these terms have to other phrases or key words. This 'filtered' text is then translated by a parser into machine-readable format (Stix, 2005). Once the information is filed into a database, various query languages can be used to "ask" the database questions, after which information can also be graphically displayed, illustrating links between different variables under scrutiny. Notwithstanding the technique's niche development within the area of molecular biology, the core principle operating within this nexus of research can be applied to other fields and can be harnessed, adapted and utilised within the social sciences model.

Together with tools from artificial intelligence and statistics this endeavour has opened up a vista large enough to warrant its use within social science research. Modelling links between variables, which would otherwise remain dormant,¹⁶ can only aid in identifying veiled connections which can prove fruitful in establishing new domains and avenues to pursue. As much is conveyed in the statement by Wang (2003) in which is stated that along with energy and matter, information is the third essence for modelling the natural world and when combined amalgamate into a new transdisciplinary study of cognitive informatics (p.151). Towards just such an end within the behavioural sciences is the pioneering work being conducted within this field which currently encompasses the study of philosophy, psychology, neuroscience, neuropsychology, cognition (cognitive science), computer science, software engineering, artificial intelligence, mathematics and statistics (Chiew & Wang, 2003; Wang, 2002; Wang & Wang, 2002; Wang, 2003a; Wang, 2003b) which seeks to align these varied areas of interest in such a manner so as to allow for the discovery of how humans process information, sustain consciousness as well as memory storage and retrieval (Wang & Liu, 2003; Wang, Patel, Patel & Wang, 2003; Yao, 2004). Having as forerunners classical information theory through to contemporary classical informatics (the science of information), cognitive informatics can be said to be:

the study of computing and information processing problems by using cognitive science and neuropsychology theory and studies cognitive information processing mechanisms of the brain by using computing and informatics theories
(Wang, 2003b, p.151).

Not all proponents of this new field are, however, in agreement with exactly how such an endeavour should propel itself forward as much of the initial inherent assumptions are not entirely agreed upon (Bryant, 2003). As the above-mentioned list of fields of enquiry suggests, cognitive informatics can be understood as a loose term encompassing myriad definitions, either broad in range or narrowly focused on specified topics. This multi-faceted approach that cognitive informatics takes may result in the evasion of the broader cultural context in which this area of research finds its niche. In an attempt to bring back to this emerging discipline a core ideal of inclusiveness of other co-constructed realities (and not merely an endeavour to understand the transmission of data in a purely technical sense), Bryant argues that before this field becomes too immersed in terminology and

¹⁵ Of course the degree to which one is dependent on the accurate and reliable functioning of just such a search tool does not obviate the need to perhaps control for consistency within the programme in terms of how it might code for certain aspects; an inter-rater reliability scale of sorts could be devised. Meta-analytic techniques frequently make use of such inter-rater reliability scales to ensure the accuracy and reliability of coding across studies, both within and between researchers. However, with the past advances in the artificial intelligence field, this may not necessarily pose such a threat (Mullins, 2005). The process of curation however takes care of assigning levels of confidence to data by annotating original statements with statements of confidence (Rzhetsky et al., 2004).

¹⁶ The fact remains that however global academic society becomes, no-one single human researcher can possibly be cognisant of every other researcher in the world who is actively engaged within the same research area. The need to establish potential links using automated software is one solution that should be encompassed by social scientists who would seek to remain at the forefront of their respective fields of enquiry.



research programmes which may flounder from the broader framework, the field should be recast in a new light which takes cognisance of the information disseminated as well as the *observers* present during such dissemination (Bryant, 2003).

The example of how data and information is transmitted along a conduit is often used to construe the nature of how language and the written word is conveyed which can at times be misleading as a representation of how in fact communication is achieved between human beings for instance. Another major point of contention detailed by Bryant, is the notion of a relational brain as evidence of how in fact cognition takes place within the brain whereas cognitive informatics as understood and propounded by more deterministic and analytic researchers, seems to relate to cognition from a more componential outlook stressing the symbolic nature of cognition. For instance, when recalling from memory one might believe that information is stored in symbolic networks as opposed to recalling a relational structure in which the memory is reactivated within a network and reconstructed each time the memory is needed.¹⁷ A stern warning is issued by Bryant to those hoping to embark on a career within cognitive informatics;

we must be constantly aware of the role of the observer, the embodied nature of information, the lure of the conduit metaphor, the tendency to mechanize and technicize, and the weakness of the functionalist view of representation

(2003, p.227).

The final question to be answered though is what this new emerging field of cognitive informatics has to offer dynamic assessment within intelligence research. How is the power of this methodology to be harnessed and what are the end goals? Purely qualitative behavioural assessments are not truly the domain of quantitative high-speed amalgamated databases and the work carried out within neuroscience laboratories seems too distant to bridge within the assessment of human beings from a dynamic assessment point of view. However, the scope of dynamic assessment, being entrenched within the broader intelligence research field allows for myriad branches of these pursuits which can easily be brought into contact, namely through

- the power of meta-search tools as evidenced by GeneWays;
- the alluring appeal of a multidisciplinary co-operative focus area such as cognitive informatics which can aid in how knowledge is stored and represented within the brain;
- and the manner in which qualitative assessments can evolve alongside such endeavours.

This at least allows practitioners the opportunity to borrow from and inform other emerging and emerged disciplines from across the academic board. In essence, dynamic assessment, like any other research field, has something to offer human beings. It can also learn from other fields. It is the opinion of this author that the confluence of these above-mentioned areas will buoy dynamic assessment in the years to come.

¹⁷ Of course by implication if memory were to work in this manner (i.e. reconstructed each time), reconstructing memories would entail past knowledge and feelings being evoked and thus distorting these memories. How accurate such memories would be is at present a contentious issue.

Section B Further information regarding studies used in the meta analysis

1. Introduction

The purpose of including this summarised information on the primary studies is to allow for replication of meta-analytic results making use of a variety of choice statistics. This also makes provision for those readers seeking more qualitative information regarding each study as this was not included in the meta-analytic analysis. Studies with asterisks are those included in the final analysis (seven in total). It also includes main characteristics of each study and takes the following format:

- (1) Author
- (2) Year
- (3) Title
- (4) Publisher
- (5) Type of publication
- (6) Geographical area
- (7) Short description
- (8) Sample size
- (9) Number of groups
- (10) Between/within groups design
- (11) Ability group
- (12) Chronological age
- (13) Tests used
- (14) Type of feedback
- (15) Design and procedure
- (16) Nature of sample
- (17) Time between pre and posttest
- (18) Hypotheses
- (19) Mediation
- (20) Measures
- (21) Summary of findings
- (22) Criterion and predictor variables
- (23) Limitation

Each study is then summarised using the key number to reflect the data concerned. To ensure a smoother flow when reading the below mentioned results it is deemed better to make use of some shorthand:

Exp = experimental group; cont = control group; X = mean; SD = standard deviation; % = percentage; N = total number; n = number per group; t = t test result; F = f test result, $p < /p>$ = p value at the chosen significance level; btw = between; sig = significant; df = degrees of freedom; H0 = null hypothesis; HA or H1 = alternative hypothesis; Matric = matriculation level (last year of schooling)

(1) Andrews, S.G.**

(2) 1996

(3) A small-scale investigation of the group administration of Feuerstein's learning potential assessment device

(4) University of Natal, Pietermaritzburg

(5) Unpublished Masters dissertation

(6) Kwa-Zulu Natal

(7) To determine if cognition is modified in a group administration of the LPAD; and to detect differences in the degree of modifiability

(8) 21 (originally 24)

(9) 2

(10) Within and between groups design - matched experimental and control groups

(11) Homogenous group - matched pairs in experimental and control groups

(12) Std 7 (new grade 9); therefore average of 14

(13) LPAD - numerical progressions, organizer and complex figure drawing

(14) Assisting questions from pupils, modelling, feedback and cognitive structuring

(15) Pretest-(mediation)-posttest of experimental group; control group received no mediation

(16) Black female high school pupils

(17) 14 days

(18) H01: Mediation will result in modified cognition and improved performance of the experimental group; H02: Group administration will detect differences in degree of modifiability (Andrews does not refer to the hypotheses as "null" but refers to them as main aims)

(19) Mediation in cognitive operations, assisting questions, modelling, contingency management, feedback, cognitive restructuring and assistance in the functions required by instrument - completed in a 2 hour session

(20) Raw scores used for Organizer and Numerical Progressions as presented in Andrews' appendix. Numerical progressions: Exp pretest $X=25$, $SD=15.88$, $n=10$; Exp posttest $X=60$, $SD=15.32$, $n=10$; $t=2.97$ ($p < 0.01$). Cont pretest $X=20.8$, $SD=17.64$, $n=11$; Cont posttest $X=42.5$,

SD=14.65, n=11; t=0.02 (p>0.1). Organizer: Exp pretest X=21.9, SD=18.82, n=10; Exp posttest X=59.5, SD=19.5, n=10; t=4.05 (p<0.005). Cont pretest X=23.3, SD=9.49, n=11; Cont posttest X=46, SD=22.06, n=11; t=0.654 (p>0.1). Complex figure drawing - two phases (copy and recall): Copy: Exp pretest X=16.2, SD=2.23, n=10; Exp posttest X=15.5, SD=1.96, n=10. Cont pretest X=16.7, SD=1, n=11; Cont posttest X=13.5, SD=3.14, n=11. Recall: Exp pretest X=13.4, SD=4.86, n=10; Exp posttest X=7.2, SD=3.10, n=10; Cont pretest X=12.5, SD=2.93, n=11; Cont posttest X=6.8, SD=3.13, n=11. T test performed in order to determine whether there was any sig improvement from pre to posttest scores across all three instruments: Exp group t=1.75 at p<0.1 and for cont group: t= -6.87 at p<0.0005 - however, Andrews discards this aggregate across three tests due to an extraneous variable (complex figure drawing being more difficult in posttest than in pretest)

(21) Regarding H01 and H02 for numerical progressions and organizer both null hypotheses are supported (recall that Andrews refers to these hypotheses as "aims" and not "null", hence support for these aims translates into support for the null hypotheses although this would not be the conventional manner of describing it). Only limited support for both hypotheses for the complex figure drawing due to unforeseen factors.

(22) Criterion: only results on the posttest - compared with results on the pretest

(23) Small sample sizes, terminology of Organizer not familiar to students and first language of students was not English (the language of mediation). Complex figure drawing in posttest was more difficult than the pretest drawing.

(1)Shochet, I.M.

(2)1986

(3) Manifest and potential performance in advantaged and disadvantaged students

(4) University of the Witwatersrand

(5) Unpublished PhD thesis

(6) Johannesburg

(7) To search for culture-fair alternative predictors of tertiary academic success by making use of Feuerstein's philosophy (as encapsulated within the LPAD) and to operationalise the LPAD constructs for this environment by enriching static assessment tools

(8) 156

(9) 2

(10) Between and within groups design (dependent samples used)

(11) Advantaged and disadvantaged first year Arts faculty (BA) students at the University of the Witwatersrand. 52 male and 104 female

(12) Mean age 19.56 and range 17-31

(13) Deductive reasoning test (DRT) (both static and enriched versions, i.e. DR/T and DR/E) and Pattern relations test (PRT) (both static and enriched versions, i.e. PR/T and PR/E), both tests developed in South Africa. The differences in scores between the two yielded the difference score (DR/D and PR/D)

(14) Immediate. Mediation followed straight after the pretest. Feedback is given after each problem as well as modelling some of the logic and thinking processes necessary to solve the particular problems, giving advance warning about typical errors that could be made, giving students the opportunity in subsequent items to transfer this learning and thus demonstrate learning potential

(15) Pretest-mediation-posttest design for both DRT (static and enriched) and PRT (static and enriched) tests - Group administered to both the advantaged and disadvantaged groups (therefore 4 administration groups)

(16) Female n = 104 in total and divided into 77 advantaged and 27 disadvantaged students; male n = 52 in total and divided into 27 advantaged and 25 disadvantaged students

(17) immediate - pretest followed by mediation and posttest

(18) HA1: Advantaged and disadvantaged students will have different predictors correlating significantly with the criterion of university success. HA2: The prediction of university success will be significantly enhanced through the introduction of the enriched testing conditions as conceptualised by Feuerstein and operationalised for the purpose of the present study. Sub-hypothesis HA2 - The prediction of university success will be significantly enhanced for disadvantaged students through the enriched testing conditions

(19) Feedback is given after each problem and modelling some of the logic and thinking processes necessary to solve the particular problems, giving advance warning about typical errors that could be made, giving students the opportunity in subsequent items to transfer this learning and thus demonstrate learning potential

(20) To determine whether there was a sig difference btw the traditional and enriched versions of the tests (i.e. in order to determine whether mediation was effective), t-tests for paired differences for dependent samples was used. DRT Whole group: N=156, mean DRT (static) = 16.89, mean DRT (enriched) = 20.93 and mean DRT (difference btw static and enriched) = 4.13; t=10.94 significant at p=0.001. Disadvantaged group n=52, mean DRT (static) = 12.06, mean DRT (enriched) = 16.42 and mean DRT (difference btw static and enriched) = 4.23; t=6.62 significant at p=0.001. Advantaged group n=104, mean DRT (static) = 19.31, mean DRT (enriched) = 23.18 and mean DRT (difference btw static and enriched) = 4.07; t=8.69 significant at p=0.001. PRT Whole group: N=156, mean PRT (static) = 10.56, mean PRT (enriched) = 17.32 and mean PRT (difference btw static and enriched) = 6.75; t=19.87 significant at p=0.001. Disadvantaged group, n=52, mean PRT (static) = 7.29, mean PRT (enriched) = 12.85 and mean PRT (difference btw static and enriched) = 5.19; t=7.78 significant at p=0.001. Advantaged group, n=104, mean PRT (static) = 12.20, mean PRT (enriched) = 19.56 and mean PRT (difference btw static and enriched) = 7.53; t=20.68 significant at p=0.001.

Measures for investigating HA1: Pearson product-moment correlations were conducted: all predictor variables were correlated with both criterion variables (credits and average scores, see number 22 below) for each group (advantaged and disadvantaged groups). Advantaged group: Matric & credits = 0.34 (sig at p=0.01) two-tailed, matric & average 0.55 (sig at p=0.01), DRT (static) & credits -0.13, DRT (static) & average -0.04; DRT (enriched) & credits -0.19, DRT (enriched) & average -0.08; DRT (difference btw static and enriched) & credits -0.07, DRT (enriched) & average -0.07, PRT (static) & credits -0.12, PRT (static) & average -0.06, PRT (enriched) & credits -0.07, PRT (enriched) & average -0.04, PRT (difference btw static and enriched) & credits 0.10, PRT (difference btw static and enriched) & average 0.05. Disadvantaged group: Matric & credits = 0.20, matric & average 0.15, DRT (static) & credits 0.25, DRT (static) & average 0.26; DRT (enriched) & credits 0.03, DRT (enriched) & average 0.03; DRT (difference btw static and enriched) & credits -0.26, DRT (enriched) & average -0.29 (sig at 0.05 two tailed), PRT (static) & credits 0.16, PRT (static) & average 0.21, PRT (enriched) & credits 0.05, PRT (enriched) & average 0.0, PRT (difference btw static and enriched) & credits 0.06, PRT (difference btw static and enriched) & average 0.03. Therefore different predictors for both groups. In order to determine whether there were sig diff btw the sig correlations for both groups, the correlations underwent Fisher's r-z transformations. Matric & credits 0.829, matric & average 2.565 (sig at p<0.05 two tailed), DRT (difference btw static and enriched) & credits -1.846, DRT (difference btw static and enriched) & average -2.027 (sig at p<0.05 two tailed), DRT (static) & credits -2.126 (sig at p<0.05 two tailed), DRT (static) & average -1.681. Hence some differences are significant for the two groups: therefore support for HA1.

Investigating HA2 through a main effect: i.e. to establish whether Feuersteinian mediation would significantly improve prediction from traditional measures. ADVANTAGED students only: Stepwise multiple regression models with unstandardised beta weights: dependent variable = CREDITS traditional measures - matric 0.08(beta), F(df1/88) 16.93, p = 0.000(sig at 0.05), PRT (static) -0.04(beta), F(df1/88) 1.98, p = 0.163, DRT (static) -0.05(beta), F(df1/88) 3.91, p = 0.051. Multiple r squared = 0.184(sig at 0.05 and 0.01). No Feuersteinian measure met the 0.10 sig level for entry into the model after the above model had been included therefore no Feuersteinian measures enhanced prediction. The same analysis was run with dependent variable: AVERAGE: traditional measures - matric 0.70(beta), F(df1/88) 45.76, p0.000(sig at 0.05 and 0.01), PRT (static) -0.22(beta), F(df1/88) 1.82, p=0.181, DRT (static) -0.22(beta), F(df1/88) 2.98, p=0.088 multiple r squared = 0.345(sig at 0.05 and 0.01). No Feuersteinian measure met the 0.10 sig level for entry into the model after the above model had been included therefore no



Feuersteinian measures enhanced prediction. DISADVANTAGED students only: Stepwise multiple regression models with unstandardised beta weights: dependent variable = CREDITS traditional measures - matric 0.05(beta), $F(df1/44)$ 0.96, $p=0.332$, PRT(static) 0.04(beta), $F(df1/44)$ 0.59, $p=0.448$, DRT (static) 0.05(beta), $F(df1/44)$ 1.78, $p=0.190$ multiple r squared = 0.094. No Feuersteinian measure met the 0.10 sig for entry into the model after the above model had been included level therefore no Feuersteinian measures enhanced prediction. The same analysis was run with dependent variable: AVERAGE: step 0 traditional measures - matric 0.16(beta), $F(df1/44)$ 0.34, $p=0.562$, PRT(static) 0.33(beta), $F(df1/44)$ 1.01, $p=0.320$, DRT (static) 0.30(beta), $F(df1/44)$ 2.26, $p=0.140$ multiple r squared=0.096. Step 1 traditional measures - matric 0.11(beta), $F(df1/44)$ 0.16, $p=0.689$, PRT(static) 0.29(beta), $F(df1/44)$ 0.81, $p=0.373$, DRT (static) 0.30(beta), $F(df1/44)$ 2.48, $p=0.122$ DRT(enhanced) -0.42 (beta), $F(df1/44)$ 3.52 $p=0.067$ r squared=0.165. No Feuersteinian measure met the 0.10 sig level for entry into the model therefore no Feuersteinian measures enhanced prediction. No predictors were significant for disadvantaged students. HA2 is not supported through a main effect for either group.

Investigating HA2 through a moderator effect: ADVANTAGED students: two models used with different criteria: model 1 - predictors include DRT (static), DRT (difference) DRT (static) X DRT (difference): CREDITS: DRT (static) -0.06(beta) $F(df1/88)$ 2.86 $p=0.094$, DRT (difference) -0.12 (beta) $F(df1/88)$ 1.14 $p=0.288$, DRT(static) X DRT (difference) 0.00 (beta) $F(df1/88)$ 0.61 $p=0.438$ multiple r squared = 0.038. AVERAGE: DRT (static) -0.23(beta) $F(df1/88)$ 1.16 $p=0.284$, DRT (difference) -0.76 (beta) $F(df1/88)$ 1.31 $p=0.256$, DRT(static) X DRT (difference) 0.03 (beta) $F(df1/88)$ 0.91 $p=0.342$ multiple r squared = 0.019. Model 2 predictors include PRT (static), PRD (difference), PRT (static) X PRD (difference). CREDITS: PRT (static) 0.00(beta) $F(df1/88)$ 0.00 $p=0.979$, PR (difference) 0.09 (beta) $F(df1/88)$ 0.92 $p=0.341$, PRT(static) X PRT (difference) -0.01 (beta) $F(df1/88)$ 0.60 $p=0.442$ multiple r squared = 0.026. AVERAGE: PRT (static) -0.08(beta) $F(df1/88)$ 0.06 $p=0.805$, PRT (difference) 0.10(beta) $F(df1/88)$ 0.03 $p=0.859$, DRT(static) X DRT (difference) 0.00 (beta) $F(df1/88)$ 0.00 $p=0.965$ multiple r squared = 0.005 HA2 is not supported through a moderator effect for advantaged students. DISADVANTAGED students: two models used with different criteria: model 1 - predictors are DRT (static), DRT (difference) DRT (static) X DRT (difference): CREDITS: DRT (static) 0.12(beta) $F(df1/45)$ 6.81 $p=0.012$ (sig at 0.05), DRT (difference) 0.14 (beta) $F(df1/45)$ 1.37 $p=0.248$, DRT(static) X DRT (difference) -0.02 (beta) $F(df1/45)$ 3.47 $p=0.069$ multiple r squared = 0.192 (sig at 0.05). AVERAGE: DRT (static) -0.53(beta) $F(df1/45)$ 3.72 $p=0.060$, DRT (difference) 0.16 (beta) $F(df1/45)$ 0.05 $p=0.822$, DRT(static) X DRT (difference) -0.05 (beta) $F(df1/45)$ 0.85 $p=0.363$ multiple r squared = 0.165 (sig at 0.05). Model 2 predictors are PRT (static), PRD (difference), PRT (static) X PRD (difference). CREDITS: PRT (static) -0.01(beta) $F(df1/45)$ 0.01 $p=0.929$, PR (difference) -0.09 (beta) $F(df1/45)$ 0.85 $p=0.362$, PRT(static) X PRT (difference) 0.02 (beta) $F(df1/45)$ 1.95 $p=0.170$ multiple r squared = 0.079. AVERAGE: PRT (static) -0.02(beta) $F(df1/45)$ 0.00 $p=0.972$, PRT (difference) -0.62(beta) $F(df1/45)$ 1.31 $p=0.259$, DRT(static) X DRT (difference) 0.11 (beta) $F(df1/45)$ 2.46 $p=0.124$ multiple r squared = 0.100. For the deductive reasoning models, the multiple r squared on both criteria is significant at 0.05. This is not the case for the pattern relations model. DRT (difference) functions as a moderator of the relationship between DRT (static) and success. In order to determine whether the moderator effect is sufficiently strong to support HA2, the multiple r squared of DRT (static) must be compared to the multiple r squared of the moderated model. DRT (static) r squared = 0.063 against criterion of CREDITS and DRT (static) has an r squared = 0.068 against the criterion of AVERAGE. Thus for CREDITS the increment to r squared when DRT (static) is moderated by DRT (difference) is 0.063 subtracted from 0.192, increment to r squared = 0.129, $F(df2/45) = 3.64$ (sig at 0.05). For AVERAGE the increment to r squared when DRT (static) is moderated by DRT (difference) can be determined by subtracting 0.068 from 0.165, increment to r squared = 0.097, $F(df2/45) = 2.6$ which is not sig. Thus HA2 for disadvantaged students has been supported on the criterion of CREDITS. For every increase in modifiability, predictability on DRT (static) decreased. Other findings such as whether age and gender were predictors were also calculated as well as for attendance at an academic support programme.

(21) Enhanced results provided greater predictive value for those students who were more modifiable - but acted as moderator variables. The more the disadvantaged students were able to modify their cognitive functioning, the less predictable was their university success on the basis of the traditional intelligence test and vice versa. Mediation of tests can aid in predicting tertiary success.

(22) Criterion variables: University success at the end of the first year of studies - number of credits obtained in first year of study and as an alternative, the average marks obtained in BA 1. Predictor variables: both static and enriched versions of the DRT and PRT as well as Matric results. Also the difference btw the static and enriched conditions for both the DRT and PRT. Subject variable: level of disadvantage (advantaged and disadvantaged); sex, age, attendance at a special programme (attendees and non attendees).

(23) All four administration groups were of different sizes, which may have impacted on type and quality of mediation, hence ANOVA's were conducted to determine whether means differed significantly for the smaller groups, however, no sig. difference was found; $F1.96$ (for DRT/D) and $F0.49$ (for PRT/D) ($df=3$ for each) and $p=0.120$ and 0.696 respectively. PRT and DRT difference scores used as index of modifiability in the ANOVA. No difference in Scheffe post-hoc test. 15 students cancelled their studies hence sample size decreased from 156-141 (therefore no criterion results for them).

(1) Zolezzi, S.A.

(2) 1992

(3) Alternative selection measures for university undergraduate admissions

(4) University of the Witwatersrand

(5) Unpublished Masters dissertation

(6) Johannesburg

(7) The study aimed to assess the effectiveness of both traditional and learning process selection measures (including dynamic assessment) among a group of advantaged and disadvantaged students

(8) 26

(9) 1 group divided into two and further sub-divided into 4

(10) Within groups

(11) Students enrolled in the pre-university bursary scheme at the University of the Witwatersrand within the commerce Faculty. 14 male and 12 female. Whole group divided into advantaged and disadvantaged; also into low modifiable and high modifiable

(12) 17-25

(13) Static tests used: Biographical questionnaire (BQ); Intelligence test (MAT); inductive reasoning test static (PRT/T); an interview measure (IM), and Matriculation marks (MATRIC). Dynamic tests used: Pattern relations test enriched (PRT/E); the learning process measure (LSP); study process questionnaire of Biggs; the learning and study strategies inventory (LASSI). Learning potential is defined as the difference between the enriched and traditional scores.

(14) Enriched testing condition comprised four stages: introductory patter, intensive mediation, minimal mediation and no mediation.

(15) Group administration of tests. Pretest-mediation-posttest for whole group, statistically analysed afterwards according to group membership i.e. advantaged, disadvantaged and high/low modifiability

(16) Advantaged and disadvantaged commerce students

(17) Both static and dynamic testing took two days

(18) HA1: Learning potential is a better predictor of academic competence for the disadvantaged students than a traditional measure of general intelligence. HA2: learning potential is a better predictor of academic competence for the disadvantaged students than school marks. HA3: learning potential together with the learning process measures is a better predictor of academic competence for both advantaged and disadvantaged students than only learning potential or static measures alone.

(19) The criteria for mediation followed Feuerstein's principles of mediated learning experience

(20) Simple statistics for predictor and criterion variables. Whole sample. $N = 26$. First figure = Mean and second figure = standard deviation (incl. the sub-groupings within each predictor test, for instance LASSI includes L1-L10, and SPQ Biggs includes B1-B9). For ease of reading, the variables have been underlined. Matric 21.11, 6.80 , IQ 26.96, 6.35, PRT/T 8.88, 3.74, PRT/E 16.11, 4.81, LP 8.07, 3.49, BQ 57.15, 6.86, INTERVIEW 48.26, 11.61, LSP 49.57, 11.43, LASSI L1 32.07, 6.38, L2 30.03, 5.21, L3 25.57, 4.64, L4 23.53, 5.96, L5 28.8, 5.51, L6 27.26, 6.74, L7 17.42, 3.61, L8 25.92, 4.16, L9 26.61, 5.67, L10 27.65, 5.15; SPQ - B1 26.11, 3.99, B2 21.61, 4.68, B3 23.61, 4.8, B4 23.53, 5.47, B5 26.42, 4.65, B6 25.15, 5.55, B7 47.73, 7.32, B8 47.15, 8.71, B9 51.57, 7.93; Criterion: June Accounting 41.5, 17.25, June Stats 45.19, 15.72, June Maths 52.15, 13.6, June Average 46.07, 12.38, June Business studies 64.6, 14.25. ADVANTAGED students only: $N = 8$. First figure = Mean and second figure = standard deviation (incl. the sub-groupings within each predictor test): Matric ($N=7$) 22.28, 3.72 , IQ 30.50, 5.58, PRT/T 11.37, 4.92, PRT/E 18.37, 2.87, LP 6.25, 2.81, BQ 62.25, 7.30, INTERVIEW ($N=7$) 52.00, 6.08, LSP 49.5, 12.61, LASSI L1 34.12, 4.51, L2 30.62, 4.47, L3 25.00, 4.59, L4 21.12, 6.49, L5 26.25, 5.39, L6 25.12, 8.85, L7 18.00, 3.54, L8 24.37, 4.10, L9 26.00, 7.15, L10 28.25, 6.08; SPQ - B1 27.5, 2.67, B2 23.75, 3.49, B3 23.62, 3.33, B4 22.37, 5.70, B5 25.62, 4.95, B6 23.75, 3.57, B7 51.25, 3.57, B8 46.00, 6.04, B9 49.37, 6.47; Criterion: June Accounting 42.87, 19.04, June Stats 46.5, 20.24, June Maths 55.25, 15.58, June Average 48.00, 16.86, June Business studies 65.37, 18.50. DISADVANTAGED students only: $N = 18$. First figure = Mean and second figure = standard deviation (incl. the sub-groupings within each predictor test): Matric ($N=17$) 23.11, 2.57 , IQ 25.38, 6.16, PRT/T 7.77, 2.53, PRT/E 16.33, 3.88, LP 8.55, 3.79, BQ 54.88, 5.44, INTERVIEW 49.5, 6.39, LSP 49.61, 11.25, LASSI L1 31.16, 6.98, L2 29.77, 5.61, L3 25.83, 4.76, L4 24.61, 5.56, L5 29.94, 5.31, L6 28.22, 5.61, L7 17.16, 3.71, L8 26.61, 4.11, L9 26.88, 5.09, L10 27.38, 4.85; SPQ - B1 25.5, 4.38, B2 20.66, 4.91, B3 23.61, 5.42, B4 24.05, 5.45, B5 26.77, 4.62, B6 25.77, 6.22, B7 46.16, 8.08, B8 47.66, 9.77, B9 52.55, 8.48; Criterion: June Accounting 40.88, 16.94, June Stats 44.61, 13.92, June Maths 50.77, 12.87, June Average 45.22, 10.29, June Business studies ($N=17$) 64.23, 12.43. LOW MODIFIABLE students only: $N=17$. First figure = Mean and second figure = standard deviation (incl. the sub-groupings within each predictor test): Matric ($N=15$) 22.73, 2.31 , IQ 27.35, 6.66, PRT/T 9.58, 3.93, PRT/E 15.82, 3.50, LP 5.88, 1.79, BQ 56.82, 6.28, INTERVIEW 51.11, 6.42, LSP 50.88, 11.98, LASSI L1 33.23, 4.49, L2 30.41, 4.59, L3 24.52, 4.93, L4 22.47, 6.52, L5 28.76, 5.64, L6 28.29, 6.74, L7 17.58, 3.44, L8 25.58, 4.22, L9 26.76, 5.81, L10 28.52, 5.36; SPQ - B1 26.64, 4.10, B2 21.17, 3.74, B3 22.94, 4.84, B4 23.11, 5.77, B5 25.94, 3.99, B6 24.05, 5.29, B7 47.82, 6.39, B8 46.05, 8.66, B9 50.00, 7.80; Criterion: June Accounting 41.94, 16.80, June Stats 42.82, 17.52, June Maths 51.82, 15.28, June Average 45.29, 13.97, June Business studies 65.29, 14.52. HIGH MODIFIABLE students only: $N=9$. First figure = Mean and second figure = standard deviation (incl. the sub-groupings within each predictor test): Matric 23.11, 3.82, IQ 26.22, 6.03, PRT/T 7.55, 3.12, PRT/E 19.11, 3.10, LP 11.55, 3.32, BQ 57.77, 8.21, INTERVIEW ($N=8$) 48.25, 5.89, LSP 47.11, 10.51, LASSI L1 29.88, 8.88, L2 29.33, 6.46, L3 27.55, 3.43, L4 25.55, 4.36, L5 28.88, 5.60, L6 25.33, 6.68, L7 17.11, 4.10, L8 26.55, 4.21, L9 26.33, 5.72, L10 26.00, 4.55; SPQ - B1 25.11, 3.78, B2 22.44, 6.26, B3 24.88, 4.75, B4 24.33, 5.07, B5 27.33, 5.87, B6 27.22, 5.73, B7 47.55, 9.27, B8 49.22, 8.92, B9 54.55, 7.73; Criterion: June Accounting 40.66, 19.07, June Stats 49.66, 11.12, June Maths 52.77, 10.52, June Average 47.55, 9.20, June Business studies 63.12, 14.53. Correlations conducted for whole group, also for advantaged and disadvantaged sub-groups; and low/high modifiable sub-groups. Only noting sig correlations for the whole group ($N=26$ unless otherwise indicated): LSP and June Maths (0.55483, $p = 0.0033$); LSP and June Average (0.40393, $p = 0.0407$); L1 and June Stats (0.45108, $p = 0.0207$); B4 and June Business studies (0.42643, $p = 0.0335$, $N=25$); B8 and June Business studies (0.43326, $p = 0.0305$, $N=25$).

(21) All three hypotheses were supported, however, HA1 and HA2 were not supported through a main effect. Prediction was enhanced for academic success by supporting HA1 and HA2 through a moderator variable. Advantaged and disadvantaged students predict differently. The more modifiable the student, the less traditional predictors predicted and vice versa.

(22) Criterion variables: July examinations in three subject areas (accounting, Maths and stats) as well as the average score for all three subjects; business studies results; subject variables: advantaged and disadvantaged students; predictor variables: LP (enriched minus traditional); PRT (enriched); LSP, SPQ, LASSI, PRT (traditional), IM, BQ, MAT, MATRIC. There were various sub-groupings regarding the predictor variables, totalling 27 predictors, 5 criterion and 2 subject variables, namely, advantaged and disadvantaged students.

(23) The study did not explore the dynamics involved in the mediational process. Selection of students took place during the second semester, thus allowing students to benefit from enriched teaching. The number of predictors was larger than sample size ($N=26$ and predictors = 27).

(1) Lipson, L.E.

(2) 1992

(3) Relationship of static and dynamic measures to scholastic achievement of black pupils

(4) University of the Witwatersrand

(5) Unpublished Masters thesis

(6) Johannesburg

(7) This study aimed to compare the General Scholastic Aptitude Test (GSAT) with Feuerstein's Learning Potential Assessment Device (LPAD) in terms of the ability of each to predict achievement on a range of school subjects

(8) 27

(9) 1 divided into 2 for ease of administration

(10) Within groups

(11) Standard Four (grade 6) pupils from an English medium school serving a black community. 13 male and 14 female

(12) Mean of 11.84 and SD 0.53

(13) LPAD (selected subtests) and GSAT

(14) Mediation as prescribed by Feuerstein

(15) Pretest-mediation-posttest design for entire sample. The sample was divided into two groups to ensure effectiveness of mediation. Sample 1 $n = 13$ and sample 2 $n = 14$. Within group design. Independent variables: GSAT scores, pretest LPAD scores, posttest scores on the LPAD, length of time of enrolment at school. Dependent variable: terms marks and examination marks

(16) Standard Four black pupils from an English medium school serving a black community.

(17) Both LPAD and GSAT administered over a 4 day period. Pretesting and posttesting took place either on the same day or were presented a day apart (Raven's and set variations)

(18) No hypotheses were stated but questions were stated instead: Q1 - Which battery is the better predictor in school subjects? Q2 - Do pupils who are provided with mediation in a test situation improve in their learning and their post-mediation performance? Q3 - Is there a positive relationship between the length of time a pupil has attended a school which provides adequate learning opportunities and his/her performance on the LPAD and GSAT? Q4 - Is the distribution of scaled GSAT scores among these pupils comparable to the general population?

(19) Interaction between tester and tantee, mediation, understanding of applications and cognitive strategies underlying the tasks according to the principles of MLE as prescribed by Feuerstein

(20) Three sets of results included; namely correlations btw school results and the GSAT and LPAD batteries. Secondly, regression analysis compares the predictability of school achievement on each battery. Lastly a comparison of pre and posttest scores on the LPAD. Lipson includes other results not pertinent to the meta-analysis.

First set of results - Significant correlations - GSAT and LPAD with MARCH MARKS/JULY MARKS AND JULY EXAMINATION MARKS.

Key: only includes those variables that were sig

(* $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$)

GSAT	General Scholastic Aptitude Test
GWA	word analogies
GVERBREA -	verbal reasoning
GNUMPROB -	number problems
GWRODPRS -	word pairs
GSERIES -	number series
GPATCOMP -	pattern completion
GFIGAN -	figure analogies
GRAWV -	raw verbal
GRAWNV -	raw non-verbal
LPAD -	Learning Potential (Propensity) Assessment Device
NUMPRE -	numerical progressions pretest
NUMPOST -	numerical progressions posttest
RPRETOT -	Raven's pretest
RPRETRANS -	Raven's pretest transfer
RPOSTOT -	Raven's posttest
RPOSTRANS -	Raven's posttest transfer
PRECOPSD -	Complex Figure Drawing - sequence of drawing copied on pretest
POSCOPPD -	Complex Figure Drawing - precision of drawing copied on posttest
POSCOPPL -	Complex Figure Drawing - precision of location copied on posttest
POSCOPSD -	Complex Figure Drawing - sequence of drawing copied on posttest
POSMEMSD -	Complex Figure Drawing - sequence of drawing memorised on posttest
TRACOPSD -	Complex Figure Drawing - sequence of drawing copied on transfer
TRAMEMN -	Complex Figure Drawing - number of item memorised on transfer

Subjects include: English, Afrikaans (Afr), Maths, Geography, Health, Science and Biology.

GSAT and LPAD with March Marks

GSAT: GWA - English 0.65(***), Afr 0.54(**), Maths 0.54(***), geography 0.71(***), health 0.59 (**), science 0.65(***), biology 0.71(***). GVERBAL REASONING - English 0.54(**), Afr 0.49(*), Maths 0.46(*), geography 0.57(**), health 0.53 (**), science 0.49(*), biology 0.63(***). GNUMBER PROBLEM - English 0.41(*), Afr 0.39(*), Maths 0.42(*), geography 0.60(**), health 0.61 (**), science 0.54(**), biology 0.42(*). GWORDPROCESSING -English 0.67(***), Afr 0.57(**), Maths 0.61(***), geography 0.62(***), health 0.51 (**), science 0.55*(*), biology 0.51(**). GSERIES - health 0.38(*). GPATCOMP - health 0.52(**). GFIGAN - English 0.56(**), Afr 0.38(*), Maths 0.51(**), geography 0.67(***), health 0.60 (**), science 0.66(***), biology 0.49(*).GRAWV - English 0.68(***), Afr 0.60(**), Maths 0.60(**), geography 0.74(***), health 0.66 (**), science 0.66(***), biology 0.70(***). GRAWNV - Maths 0.43(*), geography 0.52(**), health 0.64 (**), science 0.51(**), biology 0.50(*). LPAD - NUMPRE English 0.50(**), Maths 0.76(***), geography 0.40(*), health 0.59 (**), science 0.40(*). NUMPOST - maths 0.51(**), health 0.47(*). RPRETOT maths 0.50(**), geography 0.45(*), health 0.41(*), science 0.48(*). RPRETRANS English 0.43(*), Maths 0.52(**), geography 0.58(**), health 0.46 (*), science 0.42(*), biology 0.48(*). RPOSTRANS English 0.43(*), Maths 0.52(**), geography 0.58(**), health 0.46 (*), science 0.42(*), biology 0.51(**). PRECOPSD - health 0.41 (*), science 0.41(*). POSCOPD - maths 0.43(*). POSCOPPL- maths 0.46(*). TRAMEMN - geography 0.43(*), biology 0.42(*)).

GSAT and LPAD with JULY MARKS

GSAT: GWA - English 0.70(***), Afr 0.61(***), Maths 0.61(***), geography 0.76(***), health 0.53 (**), history 0.75(***), science 0.75(***), biology 0.55(***). GVERBAL REASONING - English 0.58(**), Afr 0.43(*), Maths 0.63(***), geography 0.50(***), health 0.48 (**), history 0.71 (**), science 0.44(*), biology 0.47(*). GNUMBER PROBLEM - Maths 0.40(*), geography 0.42(*), history 0.61 (**). GWORDPROCESSING -English 0.58(**), Afr 0.59(***), Maths 0.62(***), geography 0.61(***), health 0.60 (**), history 0.70(***), science 0.54(**), biology 0.59(***). GSERIES - history 0.49(**). GPATCOMP - history 0.42(*). GFIGAN - English 0.58(**), Afr 0.56(**), Maths 0.56(**), geography 0.56(***), health 0.40 (*), history 0.64(***), science 0.56(***).GRAWV - English 0.67(***), Afr 0.60(***), Maths 0.69(***), geography 0.70(***), health 0.59 (**), history 0.84(***), science 0.64(***), biology 0.56(**). GRAWNV - English 0.48(*), Afr 0.40(*), Maths 0.42(*), geography 0.41(*), health 0.27, history 0.64(***), science 0.46(*), biology 0.30. LPAD - NUMPRE English 0.59(**), Afr 0.42(*), Maths 0.73(***), geography 0.56(**), health 0.39 (*), history 0.57(***), science 0.45(*), biology 0.48(*). NUMPOST - English 0.42(*), maths 0.47(*), history 0.55(***), science 0.22, biology 0.31. RPRETOT English 0.40 (*), maths 0.50(**), geography 0.46(*), history 0.50(**), biology 0.40(*). RPRETRANS maths 0.48(*), geography 0.47(*), history 0.48 (*), biology 0.39(*). RPOSTOT English 0.47(*), Maths 0.52(**), geography 0.47(*), history 0.56 (**). RPOSTRANS English 0.44(*), Maths 0.48(*), geography 0.46(*), history 0.57 (**), science 0.39(*). POSCOPPL- maths 0.38(*). PSOMEMSD - English 0.42(*)).

GSAT and LPAD JULY EXAMINATION

GSAT: GWA - English 0.73(***), Afr 0.58(***), Maths 0.67(***), geography 0.68(***), health 0.66 (**), history 0.56(**), science 0.86(***), biology 0.73(***), average achieve 0.77(***). GVERBAL REASONING - English 0.71(***), Afr 0.50(**), Maths 0.58(**), geography 0.52(**), health 0.60(***), history 0.67 (**), science 0.68(***), biology 0.63(***), average achieve 0.70(***). GNUMBER PROBLEM - English 0.45(*), Afr 0.52(**), Maths 0.49(**), geography 0.55(**), history 0.42(*), science 0.50 (**), biology 0.41(*), average achieve 0.51(**). GWORDPROCESSING -English 0.74(***), Afr 0.61(***), Maths 0.61(***), geography 0.55(**), health 0.71 (**), history 0.42(*), science 0.64(***), biology 0.76(***), average achieve 0.7(***). GPATCOMP - history 0.46(*). GFIGAN - English 0.59(***), Afr 0.56(**), Maths 0.61(***), geography 0.62(***), health 0.46 (*), history 0.62(***), science 0.70(***), biology 0.65(***), average achieve 0.66(***).GRAWV - English 0.80(***), Afr 0.66(***), Maths 0.71(***), geography 0.70(***), health 0.71 (**), history 0.64(***), science 0.82(***), biology 0.77(***), average achieve 0.82(***). GRAWNV - English 0.46(*), Afr 0.41(*), Maths 0.48(*), geography 0.47(*), health 0.39(*), history 0.58(**), science 0.60(***), biology 0.55(**), average achieve 0.53(**). LPAD - NUMPRE English 0.55(**), Afr 0.47(*), Maths 0.69(***), geography 0.49(**), health 0.50 (**), history 0.43(*), science 0.44(*), biology 0.55(***), average achieve 0.63(***). NUMPOST-maths 0.42(*), geography 0.39(*), health 0.42(*), history 0.46(*), biology 0.44(*) average achieve 0.43(*). RPRETOT English 0.43 (*), Afr 0.48(*) maths 0.49(**), geography 0.54(**), history 0.46(*), biology 0.40(*), average achieve 0.50(**). RPRETRANS English 0.44(*), Afr 0.50(**), maths 0.48(*), geography 0.56(***), health .39(*), history 0.48 (*), science 0.40(*), biology 0.41(*), average achieve 0.50(**). RPOSTOT English 0.53(**), Afr 0.56(**), Maths 0.53(**), geography 0.60(***), history 0.50 (**), science 0.45(*), biology 0.42(*), average achieve 0.56(***). RPOSTRANS English 0.51(**), Afr 0.56(**), Maths 0.50(***), geography 0.60(***), history 0.50(**), science 0.48(*) biology 0.40(*) average achieve 0.54(**). POSCOPSD - biology 0.40(*). PSOMEMSD - English 0.47(*), health 0.40(*), biology 0.45(*), average achieve 0.42(*). TARCOPSD - biology 0.40(*). TRAMEMN - health 0.43(*)).

Second set of results(a) - Stepwise regression analysis regressing July results for each subject and composite result, firstly on GSAT and then LPAD batteries - Order of variables is criterion, battery, predictors at each step, R square, F(model) and df. ***p<0.001, **p<0.01, *p<0.05.

English GSAT GRAWV0.65, 45.98***, 1 & 26; GRAWV, GPATCOMP 0.70, 28.28***, 2 & 26 - English LPAD LNUMPRE 0.31, 11.08**, 1 & 26; LNUMPRE, POSMEMSD 0.43, 8.96**, 2 & 26, LNUMPRE - POSMEMSD-PRECOPN 0.53, 8.78**, 3 & 26; -, Maths GSAT GRAWV 0.50, 25.27***, 1 & 26 Maths LPAD - LNUMPRE 0.48, 22.63***, 1 & 26, Geog GSAT GRAWV 0.48, 23.37** 1 & 26 - Geography LPAD RPOSTOT 0.36, 14.12***, 1 & 26, Biology GSAT GRAWV 0.59, 36.57***, 1 & 26 - Biology LPAD LNUMPRE 0.31, 11.04**, 1 & 26, Science GSAT GWA 0.73, 68.05***, 1 & 26; GWA, GFIGAN, 0.78, 42.95*** 2 & 26, GWA, GFIGAN, GRNV, 0.83, 38.10***, 3.26; GWA, GFIGAN, GRNV, GVERBRE, 0.87, 36.07***, 4 & 26 - Science LPAD RPOSTRANS, 0.23, 7.41*, 1&26; POSTRANS, POSCOPN, 0.37, 6.93**, 2&26; Health GSAT GRAWV, 0.51, 26.01***, 1&26; - Health LPAD LNUMPRE, 0.25, 8.46**, 1&26; LNUMPRE, TRAMEMN, 0.40 8.14**, 2&26; History GSAT GVERBRE 0.44, 20.42***, 1&26; GVERBRE, GFIGAN, 0.56, 15.30, 2&26 - History LPAD RPOSTRAN, 0.25, 8.53**, 1&26, Total GSAT GRAWV, 06.67, 51.43***, 1&26; GRAWV, GPATCOM, 0.70, 28.47***, 2&26, Composite LPAD LNUMPRE 0.40, 16.87***, 1&26.

Second set of results (b) Summary of stepwise regression analyses of July subject and composite results on combined gsat and lpad batteries. Order of variables is: criterion, predictor(s) at each step, r square, f(model) and df.

Eng (i) GRAWV 0.65, 45.98***, 1&25(ii) GRAWV, PRECOPPL, 0.74, 33.87***, 2&24 (iii) GRAWV, PRECOPPL, GNUMPROB, 0.81, 33.09***, 3&23, Maths (i)GRAWV 0.50, 25.27***, 1&25, (ii)GRAWV LNUMPRE 0.61, 18.72***, 2 & 24, (iii)GRAWV, LNUMPRE, LNUMPOST 0.68, 16.32***, 3&23, Geography (i)GRAWV 0.48, 25.37***, 1&25(ii) GRAWV, RPOSTOT 0.56, 15.50***, 2&24, Biology (i)GRAWV 0.59, 36.57***, 1&25 (ii) GRAWV, POSCOPSD 0.66, 22.81***, 2&24 (iii) GRAWV, POSCOPSD, TRAMEMPL 0.71, 18.82***, 3&23, Science (i) GAW 0.73, 68.05***, 1&25, (ii)GWA, GFIGAN 0.78, 42.95***, 2&24, (iii)GWA, GFIGAN, GRNV 0.83, 38.10***, 3&23, Health (i)GRAWV 0.51, 26.01***, 1&25 (ii) GRAWV, TRAMEMPD 0.62, 19.32***, 2&24, History (i) GVERBRE 0.45, 20.42***, 1&25 (ii) GVERBRE, GFIGAN 0.56, 15.30***, 2&24, Composite average achievement (i)GRAWV 0.67, 51.43***, 1&25.

Third set of results: DIFFERENCE SCORES on pre and posttest mediation on LPAD

Multiple analysis of variance repeated measures, extreme scores were excluded to avoid regression to the mean. Non-adjusted (cases were excluded, hence adjusted) and adjusted scores of the LPAD battery, examining pre and posttest differences. Order of variables: subscale, mean pretest, mean posttest, non adjusted SD pretest, nonadjusted mean posttest SD, F, mean pretest, mean posttest, adjusted scores, SD pretest, SD posttest, F (df = 1&26). *p<0.05 ** p<0.01 ***p<0.001.

NUMPROG 48.65, 46.30, 19.51, 17.45, 1.06 (nothing for rest of variables), RAVTOT, 39.19, 41.26, 7.93, 6.81, 5.17* (nothing for rest of variables) RAVTRANS 33.41, 35.00, 6.51, 5.58, 4.23* (nothing for rest of variables), CFD COPN 17.56, 17.78, 0.58, 0.51, 5.20* (nothing for rest of variables), MEMN 15.30, 17.70, 2.51, 0.47, 28.95*** (nothing for rest of variables), COPYPD 13.44, 15.15, 2.68, 2.09, 10.86**, 13.80, 15.04, 2.45, 2.09, 8.34**, MEMPD 10.19, 15.00, 4.24, 1.47, 36.32***, 11.48, 14.96, 2.94, 1.43, 38.85***, COPYPL 16.15, 16.67, 1.70, 1.14, 4.63* (nothing for rest of variables), MEMPL 12.07, 16.30, 4.14, 1.59, 29.33***, 13.52, 16.39, 2.06, 1.53, 34.55***, COPSD 2.52, 1.33, 1.12, 0.55, 30.74*** (nothing for rest of variables), MEMSD 2.37, 1.22, 1.55, 0.51, 15.07*** (nothing for rest of variables).

(21) Regarding Q1 - Based on correlation and regression, the subtests of the GSAT are more highly significant than the LPAD in terms of predicting the criterion. GSAT total verbal raw score was most correlated. Even when combined with LPAD, no single LPAD sub test can substantially complement GSAT sub tests. Q2 - All results were significant bar the numerical progressions - hence mediation worked - but due to a lack of control group this cannot be stated for certain.

(22) Criterion variables: composite term results and examination marks were collected for each subject at three time-intervals. These were collated and each school subject was weighted. Predictor variables: GSAT and LPAD results, as well as length of enrolment at school. GSAT: verbal - word analogies, verbal reasoning, number problems and word pairs. Non verbal - number series, pattern completion and figure analogies. LPAD: complex figure drawing, numerical progression, Raven's progressive (pretest and posttest) matrices/set variations 1 (used for the mediation phase).

(23) Due to fatigue and time of day when numerical progressions was administered (LPAD) scores were lower. Lack of control group which would be necessary to establish whether or not mediation worked. Sample size small. Lack of fit of the scaled scores of the population tested, to the normal distribution was also evident.

(1) Henley, S.J.

(2) 1989

(3) An investigation of Feuerstein's theory of mediated learning experience with a disadvantaged community

(4) University of the Witwatersrand

(5) Unpublished Masters thesis

(6) Johannesburg

(7) The aim of the study was to investigate Feuerstein's theory of MLE as the underlying component of structural cognitive modifiability. This study does not make comparisons with static instruments - its purpose is to validate (or not) the theory behind this particular branch of dynamic assessment hence it is included here as it deals with the effectiveness of dynamic assessment as a method of valid assessment. What is being tested here is the theory behind the practice of dynamic assessment

(8) 100 were tested but analyses was only conducted on 30

(9) Originally 2 then 4 groups

(10) Mixed design - 100 subjects randomly sampled from population of 300. Random assignment to experimental and control groups. Regarding within groups comparison, 30 pupils (16 boys and 14 girls) from the experimental group were allocated to high and low Mediated Learning Experience (MLE) based on results from a questionnaire assessing their MLE is early childhood. Both the high and low MLE groups received mediation

(11) Disadvantaged pupils from a gifted child programme. 54 girls and 46 boys

(12) Range 12-19 with mean = 14.8

(13) LPAD - Organizer (A), Verbal analogies test and Set Variations I and II. Independent measures (transfer measures): Organizer (B), similarities subtests of the WISC-R and WAIS-R as well as Raven's standard progressive matrices. i.e. Pretest measures and their transfer counterparts included organizer A with an alternative form as well as organizer B; Verbal analogies transfer counterpart included similarities from the WISC-R and WAIS-R; set variations I and II transfer counterparts included the Raven's progressive matrices. A questionnaire was developed in order to determine the extent of MLE received at home, and on the basis of this the grouping of low and high MLE was based.

(14) As per Feuerstein's method of test-teach-test

(15) Solomon four group design, group administered- experimental and control groups; then half the experimental group and half the control group were tested on tasks similar to LPAD to control for practice effects. The experimental group pretest (group 1 n=24); Experimental group no pretest (group 2 n=28); Control group pretest (group 3 n=23); control group no pretest (group 4 n=25)

(16) Disadvantaged children on the gifted child programme in Soweto

(17) Mediation had already taken place in the study of Hoffenberg (1988), Henley made use of the same data and reanalysed the data based on the groupings of high and low MLE students (obtained from the questionnaire)

(18) H1: the group identified as having had high MLE will have a higher level of academic achievement than the low MLE group. H0 there will be no difference between the groups. H II: The group identified as having had high MLE will have a higher level of achievement on LPAD tasks than the low MLE group. H0: there will be no difference between the groups. H III: The group identified as having had high MLE will have a higher level of achievement on transfer measures comparable to the LPAD tasks than the low MLE group. H0: There will be no differences between groups

(19) As prescribed by Feuerstein

(20) Scores on the LPAD pre and posttest phases were compiled as follows: LPAD Index scores were converted to percentages - sum of Organizer A% + Verbal analogies% + set variations I and II and all divided by 3. The same was done for the posttest scores: Organizer B% + Similarities% + Raven's% all divided by 3. These two indices were used in comparing groups. Based on the questionnaire that was designed to ascertain MLE status, 66.67% of high MLE were high achievers and 75% of the low MLE group were low achievers. Overall 70.8% of the MLE groups contained subjects that were expected to be in each group (discriminant analysis could not be conducted due to the small sample size).

LPAD scores (pretest scores): t test conducted to determine whether there was a significant difference between achievement groups. Mean of high achievement group (n=15) = 69.5333% (SD 5.79244) and low MLE group (n=15) = 45.4667%. (SD 7.94505) $t=9.47986$, $df=28$, $p<0.05$. Therefore sig difference.

H1t test used to compare academic performance of high and low MLE groups: mean of high MLE group (n=12) = 61.3333% (SD 12.8015) low MLE group (n=12) mean = 50.667% (SD 12.1381) $t=2.09455$, $df=22$, $p<0.05$. Thus H1 confirmed.

H2 t test to compare the two MLE groups in terms of scores on the LPAD tasks t test: high MLE group mean = 58.8942% (n=12) (SD 11.7225) and low MLE group (n=12) (SD 14.0979) mean = 55.8408% $t=0.576881$, $df=22$ but $p>0.05$, not significant therefore H2 cannot be confirmed.

t test to compare means between high and low achievement groups on the LPAD tasks: high achievement group mean (n=15) (SD 10.8878) = 65.1487% and low achievement group mean (n=15) (SD 10.8267) = 52.146%. $t=3.279676$, $df=28$ and $p<0.05$. Therefore significant difference between groups.

Posttest scores: Regarding H3; t test to compare differences on posttest scores on both high and low MLE groups. High MLE group mean (n=11) (SD 8.92848) = 65.3509%, low MLE group mean (n=11) (SD 12.2489) = 56.3309%. $t=1.97366$, $df=20$ and $p<0.05$, significant therefore H3 confirmed. Same t test conducted on high and low achievement groups: mean high MLE group (n=14) (SD 10.4939) = 68.2921% mean low MLE group (n=14) (SD 9.21918) = 55.0536%. $t=3.54617$, $df=26$ and $p<0.05$

(21) Support given to hypotheses I and III but not to Hypothesis II. Feuerstein's theory of higher amounts of early MLE leading to improved cognitive development and cognitive performance is confirmed. Support was given to Feuerstein's theory that mediated learning experience is the proximal factor in cognitive development

(22) Predictor variable: MLE status and criterion variables: LPAD scores, academic achievement and posttest transfer measures

(23) Two MLE groups were not established as comparable so differences between them may be due to differences in intelligence and not MLE alone. Small sample size. Questionnaire design and false answers on items etc. There was a significant difference between high and low MLE groups on academic performance and on posttest tasks, there was also no significant difference on the LPAD tasks for these two groups.

(1)Boeyens, J.**

(2) 1989

(3) Learning potential: an empirical investigation.

(4) Human Sciences Research Council

(5) Published

(6) Pretoria

(7) To evaluate the performance of a learning potential instrument and to investigate the relationship between academic performance and learning potential. Stages of the study: stage one is a pilot study (N=183), the second stage concerns itself with the scalability of the test (not included here) and the third phase deals with the reliability of the test (N=202). Two separate sets of results are given.

(8) First study: 183. Third study: 202

(9) 2

(10) First study n=91 and n=92, the first group was administered mediation, the second group was not administered mediation. Groups were further subdivided and analysed as within-groups designs.

(11) For the first group: disadvantaged black Matric pupils undergoing education during the time of apartheid under the Department of Education and Training and for the second group older adults who wished to obtain a matriculation pass

(12) Age was not given for first study but a figure of 17 is estimated. For the second study age ranged between 17-62 and the mean was = 25.5

(13) First study: The Learning Potential Test (LP) was used - a test developed by the author for the purposes of this study. For the second study the predictor variables consisted of the Learning potential test, pretest of the LP, the Intermediate Mental Alertness test and the Number Comprehension Test. The criterion variable consisted of academic maths performance on a mathematics achievement test taken twice in the year

(14) Administration of the lesson involved reading the text to the testees, who followed in their answer books

(15) The experimental group followed a pretest, mediation and posttest programme but no mediation was administered to the control group

(16) For the first study: Black matric pupils (DET) who were classified as "disadvantaged" and for the third study the sample consisted of black teachers who wanted to obtain a matriculation

(17) This was not indicated

(18) It is predicted that learning potential scores would be indicative of the amount of improvement in academic performance that a previously disadvantaged individual currently exposed to an enriched educational environment will evidence during the course of the academic year. H1: A positive relationship exists between improvement in academic competence and learning potential. H2: Learning potential scores correlate higher with improvement in mathematical competence as opposed to ability scores. Thirdly the relationship between learning potential and traditional ability test scores are investigated, fourthly the relationship between ability, learning potential and school marks is investigated and lastly, the investigation into low and high Learning potential groups' school results and how they will differ when correlated with static tests is undertaken.

(19) The administration of the lesson involved reading the text to the testees, who followed in their answer books

(20) The test consisted of the completion of a letter series by following patterns etc. Two systems of scoring were used, a micro and item scoring. The micro scoring consisted of assigning a score of one to each letter correctly supplied thus 30 items resulted in a score of 90 as there were three alternatives. The second scoring system takes each series as either being correct or wrong irrespective of whether or not some of the letters are correct. Therefore a total score of 30 was obtained. The following is presented: Exp n=91 and control n= 92 each group has scores representing micro-item and item scores, with means and standard deviations for each. Each group also has a pretest, posttest and differences scores in this order. Exp group pretest(n=91): Micro: mean 38.78, sd 11.53, item: mean 12.58, sd 4.34. Exp group posttest: Micro:

mean 46.90, sd 13.60, item: mean 15.66, sd 5.02. Exp group difference score: Micro: mean 8.12, sd 10.48, item: mean 3.08, sd 4.02. Control group (n=92) pretest: Micro: mean 38.70, sd 13.59, item: mean 12.53, sd 5.02. Control group (n=92) posttest: Micro: mean 47.03, sd 17.76, item: mean 15.42, sd 6.41. Control group (n=92) difference score: Micro: mean 8.34, sd 8.39, Item: mean 2.89, sd 2.99. No significant difference between ability groups and no significant difference between mean improvement. However, experimental group attempted fewer items on posttest than control group even though they did not differ significantly. Experimental group attempted 60.47 items with sd 15.44 on pretest and 61.74 with sd 13.32 on posttest while control group attempted 64.27 with sd 15.31 on pretest and mean 73.10 and sd 12.78 on posttest.

For further analyses experimental group was divided into (i) group who showed no sign that lesson (mediation) had an impact (ii) group who made use of lesson (mediation) and (iii) a group who too rigidly applied the lesson (mediation) in their tasks. The following is available for means and standard deviations of the DIFFERENCE scores in the groups : experimental group (n=92). Micro: mean 8.34, sd 8.39 and item: mean 2.89, sd 2.99. Group (i) (n=26) Micro: mean 11.00, sd 9.12 and item: mean 3.85, sd 3.65. Group (ii) (n=44) Micro: mean 11.82, sd 7.62 and item: mean 4.45, sd 3.36 Group (iii) (n=21) Micro: mean -3.19, sd 9.52 and item: mean -0.76, sd 3.37. t test indicated significant difference between group (ii) and control group when scored out of 90 $p < 0.02$ and when scored out of 30 $p < 0.01$. Mediation seems to have had some impact.

Third study's results - H1 Correlations between LP and maths scores (n=40) students who took maths: (** $p < 0.01$ and * $p < 0.05$). maths score 1 with LP pretest 0.10 and with posttest 0.11, and with LP (post-pre LP score) 0.03; maths score 2 with LP pretest 0.42** and posttest 0.47** and with LP 0.14 and difference score in maths (between score 1 and 2) and LP pretest 0.36* and posttest 0.40** and LP 0.12. Investigation for the group (ii) (n=29) who made use of the mediation offered - their maths scores 1 and 2 are correlated with pre and posttest LP: (** $p < 0.01$ and * $p < 0.05$). maths score 1 with LP pretest 0.02 and with posttest -0.06 and LP -0.12, maths score 2 with LP pretest 0.18 and posttest 0.35 and with LP 0.32 and difference score in maths (between score 1 and 2) and pretest 0.17, with posttest 0.44** and with LP 0.47**. Therefore there is a significant correlation between improvement in maths achievement and learning potential only for the group who used the mediation strategies though. Support for H1.

Regarding H2: Which scores correlate higher with improvement in maths achievement? LP or traditional tests? Propensity to improve should be correlated higher with LP than with current ability. traditional scores are correlated for the following maths improvement score (* $p < 0.05$): Mental Alertness: all maths pupils (n=40) = 0.06 and those who applied mediation properly (n=29) = -0.15; Number comprehension: all maths pupils (n=40) = 0.15 and those who applied mediation properly (n=29) = -0.01; Learning potential pretest (recall that LP pretest is traditional form of testing): all maths pupils (n=40) = 0.36** and those who applied mediation properly (n=29) = 0.17. Due to the two sample sizes, the correlation between improvement in maths achievement (n=29) (0.47**) and number comprehension with improvement in maths (n=29) (-0.01) is taken and compared to see if 0.47 is significantly higher than -0.01. According to the author it is. Therefore H2 supported.

As current ability and LP are not the same construct, the two should not correlate and the following correlations for learning potential are presented: Entire sample (n=202): number comprehension 0.13, mental alertness 0.14, learning pretest 0.05. Group (i) (n=63): number comprehension 0.06, mental alertness 0.11, learning pretest 0.06. Group (ii) (n=127): number comprehension 0.13, mental alertness 0.15, learning pretest -0.00. Group (iii) (n=12): number comprehension 0.32, mental alertness 0.01, learning pretest 0.08. No significant correlations between any variable, therefore H3 supported (two concepts of ability and potential are independent).

Regarding H4, LP will add significantly to prediction of academic success over and above traditional tests. Stepwise regression conducted with March test results as dependent variable and Mental Alertness battery (MA), Number comprehension (NC) and LP as independent. Following are correlations for these variables (N=165) (** $p < 0.01$): LP-MA=0.12; LP-NC=0.10; LP-March=0.23**; MA-NC=0.62**; MA-March=0.53**; NC-March=0.44**. Stepwise regression (N=165) for March as dependent variable: order of variables - test, partial r, multiple r, F and p.MA - 0.28, 0.28, 64.95, 0.0001. LP- 0.03, 0.31, 6.35, 0.01. NC- 0.01, 0.33, 4.65, 0.03. Better prediction can be obtained when using both MA and LP. Similar correlations carried out for the group (i) who integrated their mediation (N=109) : (** $p < 0.01$): LP-MA=0.17; LP-NC=0.12; LP-March=0.29**; MA-NC=0.54**; MA-March=0.47**; NC-March=0.38**. Stepwise regression also conducted for this group (N=109) for March as dependent variable: order of variables - test, partial r, multiple r, F and p.MA - 0.22, 0.22, 30.18, 0.0001. LP- 0.04, 0.26, 6.44, 0.01. NC- 0.02, 0.28, 3.00, 0.09. For this group LP improved prediction but NC did not add much. Thus LP added more than static tests for this group.

Regarding H5 - low and high LP groups will differ when correlated with static tests - i.e. low LP group will have significantly higher correlations with static tests than high LP group. Correlations are presented, first figure is correlation with high LP group followed by n and second figure is correlation with low LP group followed by n. (* $p < 0.05$ ** $p < 0.01$): Eng (March) 0.46** (88), 0.54** (77), Eng (June) 0.46** (88), 0.63** (77); Afr (March) 0.52** (88) 0.39** (75), Afr (June) 0.45** (88) 0.53** (73); Biology (March) 0.50** (64) 0.65** (54); Maths (March) 0.30 (27) 0.68** (14); Business (March) 0.36* (41) 0.47** (43), Business (June) 0.25 (41) 0.42** (43); Average (March) 0.48** (88) 0.56** (77). Although there were differences between the correlations for each subject, after Fisher's Z transformations, no significant differences between correlations for the two groups were found. The same correlations were run but for group (i) again - those who integrated the mediation: first figure is correlation with high LP group followed by n and second figure is correlation with low LP group followed by n. (* $p < 0.05$ ** $p < 0.01$): Eng (March) 0.36** (66), 0.53** (43), Eng (June) 0.36** (66), 0.58** (43); Afr (March) 0.38** (66) 0.40** (41), Afr (June) 0.36** (66) 0.41** (39); Biology (March) 0.33** (45) 0.69** (31); Maths (March) 0.10 (21) 0.85** (10); Business (March) 0.37* (33) 0.52** (22), Business (June) 0.23 (33) 0.15 (22); Average (March) 0.33** (66) 0.57** (43). Once again, differences in correlations are not significant, but they are consistently higher for the low LP group than for the high LP group.

(21) Learning potential is an effective predictor of the potential to benefit from instruction. Improvement between pre and post testing suggested that there was a higher correlation between those pupils receiving mediation as opposed to those who did not. This supports the first hypothesis. Also, no relation exists between learning potential and static tests and although not significant, correlations between high and low learning potential groups regarding static tests evidenced higher correlations for the lower potential group, thus indicating that ability and potential are different constructs; also there were no significant correlations between static measures and learning potential measures, emphasising the difference in constructs. There were quantitative and qualitative differences between the three groups identified in terms of learning potential i.e. those who integrated the mediation, those who did not and those who applied the mediation too rigidly. Those who applied mediation properly benefited the most, their correlations between static scores and improvement in mathematical competence was lower than the correlation between the number comprehension scores and improvement in mathematical competence. This group thus had different predictive value in comparison to the other groups. Therefore, learning potential is a useful predictor of the outcome of exposure to an enriched educational environment, but its predictive validity is limited to a proportion of typical students. According to results from a multiple regression analysis, it was evidenced that learning potential in combination with other static assessment yielded greater predictability than when either of two static tests were used on their own (the learning potential results contributed significantly to the model). The lower learning potential group consistently averaged higher correlations between learning potential and static assessments contrary to what was expected, but Boeyens points out that this needed further investigation.

(22) Criterion variable: march and June school results; predictor variable: mental alertness battery, number comprehensions and LP test

(23) The qualitative differences between the three groups could receive more attention regarding non cognitive aspects.

(1) Murray, D.**

(2) 1988

(3) Effectiveness of Feuerstein's learning potential assessment device in a South African context

(4) University of the Witwatersrand

(5) Unpublished. Masters dissertation

(6) Johannesburg

(7) To test the effectiveness of a dynamic assessment approach (LPAD) among groups of socio-politically and educationally disadvantaged Indian and Coloured adolescents

(8) 108

(9) 4

(10) Mixed design, investigating both between and within-group aspects

(11) Disadvantaged Indian and Coloured high school pupils (from two different schools)

(12) Age ranged between 13-15. n=60 for Coloured sample and n=60 for Indian sample

(13) Two sets of measures were used, namely several tests from the Learning Potential Assessment Device (LPAD) and a set of independent measures to assess transfer of mediation. Dependent measures include the LPAD - Set variations I and II; Complex figure drawing (CFD) and the Lahi. Independent measures (transfer tests) include: Ravens' standard progressive matrices (RSPM) (parallel to set variations); equivalent complex figure (ECF) (parallel to CFD); coding subtest from the WISC-R (parallel to Lahi); and similarities subtest of the WISC-R (parallel to comparisons). A biographical questionnaire was developed to assess and compare differences between sub groups

(14) Mediation on all LPAD tasks except for set variations were administered by students, set variations were administered by a psychologist. Mediation involved the principles as described by Feuerstein. The control group were exposed to training, but no mediation other than having the tasks explained to them.

(15) The experimental group received a pretest, mediation and posttest programme but no mediation was administered to the control group. n=30 chosen from the top achievement and n=30 chosen from the lower achievement scale for both schools. Thus four groups of 15 each - 15 high achievers for Coloured, 15 low achievers for Coloured were arranged as well as 15 high achievers and 15 low achievers for Indians. 10% attrition rate resulted in a total sample of N=108. Due to random assignment of pupils to experimental and control groups, no socio-economic differences were found for either experimental and control groups nor for high and low academic groups. Thirty mediators were each randomly assigned two experimental and two control group subjects who were individually mediated on the LPAD

(16) Disadvantaged Indian and Coloured adolescents

(17) Independent measures were administered one week before and two weeks after the LPAD. LPAD involved six hours of interaction (four of the six hours was individual mediation and the remaining two consisted of group mediation (n=15))

(18) H1 - mediation on the various tasks of the LPAD was expected to be effective in improving the performance of subjects on these tasks relative to subjects who did not receive such mediation. H2 - effects of mediation provided by the LPAD were expected to generalise to performance on other comparable measures (transfer effect). H3 - it is believed that the Coloured sample was "more disadvantaged" and that mediation provided by LPAD would be more effective in improving performance for them compared to the Indian sample. H4 - it is expected that mediation would be effective in improving the performance of both high and low academic status students and that higher achieving students who were nevertheless deprived would show higher levels of potential

(19) As prescribed by Feuerstein

(20) Analysis of covariance was used as more than one dependent variable was included resulting in multiple analysis of variance and in order to test for main and interaction effects. Only significant results follow. Order of variables: variable; CFD (memory organisation), (memory accuracy), (production accuracy); Set variations I, set variations II (* p<0.05 ** p<.01, *** p<0.001) DF = 1 and 107. Race group (Indian and Coloured) 0.01, 0.01, 0.67, 1.32 6.85*. Group (Exp/control) 9.43***, 0.76, 11.10**, 1.48, 4.56*. Academic status (high/low) 6.20*, 4.72*, 0.13, 28.42***, 27.16***. Race group X exp/con group 0.90, 0.30, 0.38, 0.14, 0.43. Race group X academic status 0.22, 0.07, 0.07, 0.06, 0.20. Exp/Con group X academic status 0.12, 0.79, 1.67, 0.43, 0.09. Race group X Exp/Con group X academic status 2.09, 0.49, 1.45, 0.06, 2.09. Effects of mediation on transfer tasks, i.e. independent variables' means and standard deviations given for Exp/Con groups; high and low academic achievement groups and race group. Five measures used - RSPM, ECF (Accuracy (ACC); ECF Organisation (ORG), coding (COD) and similarities (SIMS). Means are given and in brackets standard deviations. First mean is pretest and second mean is posttest. First group is Exp/Control group: Exp group: RSPM 46 (8) and 49 (7) and control group 47 (6) and 49 (6). ECF (ACC) Exp group 30 (5) and 32 (4) control group 30 (5) 31 (4) ECF (ORG) Exp group 4 (1), 6 (1) control group 4 (1) 5 (1). COD exp group 51 (12) and 63 (11) control group: 53 (13) and 62 (13). SIMS Exp group 20 (5) and 23 (5) and control group 20 (6) and 22 (5).

Second group is high and low achievement status (Murray's heading on pg 25 do seem to be slightly confusing) RSPM high group: 50 (5) and 53 (4) and low group 43 (8) and 45 (6). ECF (ACC) high group 32 (4) and 32 (4) low group 29 (5) 31 (4) ECF (ORG) high group 6 (1), 5 (1) low group 4 (1) 5 (1). COD high group 53 (11) and 67 (11) low group 52 (13) and 58 (11). SIMS high group 22 (4) and 25 (4) and low group 17 (5) and 20 (4).

Third group is Indian and Coloured. RSPM Indian: 49 (6) and 51 (7) and Coloured 44 (8) and 48 (6). ECF (ACC) Indian 30 (4) and 31 (4) Coloured 30 (5) 32 (4) ECF (ORG) Indian 4 (1), 4 (1) Coloured 4 (1) 5 (1). COD Indian 58 (12) and 65 (10) Coloured: 48 (10) and 61 (13). SIMS Indian 20 (5) and 24 (4) and Coloured 20 (6) and 22 (5).

There were no significant difference between experimental and control groups even though there was a tendency for the experimental group to achieve higher scores on pretest scores. T tests for independent groups performed for high and low status groups were all significant except CODING. For RSPM t(df=106=5.6, p<.0001) for ECF (accuracy) t(df 106 = 3.7, p<0.001) for ECF (organisation) t = (df 106= 3.6, p<0.001) and for Similarities t (df 106=5.2, p<0.0001). Therefore on four of the five tests there was a significant difference between high and low academic achievement status groups both prior to and after intervention. For Indian and Coloured groups pretest differences were found to be significant on three of the five measure, ECF (organisation) t (df 106=2.0, p<0.05) RSPM t (df 106=4.1, p<0.0001) and coding t (df 106=4.8 p<0.0001). In other words, there were no significant differences to begin with for the experimental and control groups, but there were significant differences between race group and academic achievement groups.

To investigate the effects of the intervention on transfer (posttest), analysis of covariance was run. The following are results from an analysis of covariance on posttest tasks for race group, exp/con group and academic achievement group. Order of variables: variable stated, RSPM, ECF (ACC), ECF (ORG), Coding, Similarities. df = 1.07 *p<0.05 **p<0.01 ***p<0.001. Race group 7.02*, 0.00, 6.90*, 1.59, 4.82*. Exp/con group 0.67, 0.01, 0.00, 0.48, 1.07. Academic achievement 14.66***, 5.20*, 0.00, 16.92***, 9.98**. Race X exp/con group 1.88, 0.71, 0.02, 10.43**, 0.05. Race X academic achievement 5.14, 0.03, 0.77, 0.37, 8.33. Exp/con group X academic achievement 0.52, 2.16, 2.10, 0.17, 4.04*. Race X academic achievement X exp/control groups 1.12, 3.29, 0.32, 2.35, 0.81. In no case was the result for the total experimental group significantly better than the control group. However, the Indian experimental group did significantly better than the Indian control group (mean for exp 64.3 and for latter 56.9); F (1.53)= 10.43 p<0.01. Thus for Indians mediation seems to have worked for coding. Mediation had no effect for the total sample on any of the transfer effects.



(21) Academic performance was the most significant variable differentiating among subjects. Bright students benefited more from mediation than their non-coping counterparts regardless of exp/con group status. Mediation was not generally effective but there was an interaction effect with other variables. H1 not supported through significant effects but trends are evident. H2 partially supported by interaction effects for transfer tasks (posttests). H3 only partially supported as both Indian and Coloured groups differed in terms of how they improved on various tasks. H4 accepted.

(22) Independent variables: academic performance, cultural group, and exp/ control group. Dependent variable: LPAD and transfer tasks

(23) Only two hours were allocated for mediation and the use of untrained LPAD testers

(1) De Villiers, A.B.**

(2) 1999

(3) Disadvantaged students' academic performance: analysing the zone of proximal development

(4) University of Cape Town

(5) Unpublished doctoral thesis

(6) Cape Town

(7) The aim of the study is to investigate the practical application of Vygotsky's construct of the zone of proximal development to the selection of disadvantaged students in higher education and to determine alternative predictors of academic performance other than the traditional matriculation examination results used in South Africa

(8) 400

(9) Two groups but further sub-divided for various analysis purposes; including the first study which comprised the use of two dynamic assessment batteries and the effect of the mediated lesson; the second study looked at correlations between the static and dynamic tests; while the third study sought to find predictors of academic performance

(10) Mixed design, incorporating both between and within study populations

(11) Disadvantaged first year Technikon students

(12) Not stated but an approximation 18 is estimated

(13) The Conceptual reasoning test (CRT) which was regarded as a cognitive dynamic assessment instrument (group administered), the Learning potential test (LPT) also considered a cognitive dynamic assessment instrument (group administered), the Reading comprehension test (RCT) used as a static cognitive assessment instrument, the Study process questionnaire (SPQ) used as a static non-cognitive assessment instrument, the Motivated strategies for learning questionnaire (MSLQ) also used as a static non-cognitive assessment instrument, the Mental alertness test (MA) used as a static cognitive assessment instrument, and Electrical aptitude test (EAT) also used as a static cognitive assessment instrument all regarded as static assessment instruments. In sum two tests used as dynamic assessments and seen tests used as static instruments; five used as cognitive tools and two used as non-cognitive tools

(14) Both the CRT and LPT have a mediated lesson incorporated into the test and include written material and verbal interaction

(15) Pretest-standardised training-posttest. Only the posttest score of the LPT used and one score for the CRT is used. An investigation was done in the form of five different studies of first year Technikon students. Study no 1 focused on the effectiveness of the mediated lessons that form part of the two dynamic tests using a Solomon four group design. Solomon four group design for both (a) and (b) studies. LPT1 served as the pretest and LPT2 served as the posttest. Study no 2 compared the predictive validity of past academic achievement, conventional tests, noncognitive variables and two dynamic tests. Study no 3 analysed students' response to a period of mediation. Study no 4 compared different groups of students according to the following classification: schooling, gender, language, type of course and assessment and level of course to investigate whether any of the variables would have a moderator effect. Study no 5 differentiated between more and less successful students.

(16) Study 1 (a) Black students who had not yet matriculated (b) Black students in their first year education in commerce course (c) Black and Coloured first year electrical engineering students (d) Black and Coloured first year information technology students. Study 2 (a) Black and Coloured first year business students (b) first year electrical engineering students Study 3-5 used the same sample and further sub-divided for analyses purposes

(17) Study 1: one day between pre and posttest with mediation in between. The rest of the studies included statistical analyses of results

(18) H1 - there will be a significant difference between the scores of those students who received the mediated lesson as part of dynamic assessment as opposed to those who did not receive the lesson in both dynamic assessment tests. H2 - There will be a significant enhancement in the prediction of students' academic performance with the use of dynamic assessment as opposed to the use of static tests and previous academic achievement (matriculation marks). H3 - There will be significant difference in patterns of correlation with academic performance as well as in learning and cognitive profiles between the following groups: (i) "modifiability" i.e. a moderator effect will be evident on tests predicting academic success; static marks will be a better predictor for less modifiable group and dynamic assessment will be a better predictor for the more modifiable group (ii) "schooling" will also evidence a moderating factor; dynamic assessment will be a better predictor of matriculation marks for non-DET (non-Department of Education and Training schools) (iii) "Socio-economic status" (SES), gender and language will not have a moderator effect and there will be no significant difference between high and low SES correlations nor male/female groupings nor language (iv) "year of study" and area of discipline - third year students will score higher than first year students on some factors and engineering students will have higher scores than business students

(19) Written material and verbal feedback

(20) Results analysed according to the number of sub-studies (numbers 1-5 as indicated in point 16 above). For the purposes of this study (in terms of effectiveness of dynamic assessment as opposed to the effectiveness of static assessment, statistical results pertaining to other aspects of De Villiers' study will not be documented in this section)

Study no 1: effectiveness of mediated lessons for (a) LPT and (b) LPT.

(a) LPT four groups, (i) n=19 engineering (exp group); (ii) n=13 science (con); (iii) n=16 business(con), (iv) n=14 business and building(con). Means and sd's for each group - group (i) (n=19): pretest 12.15 (3.84), posttest 14.57 (4.76), difference 2.47 (3.84). group (ii) (n=13): pretest 11.31 (3.98), posttest 14.53 (6.09), difference 3.25 (2.65). group (iii) (n=16): posttest 10.50 (3.24), group (iv) (n=14) posttest 8.21 (4.75). ANOVA conducted for pretested groups, no significant difference ($F=0.32$, $p>0.05$ $df=1&30$). groups (i) and (iii) who received mediation scored higher means than the other two but were not significant. Two way ANOVA conducted for the effect of mediation on posttest. Mediated lesson had no effect ($F=0.88$, $P>0.05$, $df=2;58$). No interaction between treatment and pretesting ($F=0.85$, $P<0.05$, $DF=2&58$). Groups (i) and (ii) compared to (iii) and (iv) i.e. pretested and not pretested groups - pretested groups (n=32, mean 14.56, sd 5.248) and without pretests (n=30, mean 9.43, sd 4.116). Total group (n=62, mean 12.08 sd 5.36). The pretest did sensitise the pretest group ($F=17.54$, $p<0.05$, $df=2&58$). For those students in the exp group (minus those who rigidly applied mediation tactics) (n=14): pretest mean 12.21, sd 4.14 posttest mean 16.36 sd 3.73, difference score mean 4.14 sd 2.93. Significant difference between pretest and posttest for the exp group ($t=-2.78$, $p=0.001$, $df=26$).

(b) LPT same design as with (a) four groups (all commerce students), Means and sd's for each group - group (i) (n=15): pretest 5.53 (5.19), posttest 8.73 (5.77), difference 3.2 (3.05). group (ii) (n=16): pretest 3.5 (4.25), posttest 4.43 (5.08), difference 0.94 (2.05). group (iii) (n=16): posttest 7.52 (3.65), group (iv) (n=14) posttest 6.13 (6.39). Two pretested groups compared in ANOVA: no significant difference ($F=1.43$, $p>0.05$, $df=1&29$). Two way ANOVA: effect of mediated lesson on posttest LPT scores. There was a significant effect ($F=4.51$, $p<0.05$, $df=1&59$) on the group who received the lesson. No interaction between lesson and pretesting ($F=1.19$ $p>0.05$, $df=1&59$), pretesting had no significant effect ($F=0.03$, $p>0.05$, $df=1&59$). Difference score of group (i) significantly higher than group (ii) ($t=2.44$, $p=0.02$, $df=29$). Investigating the

retest effect (similar to Cohen's D) between (a) and (b): standardised mean change between pre and posttest for control group who did not receive mediation $z(g) = \text{mean of posttest} - \text{mean of pretest} / \text{sd pretest} = 0.81$ for (a) and 0.22 for (b), hence gains in the first group are due to retest and gains in second group are not due to retest effect. Remaining with study 1 but for (c) CRT and (d) CRT:

(c) CRT two groups $n=79$ electrical engineering (exp group); $n=23$ electrical engineering (control group). Means and sd's: exp group $n=79$: 1826 (5.06) and control group $n=23$: 13.47 (3.64). $t=4.22$, $p<0.001$ $df=100$. Recall that in the CRT test there is no pretest, just mediation and test. There was a significant difference.

(d) CRT: information technology exp group ($n=34$) and information technology control group ($n=35$). means and sd's: exp ($n=34$), 18.38 (6.61) and control group ($n=35$), 15.91 (5.24). $T=1.4$, $P=0.09$, $DF=68$. No significant difference. For study (d) another analysis was conducted, the sample was divided into DET and non-DET (i.e. "disadvantaged" and "advantaged"). Non-DET (exp group) $n=24$ (mean 20.83 and sd 6.01); control group $n=32$ (mean 16.56 and sd 4.96); DET exp group $n=10$ (mean 12.5 and sd 3.68) and control group $n=3$ (mean 9 and sd 2.64). $t=2.91$, $p<0.01$ and $df=54$ for non DET exp and con groups - there is a significant difference between non DET exp and con groups but not for DET groups.

Study no 2: a comparison between past achievement, static tests and dynamic assessment as predictors.

(a) $n=33$ business management students. Pearson correlations conducted for MA (static), RCT (static), CRT (dynamic) and SR (Swedish rating) with academic results (YEAR, EXAM, FINAL, CREDIT). Order of variables: academic criterion correlated with MA, CRT, SR and RCT, first the correlation, then N, then p value: YEAR (MA) 0.368* $N=33$, $p=0.035$ *; YEAR (CRT) 0.372* $N=33$, $p=0.033$ *; YEAR (SR) 0.202 $N=33$, $p=0.260$; YEAR (RCT) 0.272 $N=33$, $p=0.125$; EXAM (MA) 0.279 $N=33$, $p=0.116$; EXAM (CRT) 0.303 $N=33$, $p=0.086$; EXAM (SR) 0.288 $N=33$, $p=0.104$; EXAM (RCT) 0.090 $N=33$, $p=0.617$; FINAL (MA) 0.287, $N=33$; $p=0.106$; FINAL (CRT) 0.324, $N=33$; $p=0.066$; FINAL (SR) 0.260, $N=33$; $p=0.144$; FINAL (RCT) 0.117, $N=33$; $p=0.514$; CREDIT (MA) 0.339, $N=33$; $p=0.054$; CREDIT (CRT) 0.349*, $N=33$; $p=0.046$ *; CREDIT (SR) 0.388*, $N=33$; $p=0.026$ *; CREDIT (RCT) 0.192, $N=33$; $p=0.286$. CRT and SR together proved better predictors.

To see which variable predicted better a stepwise multiple regression was conducted. For YEAR mark: $R=0.371$, $R \text{ square}=0.138$, $r \text{ square} a=0.110$, $F=4.976$, $p=0.003$, $n=33$ variable selected CRT beta 0.371, std error of beta 0.166, $t=2.23$ $p=0.033$. For CREDITS: $R=0.387$, $R \text{ square}=0.151$, $r \text{ square} a=0.123$, $F=5.493$, $p=0.025$, $n=33$ variable selected SR beta 0.387, std error of beta 0.165, $t=2.343$ $p=0.025$. Best predictor for YEAR is CRT and best predictor for CREDITS is SR. When CRT is added to CREDITS $r \text{ squared}$ increases by 0.08 which represents a low to medium increase (based on Cohen's D). Therefore CRT and SR contributed most to prediction explaining 14% and 15% of the variance respectively, hence other static tests did not prove as robust.

(b) $n=41$ electrical engineering students, who wrote the CRT, LPT1 and LPT2 (pre and post), RCT, EAT and SR was available. Pearson correlations: YEAR (SR) 0.424* $n=41$ $p=0.006$ *; YEAR (CRT) 0.050 $n=41$ $p=0.757$; YEAR (LPT1) 0.234 $n=41$ $p=0.140$; YEAR (LPT2) 0.189 $n=41$ $p=0.238$; YEAR (RCT) 0.234 $n=41$ $p=0.142$; YEAR (EAT) 0.091 $n=41$ $p=0.573$; FINAL (SR) 0.421* $n=41$ $p=0.006$ *; FINAL (CRT) 0.047 $n=41$ $p=0.771$; FINAL (LPT1) 0.201 $n=41$ $p=0.207$; FINAL (LPT2) 0.172 $n=41$ $p=0.282$; FINAL (RCT) 0.224 $n=41$ $p=0.160$; FINAL (EAT) 0.074 $n=41$ $p=0.647$; CREDIT (SR) 0.311* $n=41$ $p=0.047$ *; CREDIT (CRT) 0.030 $n=41$ $p=0.854$; CREDIT (LPT1) 0.210 $n=41$ $p=0.189$; CREDIT (LPT2) 0.291 $n=41$ $p=0.058$; CREDIT (RCT) 0.291 $n=41$ $p=0.065$; CREDIT (EAT) 0.048 $n=41$ $p=0.767$. Only SR was significant. A stepwise regression for this group: For CREDITS: $R=0.311$, $r \text{ square}=0.096$, $r \text{ squared} a=0.074$, $F=4.188$, $p=0.47$, $n=41$. beta 0.311, std error of beta 0.152, $t=2.046$. when LPT2 is added to the model, it increases by 0.07 which is a low to moderate contribution. For this sample, SR did better than either static or dynamic assessments.

A sample of electrical engineering and information technology students were divided into those scoring less than 36 on the SR score and those scoring greater than 36 on the SR score.

For electrical engineering group: <36 (only correlations shown for static and dynamic tests). YEAR (SR) -0.315 $n=18$ $p=0.202$; YEAR (CRT) 0.440 $n=18$ $p=0.067$; YEAR (LPT1) 0.543* $n=18$ $p=0.020$ *; YEAR (LPT2) 0.552* $n=18$ $p=0.017$ *; YEAR (RCT) 0.290 $n=18$ $p=0.243$; YEAR (EAT) 0.396 $n=18$ $p=0.103$; FINAL (SR) -0.327 $n=18$ $p=0.185$; FINAL (CRT) 0.424 $n=18$ $p=0.079$; FINAL (LPT1) 0.498* $n=18$ $p=0.035$ *; FINAL (LPT2) 0.530* $n=18$ $p=0.023$ *; FINAL (RCT) 0.241 $n=18$ $p=0.335$; FINAL (EAT) 0.371 $n=18$ $p=0.129$; CREDIT (SR) -0.141 $n=18$ $p=0.576$; CREDIT (CRT) 0.300 $n=18$ $p=0.226$; CREDIT (LPT1) 0.435 $n=18$ $p=0.071$; CREDIT (LPT2) 0.515* $n=18$ $p=0.028$ *; CREDIT (RCT) 0.215 $n=18$ $p=0.391$; CREDIT (EAT) 0.187 $n=18$ $p=0.456$.

Electrical engineering group >36 : YEAR (SR) 0.318 $n=9$ $p=0.183$; YEAR (CRT) -0.541* $n=19$ $p=0.017$ *; YEAR (LPT1) -0.175 $n=19$ $p=0.472$; YEAR (LPT2) -0.152 $n=19$ $p=0.532$; YEAR (RCT) 0.036 $n=19$ $p=0.883$; YEAR (EAT) -0.074 $n=19$ $p=0.762$; FINAL (SR) 0.293 $n=19$ $p=0.222$; FINAL (CRT) -0.534* $n=19$ $p=0.018$ *; FINAL (LPT1) -0.223 $n=19$ $p=0.358$ *; FINAL (LPT2) -0.183 $n=19$ $p=0.451$; FINAL (RCT) 0.030 $n=19$ $p=0.900$; FINAL (EAT) -0.088 $n=19$ $p=0.720$; CREDIT (SR) 0.286 $n=9$ $p=0.235$; CREDIT (CRT) -0.498* $n=19$ $p=0.000$ *; CREDIT (LPT1) -0.252 $n=19$ $p=0.296$; CREDIT (LPT2) -0.091 $n=19$ $p=0.710$; CREDIT (RCT) 0.107 $n=19$ $p=0.661$; CREDIT (EAT) -0.041 $n=19$ $p=0.866$. Hence, for the lower scoring group, LPT correlated better and vice versa for the higher scoring group. Significant difference between the two groups for CRT ($p=0.007$), LPT1 ($p=0.040$) and LPT2 ($p=0.039$) using Fisher's z transformation.

Information technology sample divided into SR scores <27 and >27 .

Information technology <27 group: YEAR (SR) 0.357 $n=21$ $p=0.111$; YEAR (CRT) -0.033, $n=25$, $p=0.873$; YEAR (LPT1) -0.304, $N=25$, $p=0.139$; YEAR (LPT2) -0.181, $N=25$, $P=0.385$; YEAR (RCT) 0.223, $n=25$, $p=0.283$. EXAM (SR) 0.357, $N=25$, $P=0.076$; EXAM (CRT) -0.186, $N=25$, $P=0.371$; EXAM (LPT1) -0.182, $N=25$, $P=0.384$; EXAM (LPT2) -0.074, $N=25$, $P=0.725$; EXAM (RCT) 0.166, $N=25$, $P=0.426$. FINAL (SR) 0.415, $N=21$, $P=0.061$; FINAL (CRT) -0.189, $N=25$, $P=0.365$; FINAL (LPT1) -0.215, $N=25$, $P=0.301$; FINAL (LPT2) -0.102, $N=25$, $P=0.626$; FINAL (RCT) 0.174, $N=25$, $P=0.404$; CREDIT (SR) 0.333, $N=21$, $P=0.139$; CREDIT (CRT) -0.362, $N=25$, $P=0.075$; CREDIT (LPT1) -0.198, $N=25$, $P=0.342$; CREDIT (LPT2) -0.183, $N=25$, $P=0.342$; CREDIT (RCT) 0.313, $N=25$, $P=0.127$.

Information technology >27 group: YEAR (SR) 0.605* $n=25$ $p=0.001$ *; YEAR (CRT) 0.045, $n=25$, $p=0.829$; YEAR (LPT1) 0.108, $N=25$, $p=0.607$; YEAR (LPT2) 0.258, $N=25$, $P=0.213$; YEAR (RCT) -0.056, $n=25$, $p=0.790$. EXAM (SR) 0.560*, $N=25$, $P=0.004$ *; EXAM (CRT) -0.101, $N=25$, $P=0.629$; EXAM (LPT1) 0.089, $N=25$, $P=0.671$; EXAM (LPT2) 0.298, $N=25$, $P=0.147$; EXAM (RCT) -0.009, $N=25$, $P=0.965$. FINAL (SR) 0.580*, $N=25$, $P=0.002$ *; FINAL (CRT) 0.075, $N=25$, $P=0.721$; FINAL (LPT1) 0.086, $N=25$, $P=0.681$; FINAL (LPT2) 0.291, $N=25$, $P=0.157$; FINAL (RCT) -0.027, $N=25$, $P=0.897$; CREDIT (SR) 0.542*, $N=25$, $P=0.005$ *; CREDIT (CRT) 0.001, $N=25$, $P=0.995$; CREDIT (LPT1) -0.120, $N=25$, $P=0.566$; CREDIT (LPT2) 0.267, $N=25$, $P=0.196$; CREDIT (RCT) -0.103, $N=25$, $P=0.621$. Matriculation marks still proved to be better predictors for the higher scoring groups than dynamic assessment instruments. Neither static nor dynamic assessments were good predictors for the lower scoring group.

LPT1 and LPT2 scores for both electrical engineering and information technology students:

Information Technology (n=55) LPT1 (mean 16.56 sd=4.84) LPT2 (mean 20.98 sd 4.15) Difference score (mean 4.42 sd 2.97).
 Electrical engineering (n=41) LPT1 (mean 17.54, sd 4.6) LPT2 (mean 21.41 sd 4.13) Difference score (mean 3.88 sd 2.92). T test for independent samples: $t = -5.14$, $df=108$, $p < 0.0001$ for information technology and for electrical engineering: $t = -4.02$, $df 80$, $p < 0.0001$. therefore significant difference for both groups between pre and post test therefore mediation had an effect.

Study no 3: for both information technology and electrical engineering groups: samples were divided into those who were more modifiable (difference score >4) and less modifiable (difference score <4).

Information technology students more modifiable: YEAR (CRT) 0.167, $n=23$, $p=0.446$; YEAR (LPT1) 0.151, $N=23$, $P=0.491$; YEAR (LPT2) 0.157, $N=23$, $P=0.473$; YEAR (SR) 0.473*, $N=22$, $P=0.026$ *; YEAR (RCT) 0.213, $N=23$, $P=0.328$; EXAM (CRT) 0.091, $=23$, $P=0.667$; EXAM (LPT1) 0.105, $N=23$, $P=0.631$; EXAM (LPT2) 0.172, $N=23$, $P=0.432$, EXAM (SR) 0.430*, $N=22$, $P=0.045$ *; EXAM (RCT) 0.223, $N=23$, $P=0.305$; FINAL (CRT) 0.110, $N=23$, $P=0.617$; FINAL (LPT1) 0.123, $N=23$, $P=0.573$; FINAL (LPT2) 0.173, $N=23$, $P=0.428$; FINAL (SR) 0.458*, $N=22$, $P=0.032$ *; FINAL (RCT) 0.228, $N=23$, $P=0.295$; CREDIT (CRT) -0.110, $N=23$, $P=0.615$; CREDIT (LPT1) -0.034, $N=23$, $P=0.876$, CREDIT (LPT2) 0.010, $N=23$, $P=0.961$; CREDIT (SR) 0.507*, $N=22$, $P=0.016$ *; CREDIT (RCT) 0.159, $N=23$, $P=0.466$.

Information technology students less modifiable: YEAR (CRT) 0.190, $n=27$, $p=0.341$; YEAR (LPT1) 0.136, $N=27$, $P=0.497$; YEAR (LPT2) 0.106, $N=27$, $P=0.596$; YEAR (SR) 0.529*, $N=24$, $P=0.008$ *; YEAR (RCT) -0.062, $N=27$, $P=0.758$; EXAM (CRT) 0.087, $=27$, $P=0.666$; EXAM (LPT1) 0.124, $N=27$, $P=0.536$; EXAM (LPT2) 0.126, $N=27$, $P=0.529$, EXAM (SR) 0.505*, $N=24$, $P=0.012$ *; EXAM (RCT) -0.078, $N=27$, $P=0.696$; FINAL (CRT) 0.096, $N=27$, $P=0.634$ FINAL (LPT1) 0.137, $N=27$, $P=0.494$; FINAL (LPT2) 0.130, $N=27$, $P=0.516$; FINAL (SR) 0.534*, $N=24$, $P=0.007$ *; FINAL (RCT) -0.072, $N=27$, $P=0.720$; CREDIT (CRT) 0.011, $N=27$, $P=0.955$; CREDIT (LPT1) 0.147, $N=27$, $P=0.464$, CREDIT (LPT2) 0.145, $N=27$, $P=0.468$; CREDIT (SR) 0.461*, $N=24$, $P=0.023$ *; CREDIT (RCT) 0.077, $N=27$, $P=0.701$.

More modifiable electrical engineering students: YEAR (SR) 0.360, $N=16$, $P=0.170$; YEAR (CRT) 0.590*, $N=16$, $P=0.016$ *, YEAR (LPT1) 0.646*, $N=16$, $P=0.007$ *, YEAR (LPT2) 0.492, $N=16$, $P=0.053$, YEAR (RCT) 0.228, $N=16$, $P=0.395$; YEAR (EAT) 0.019, $N=16$, $P=0.943$; FINAL (SR) 0.356, $N=16$, $P=0.175$; FINAL (CRT) 0.590*, $N=16$, $P=0.016$ *, FINAL (LPT1) 0.638*, $N=16$, $P=0.008$ *; FINAL (LPT2) 0.481, $N=16$, $P=0.059$; FINAL (RCT) 0.204, $N=16$, $P=0.447$; FINAL (EAT) -0.013, $N=16$, $P=0.959$; CREDIT (SR) 0.312, $N=16$, $P=0.239$; CREDIT (CRT) 0.569*, $N=16$, $P=0.021$ *, CREDIT (LPT1) 0.768*, $N=16$, $P=0.001$ *; CREDIT (LPT2) 0.625*, $N=16$, $P=0.01$ *, CREDIT (RCT) 0.195, $N=16$, $P=0.467$; CREDIT (EAT) -0.080, $N=16$, $P=0.766$.

Less modifiable electrical engineering students: YEAR (SR) 0.530*, $N=20$, $P=0.16$ *; YEAR (CRT) -0.26, $N=20$, $P=0.267$, YEAR (LPT1) 0.136, $N=20$, $P=0.566$, YEAR (LPT2) 0.101, $N=20$, $P=0.669$, YEAR (RCT) 0.200, $N=20$, $P=0.396$; YEAR (EAT) 0.186, $N=20$, $P=0.431$; FINAL (SR) 0.549*, $N=20$, $P=0.012$ *, FINAL (CRT) -0.268, $N=20$, $P=0.252$, FINAL (LPT1) 0.110, $N=20$, $P=0.643$; FINAL (LPT2) 0.078, $N=20$, $P=0.741$; FINAL (RCT) 0.184, $N=20$, $P=0.435$; FINAL (EAT) 0.182, $N=20$, $P=0.441$; CREDIT (SR) 0.386, $N=20$, $P=0.093$; CREDIT (CRT) -0.106, $N=20$, $P=0.656$, CREDIT (LPT1) 0.275, $N=20$, $P=0.240$; CREDIT (LPT2) 0.238, $N=20$, $P=0.311$, CREDIT (RCT) 0.242, $N=20$, $P=0.304$; CREDIT (EAT) 0.276, $N=20$, $P=0.237$ For the less modifiable group the SR correlations evidenced better results and for the more modifiable group the CRT evidenced better correlations. Significant difference between both groups ($p=0.02$ for final marks and 0.045 for CREDITS).

Sample was further divided into low achieving but more modifiable groups (information technology and electrical engineering).

Information technology group: YEAR (CRT) 0.099, $N=16$, $P=0.714$; YEAR (LPT1) 0.221, $N=16$, $P=0.410$; YEAR (LPT2) 0.247, $N=16$, $P=0.356$; YEAR (SR) 0.204, $N=15$, $P=0.465$; YEAR (RCT) 0.056, $N=16$, $P=0.834$. EXAM (CRT) 0.024, $N=16$, $P=0.929$; EXAM (LPT1) 0.234, $N=16$, $P=0.382$; EXAM (LPT2) 0.363, $N=16$, $P=0.166$, EXAM (SR) 0.121, $N=15$, $P=0.666$, EXAM (RCT) 0.051, $N=16$, $P=0.851$; FINAL (CRT) 0.050, $N=16$, $P=0.854$, FINAL (LPT1) 0.252, $N=16$, $P=0.345$, FINAL (LPT2) 0.349, $N=16$, $P=0.184$. FINAL (SR) 0.160, $N=15$, $P=0.569$, FINAL (RCT) 0.059, $N=16$, $P=0.827$. CRED (CRT) -0.024, $N=16$, $P=0.939$, CREDIT (LPT1) 0.375, $N=16$, $P=0.152$, CRED (LPT2) 0.418, $N=16$, $P=0.107$, CRED (SR) 0.141, $N=15$, $P=0.614$, CRED (RCT) 0.034, $N=16$, $P=0.898$.

Electrical engineering group: YEAR (SR) 0.253, $N=11$, $P=0.452$, YEAR (CRT) 0.568, $N=11$, $P=0.038$, YEAR (LPT1) 0.561, $N=11$, $P=0.072$, YEAR (LPT2) 0.325, $N=11$, $P=0.329$, YEAR (RCT) 0.236, $N=11$, $P=0.483$, YEAR (EAT) 0.032, $N=11$, $P=0.925$; FINAL (SR) 0.259, $N=11$, $P=0.441$, FINAL (LPT2) 0.587, $N=11$, $P=0.057$, FINAL (LPT1) 0.588, $N=11$, $P=0.057$, FINAL (LPT2) 0.348, $N=11$, $P=0.294$; FINAL (RCT) 0.226, $N=11$, $P=0.503$, FINAL 9EAT) 0.004, $N=11$, $P=0.990$, CREDIT (SR) 0.280, $N=11$, $P=0.404$, CREDIT (CRT) 0.610*, $N=11$, $P=0.046$ *, CREDIT (LPT1) 0.851*, $N=11$, $P=0.001$ *, CREDIT (LPT2) 0.639*, $N=11$, $P=0.034$ *, CREDIT (RCT) 0.175, $N=11$, $P=0.605$, CREDIT (EAT) -0.176, $N=11$, $P=0.603$. For the total information technology and electrical engineering groups, the SR scores were better correlated, however when the groups were divided into low achieving but more modifiable, there were stronger correlations between LPT and CRT tests. Thus, SR scores do not seem to be a good predictor for the group of students who are lower achieving but who are able to benefit from mediation.

Study no 4: comparisons between some groups

Difference between DET (n=6) and Non DET (n=90) groups: (t test for independent samples at 0.05 (two tailed) significant difference found between the two groups on RCT and CRT: RCT DET MEAN 2.83 (sd=1.32) and RCT Non DET mean 6.98 (sd=2.93), $t = -3.430$, $df=94$, $p=0.001$. CRT DET MEAN 14.50 (sd3.14), CRT Non DET mean 19.48 (5.07), $t = 2.368$, $df=94$, $p=0.019$.

Study no 5:

High (n=19) and low (n=19) SES groups: significant difference between the two groups on LPT1, low SES mean 14.31 (sd= 5.706) high SES mean 18.05 (sd 4.938). $t = -2.158$, $df=36$ $p=0.037$.

More (n=37) and less (n=35) successful students: significant difference between two groups on LPT2, more successful mean 21.91 (sd3.311), less successful mean 19.82 (sd 5.321) $t = 0.012$ $df=70$, $p=0.047$. Therefore mediated lesson played an important role in improving LPT scores for the more successful students.

(21) The weight of the evidence in the study indicates that it is possible to find alternative measures to matriculation and static tests in selecting disadvantaged students by making use of the zone of proximal development. For less modifiable students, the best predictors of academic success are the conventional manner of testing, namely matriculation examination results. For more modifiable students, the results indicate that alternative predictors are more helpful. Hypothesis 1 is supported, hypothesis 2 is partially supported and hypothesis 3 is generally supported

(22) Criterion variables included four measures used for academic performance, namely year mark, exam mark, grade point average and number of credits obtained. The Swedish Rating (SR) score was also used for previous academic achievement. Static tests included the RCT, EAT and MA and the dynamic tests included the LPT, CRT (all five being cognitive tests). The SPQ and MSLQ were regarded as non-cognitive tests



(23) Small sample sizes were used in the various studies and the dynamic assessment procedures were of short duration with not enough attention given to transfer issues.

(1) Gaydon, V.P.

(2) 1988

(3) Predictors of performance of disadvantaged adolescents on the Soweto/Alexandra gifted child programme

(4) University of the Witwatersrand

(5) Unpublished Masters dissertation

(6) Johannesburg

(7) The study aimed to determine the relative effectiveness of both conventional and dynamic assessment techniques in predicting performance of culturally and socio-economically disadvantaged students on the Soweto gifted child programme, with the further aim of developing more effective selection procedures for the programme. This sample had been used during the Hoffenberg (1988) study

(8) 99

(9) 2

(10) Between subjects, experimental and control groups

(11) Black high school pupils who were chosen for the gifted programme

(12) 13-18.8 (ranged from std6-std10 (grade 8 - grade 12)). Mean age 14.8.

(13) LPAD (3 subtests: organiser [organiser b used for transfer], set variations [Ravens progressive matrices used as transfer], verbal analogies [similarities subtests from the WAIS-R and WISC-R used as transfer]), the New South African group test, School performance (1987 mid year aggregate marks in maths, English, science and home language); the Khatena-Torrance creative perception Inventory; The Piers Harris Children's self concept scale and scores on the gifted child programme

(14) Feedback as described by Feuerstein.

(15) Solomon four group design. Pretested experimental group n=24; Pretested control group n=23; unpretested experimental group n=28 and unpretested control group n=25

(16) Pre-selected black high school pupils who were chosen for the gifted programme offered by the Department of Education and Training

(17) One week before and one week after the LPAD tasks were administered, the transfer tasks were administered

(18) No formal hypothesis given: the aim was to determine whether some dynamic tasks could help identify gifted children for a specific education programme. If stated: H0 Dynamic assessment tools will aid in selecting gifted children for an educational programme.

(19) As described by Feuerstein. LPAD was conducted in three sessions of three hours each spanning two weeks

(20) Modifiability index determined by subtracting a pre-index from a post-index scores. Statistical analyses for three groups performed on the whole sample, pretested sample and unpretested sample. Pearson's correlations conducted for LPAD tasks and transfer measures (other correlations were conducted but were not useful for the purposes of this meta-analytic study. The following significant results were obtained from the LPAD tasks and their transfer equivalents in addition to school results and the New South African group test (p=0.0001 accepted as significant).

Key: VAT - Verbal analogies test; ravens - Raven's Progressive matrices, simtot - Similarities Total score, NSAGT - New South African Group Test

Whole sample (n=99) LPAD organizer-science r=0.468, p=0.0001, n=99; organizer-aggregate r=0.485, p=0.0001, n=99; VAT-English r=0.477, p=0.0001, n=95; VAT-science, 0.55045, p=0.0001, n=99; VAT-aggregate r=0.5528, p=0.0001, n=99. transfer measures: organiser b-science, r=0.556, p=0.0001, n=99; organizer b-aggregate, r=0.523, p=0.0001, n=99; ravens-science r=0.479, p=0.0001, n=99; ravens-aggregate, r=0.475, p=0.0001, n=99; simtot-English r=0.454, p=0.0001, n=95; simtot-science r=0.449, p=0.0001, n=99; simtot-aggregate r=0.514, p=0.0001, n=99. Static measures: NSAGT-science r=0.50, p=0.0001, n=94, NSAGT-aggregate r=0.50, p=0.0001, n=94

A stepwise regression was performed with those variables that correlated significantly in order to determine the relative contributions of each measure to variations in the criterion measure. Four variables accounted for 49.43% of variation in criterion: VAT LPAD 31.14%, school performance 9.66%, organiser b 5.32% and simtot 3.31%

Pretested sample (n=47): VAT-science r=0.53, p=0.0001, n=47; VAT-aggregate r=0.59, p=0.0001, n=47; transfer measures: ravens-science r=0.58, p=0.0001, n=47. Stepwise regression conducted to account for variance (46.66%): VAT LPAD 35.19%, simtot 6.45% ravens 5.02%, results of the unpretested sample (n=53): LPAD tasks: organizer-science r=0.592, p=0.0001, n=52; organizer-aggregate r=0.59, p=0.0001, n=52; VAT-science, r=0.57, p=0.0001, n=52; VAT-aggregate r=0.51, p=0.0001, n=52; transfer tasks: organizer b - science r=0.64, p=0.0001, n=52, organizer b - aggregate r=0.64, p=0.0001, n=52, simtot-english, r=0.58, p=0.0001, n=50, simtot-science, r=0.50, p=0.0001, n=52; simtot-aggregate r=0.56, p=0.0001, n=52. static measures: NSAGT-science r=0.59, p=0.0001, n=50; NSAGT-aggregate r=0.59, p=0.0001, n=50. Stepwise regression conducted: of which two variables account for 53.89% of variance - science marks (40.96) and organizer b (12.93%).

(21) In the case of the whole sample, the VAT LPAD task, overall school performance, Organizer posttest and similarities accounted for 49.43% of variance in the criterion measure. For the pretested sample, Raven's posttest contributed to the variation in the criterion measure. For both the whole sample and unpretested sample, the VAT of the LPAD was a significant contributor to variance

(22) Baseline measures: New South African group test, school performance, Khatena-Torrance, Piers Harris children's self concept, teacher temperament questionnaire, progressive matrices, LPAD tasks (all three plus their transfer measures). Criterion measures included: performance on the gifted child programme, as measured by English, maths, science and aggregate marks

(23) Organizer pretest and Organizer posttest were similar in form and the group who were pretested and posttested on the task actually performed less well compared to the group who received no mediation. This was reasoned to be due to tedium and fatigue. In conclusion these LPAD tests along with the static tests would help in identifying gifted children for the Soweto Programme. Whether or not mediation itself was effective cannot be conclusively stated, as correlations between modifiability and performance on the programme were poor (no significant correlations were evidenced between the modifiability index and programme). The mediation offered may not have been effectively administered.

(1) Gewer, A.**

(2) 1998

(3) Uncovering potential: dynamic assessment of non-verbal reasoning ability in educationally disadvantaged children

(4) University of the Witwatersrand

(5) Unpublished Masters dissertation

(6) Johannesburg

(7) The study investigated the application of dynamic assessment to a sample of black children within a South African township clinic setting

(8) 72

(9) 2 but further subdivided into 4



- (10) Between subjects, experimental and control groups
- (11) Black children who were referred to the clinic with learning difficulties
- (12) Ranged between 9-15 (mean 10.96)
- (13) Raven's Coloured Progressive Matrices (RCPM), Rey-Osterrieth Complex Figure Test (ROCFT), LPAD (set variations I), Lurie and Kozulin's qualitative analysis of matrices
- (14) As described by Feuerstein
- (15) Solomon four group design. Matched groups (pretest-mediation-posttest). Experimental group n=48 and control group = 24. The RCPM was completed by both groups, with the experimental group receiving mediation on the LPAD set variations. Mediation was given either on the same day or later than the pretest (for this reason, the sample was split accordingly for both experimental and control groups). The ROCFT equivalent measure was a drawing from the LPAD. The LPAD was group administered
- (16) Black children who were referred to the child clinic for psychoeducational assessment
- (17) All testing took place on the same day for one half of both groups and on separate days for the other half of both groups - this was done to investigate whether there would be any difference between those who received the pretesting on the same day and those who did not
- (18) H01 - children who perform below average on the RCPM will improve their performance significantly once they have undergone a brief process of mediation, utilising the teaching tasks provided in the LPAD, when compared with a group who receives no such intervention. H02 - children who perform below average on the ROCFT, a test of perceptual organisation and visual memory, will improve their performance significantly once they have undergone a brief process of mediation as outlined in the LPAD, when compared with a group which receives no such intervention. H03 - The quality of responses provided by the children who received mediation will improve significantly on both tasks, in comparison with the children who received no such mediation (as measured by the Lurie and Kozulin scoring method). H04 - the children who receive mediation on the ROCFT will be able to transfer the skills acquired to a similar figure. This transfer will be reflected in the quality of their performance on the similar figure, when compared to that of the group that receives no mediation.
- (19) Initial mediation with the set variations is guided by the subjects' level of functioning, and goals are to define the problem, focus on the task, set rules, regulate behaviour and identify the sequence required to perform on a task. Actual mediation lasted three to four hours
- (20) RCPM results: these results are divided according to the day they were completed - experimental same day, experimental different day; control group same day, control group different day. Two way ANCOVA (using pretest as co-variate), pre test scores are partialled out and the mean adjusted. The ANCOVA included posttest scores, pretest scores, experimental and control groups as well as days on which testing took place as variables. Statistical tests were conducted to see whether mediation had an effect on experimental group as well as to investigate the degree to which pretesting had any effect (even if negative) if written on the same day.

Means and standard deviations for all four groups (n=72): exp group same day pretest 16.71 (sd 4.23), posttest 22.21 (8.85); control group same day pretest 15.75 (sd 5.15), posttest 20.00 (sd 5.82) experimental group different day pretest 19.71 (sd 5.37), posttest 24.25 (sd 8.48); control group different day pretest 17.92 (sd 4.58) posttest 18.58 (sd 5.26). No significant results evidenced for mediation (F1 & 67 = 2.92, P=0.919). Pretest day did not make a difference either even though the number of responses on the posttest increased for the experimental group.

Further analyses on some of the RCPM responses (n=72): experimental group same day pretest 0.08 (sd 0.28), posttest 1.75 (1.96); control group same day pretest 0.42 (sd 0.67), posttest 1.00 (sd 1.48) experimental group different day pre test 1.00 (sd 1.47), posttest 2.21 (sd 1.98); control group different day pretest 0.67 (sd 0.98) posttest 0.42 (sd 1.16). There was a significant result for the intervention in favour of the experimental group (F1 & 67=8.56, P<0.01).

ROCFT results: one way analysis of covariance with pretest as covariate, used to explore mediation on the copying and memory tasks of the ROCFT. All subjects pretested on the same day (n=72): experimental group (copy) pretest 18.17 (sd 9.45), posttest 26.25 (8.36); experimental group (memory) pretest 11.76 (sd 6.80), posttest 23.00 (sd 9.00); control group (copy) pretest 15.85 (sd 7.83), posttest 17.00 (sd 8.09) control group (memory) pretest 9.25 (sd 12.50) posttest 4.59 (sd 6.69). A one way ANCOVA conducted on posttests of copying and memory tasks, pretest as covariate. Significant result for copying task (F1 & 69) = 30.46, P<0.0001 and MEMORY (F1 & 69)=31.96, p<0.0001 in favour of the experimental group. Qualitative measures were not included as they do not lend themselves as accurate measures (even though kappa was calculated)

- (21) The experimental group was able to benefit from mediation and was able to transfer what they had learned to similar tasks
- (22) Criterion variable: only results on the posttest compared with results on the pretest
- (23) Due to the fact that the RCPM was administered to the groups on the same day as the mediation, the control group's performance may have negatively impacted on the overall results. Limited use of the LPAD and the use of more subtests might have been valuable

(1) Tayler, J.

(2) 1996

(3) Assessment of the predictive validity of the learning ability battery

(4) University of the Witwatersrand

(5) Unpublished Masters dissertation

(6) Johannesburg

(7) The aim of the study is to assess the predictive validity of a currently widely used, commercially marketed instrument: the Learning Ability Battery (LAB)

(8) 132

(9) 2

(10) Between subjects design as the sample was further sub-divided into both literate and illiterate groups and tested on the LAB making use of pre and post testing to determine whether training and intervention made a significant difference

(11) Volunteer respondents from four South African and one Angolan organisation (three local government organisations and two organisations representing industry). Educational levels ranged from totally illiterate to matriculated

(12) Age ranged between 20-65

(13) The Learning Ability Battery (LAB)

(14) A training session which included aiding subjects with practice examples

(15) The entire LAB was first used to group the individuals for various language courses and on completion of the courses the word recognition subscale of the LAB was again administered, this time to evaluate the accuracy of the instrument's original groupings. The LAB is divided into verbal and non verbal components. Non-verbal measures included: hand-eye coordination; visual memory short term; visual memory long term; visual perception; visual insight; quantitative perception and seeing the whole picture. Scores range from 0-94 and subjects were divided into three groups based on their performance scores: 0-35; 36-65 and 66-100 (slow, average and fast pace of learning). Each subtest was preceded by coaching (training) in the form of practice examples. The verbal measures included: word recognition test which was available in seven languages consisting of fifty sentences graded into eight levels (std 1- std 8 (grade 3 - grade 10)) and testees had to select the correct word for the sentence. One mark was awarded for every correct answer (range 0-50) and scores were then used to group people into groups

(16) Urbanised black men and women employed by five participating organisations

(17) The test-train-test method took place over a period of five to seven months

- (18) No hypothesis stated: the study attempted to ascertain the predictive validity of the LAB
- (19) During the coaching or training phase the administrator reads the instructions aloud from the test manual and demonstrates each example, individuals are corrected when and where necessary but they move on as a group. There is NO test-train-test method as such is used within this study however a component of pretest teaching and familiarisation is carried out
- (20) Taylor's study was a predictive validity study and as such various factor analysis results will not be included in this section as it does not pertain to the assessment of dynamic assessment. However, in order to determine the significance between pre and posttest for both literate and illiterate groups (the sample was divided into these two groups so as to compare the results), a paired comparison using student's two tailed t tests was computed on the means of the differences between the two scores and results are: illiterate group (n=51), $t = 13.38$ ($p < 0.0001$) $df = 50$ ($t_{crit} = 3.496$), literate group (n=67), $t = 11.88$ ($p < 0.0001$) $df = 66$, ($t_{crit} = 3.460$), therefore groups performed significantly better on the posttest compared to the pretest. Also, the higher the education the higher the scores on the posttest: ($r = 0.67$, $p < 0.0001$) which is interesting from the point of view of dynamic assessment theory which tends to advocate the opposite occurring
- (21) The LAB was found to be both reliable and valid as a predictor of optimal groupings for the purpose of development of lower educated employees. High internal reliabilities are based on moderately well defined and interpretable factors
- (22) The criterion was the objective rating after initial LAB testing and a second score was also obtained for each subject on the word recognition subtest. The subjective criterion included: trainers' subjective ratings of subjects' learning
- (23) Only volunteers were used, hence leading to a biased sample and also no comparisons were made with any other objective instrument.

(1) Nel, A.

(2) 1997

(3) Die voorspelling Van akademiese sukses binne die konteks Van 'n alternatiewe universiteitstoelatingsbeleid (The prediction of academic success within the context of an alternative university admissions policy)

(4) Rand Afrikaans University

(5) Unpublished Masters dissertation

(6) Johannesburg

(7) The study attempted to assess whether the psychometric test battery which the university uses within the framework of an academic support programme can be viewed as valid in predicting future academic performance of these students. For the purpose of this study only information pertaining to the dynamic assessment instrument, the Ability and Processing of Information and Learning Test Battery (APIL) and how it functions will be discussed

(8) 171 (for those who wrote the APIL) and 476 for the multiple regression analysis which included subjects other than those who wrote the APIL

(9) 1

(10) This study was a comparison exercise and sought to determine the best battery of predictors for academic success and as part of a suite of tests, the APIL was included

(11) First year university students who proceeded through a programme which admitted students based on alternative policies

(12) Not stated but an approximation of eighteen can be estimated

(13) The General Scholastic Aptitude Test, (GSAT), APIL and a language assessment tool made up of a reading comprehension test as well as Matriculation (M) scores were also used

(14) During the administration of the APIL four familiarisation sessions are included which enable the subjects to understand what is expected of them

(15) Students were assessed on a number of instruments including static and dynamic assessments. There was no experimental nor control groups as scores were entered into regression models to find variables which proved predictive of university success

(16) New entrants into university who had gained admission via an alternative route

(17) The APIL takes a few hours to administer

(18) General hypothesis: H0 - learning potential, intelligence and language competency as measured by the university battery for admissions is a valid predictor of academic performance of students who are admitted via the alternative policy. Specific hypotheses: H01: no statistically significant relationship exists between students results on the learning potential test and academic achievement. HA1 there is a significant relationship between the two. H02: no statistically significant relationship exists between students results on their M scores and academic achievement. HA2 there is a significant relationship between the two. H03: no statistically significant relationship exists between students' combined results (learning potential test, M scores, reading comprehension and GSAT) and academic achievement. HA3: there is a significant relationship between the combined results and academic achievement. H04: there is no significant statistical way in which the GSAT, APIL and reading comprehension test can differentiate between those who will pass and those who will fail in their subjects and those who will not gain admission to exams. HA4: this can be shown to be proven true. H05: there is no significant statistical way in which the GSAT, APIL, reading comprehension test can differentiate between those who are successful and those who are unsuccessful. HA5: this can be shown to be proven true

(19) Mediation as prescribed by Tayler the developer of the APIL which is considered a learning potential test. Four sessions each included familiarisation with the task at hand

(20) Correlations between the following and academic achievement (n=171): APIL: $r = 0.279$ significant at $p = 0.0001$; GSAT: $r = 0.291$ significant at $p = 0.0001$; reading comprehension Afrikaans: $r = 0.412$ significant at $p = 0.0001$; reading comprehension English: $r = 0.309$ significant at $p = 0.0001$ and M score: $r = 0.241$ significant at $p = 0.0001$, hence null hypotheses rejected.

For regression analyses (all variables entered at once) (n=476): all three (APIL, M score and GSAT) delivered a significant contribution to the regression ($p < 0.05$) with t values of 2.073 for APIL; 2.1496 for M score and 2.125 for GSAT. b values for APIL, GSAT and M score: b for APIL = 0.009 (significant at 0.0001) and b for APIL = 0.125; b for GSAT = 0.009 (significant at 0.0001) and b for GSAT = 0.143; b for M score = 0.032 (significant at 0.0001) and b for M score = 0.126. $r^2 = 0.107$ and standard error = 0.919.

Correlations: APIL and achievement = 0.275; GSAT and achievement = 0.292 and M score and achievement = 0.248, so although M score correlated the least in comparison to GSAT and APIL it still contributed more to the regression model. Hence, null hypothesis rejected. There is a significant relationship between results on the APIL, GSAT and M scores combined and their academic performance.

Hierarchical regression: step one: M score $r = 0.248$, $r^2 = 0.061$ standard error 0.940, step two: GSAT $r = 0.315$, $r^2 = 0.099$ standard error 0.922, step three: APIL $r = 0.327$, $r^2 = 0.107$ standard error 0.919. The second hierarchical regression: step one: M score $r = 0.248$, $r^2 = 0.061$ standard error 0.940, step two: APIL $r = 0.312$, $r^2 = 0.098$ standard error 0.923, step three: GSAT $r = 0.327$, $r^2 = 0.107$ standard error 0.919.

Intercorrelations between different tests: APIL-GSAT = 0.684 (n=488); APIL-M score = 0.416 (n=488); APIL-reading comprehension (Afrikaans) = 0.667 (n=83), APIL reading comp (English) 0.594 (n=190).



Discriminant analysis (predictors are APIL, GSAT and M score) correctly placed 50% of students into the three groups namely: no admission to exam, failed and passed. Closer inspection yielded that GSAT correctly placed 50.82% of cases, (and the level of chance = 38%), M scores correctly placed 50.20% and APIL correctly placed 49.20%. When the three classification levels are reduced to two, namely, failed or successful the total percentage correctly classified is 64.29%.

(21) GSAT and APIL correlated to the same extent with academic performance and the relationship between the two instruments was also relatively high which implies that the two probably measure the same construct (evidencing a correlation is 0.684). When used in combination, the M score, APIL and GSAT explain more of the variance in academic performance than any of the factors individually. GSAT and APIL are equally good predictors

(22) Criterion is achievement or academic success, predictor variables include the GSAT, APIL and reading comprehension test as well as M scores

(23) Different testers were used and not all testing took place on the same day, also the APIL is meant for a small scale testing situation not a large scale one and it is also quite a long test.

(1)Hoffenberg, S.R.**

(2) 1988

(3) Effectiveness of the learning potential assessment device with high achieving adolescents from a disadvantaged community

(4) University of the Witwatersrand

(5) Unpublished Master's dissertation

(6) Johannesburg

(7) The study aimed to assess the effectiveness of dynamic assessment among a group of academically superior individuals from a disadvantaged black community in South Africa. This sample had been used for other statistical analysis purposes in the study by Gaydon (1988)

(8) 100

(9) 4

(10) Solomon four group design (experimental and control group divided on the basis of pretesting)

(11) Adolescents who were achieving in the top of their academic classes. As determined by the New South African group test, the average IQ for the sample was 90.7 (ranged between 74 - 116), 46% male and 54% female

(12) Age ranged between 12-19 mean age 14.8 sd = 1.65

(13) LPAD: organiser (a) verbal analogies (VAT) and set variations I and II. Transfer measures included the organiser (b), similarities subtests from the WISC-R and WAIS-R and Raven's standard progressive matrices (RSPM)

(14) As outlined by Feuerstein

(15) A Solomon four group design was used to ensure that pretesting did not have a practice effect on the control group. The experimental group was exposed to dynamic testing (mediated learning experience (MLE) and its resultant effects) and control group not exposed to MLE. The two groups were compared on their scores on the LPAD tasks and on transfer tasks comparable to the LPAD. The pretested experimental group 1 n=24; pretested control group 2 n= 23; unpretested experimental group 3 n=28 and unpretested control group 4 n=25. The group administered pretesting was conducted by post graduate students in groups of 35-45 subjects. The LPAD was administered in smaller groups of nine subjects and both the experimental and control groups were divided into smaller groups for purposes of mediation

(16) Pre-selected group of high school top achievers in their respective classes who attended the gifted child programme

(17) The transfer tasks were administered one week before and one week after the LPAD tasks and in total the testing took four weeks. Administration of the LPAD took two weeks

(18) H01: following mediation, the performance of the experimental group will be significantly better than that of the control group on certain LPAD tasks, namely the Organiser (A), verbal analogies and set variations 1 & 2. H02: Following mediation on the LPAD, the performance of the experimental group will be significantly better than the control group on the transfer measures, namely the Organiser (B), similarities subtests of the WISC-R and the WAIS-R and the Raven's standard Progressive Matrices

(19) As prescribed by Feuerstein

(20) ANOVA and chi square tests revealed no differences between all four groups of the Solomon design on SES, gender, academic achievement nor IQ. ANCOVA conducted on pretested groups to determine the effects of mediation on LPAD and transfer tasks taking pretests into account and ANOVA on the unpretested groups to determine mediation on LPAD and transfer measures and ANOVA on whole sample experimental vs control groups.

Means and sd's of LPAD results for pretested and unpretested groups:

PRETESTED GROUP (n=47)

experimental group (n=24) organiser mean 9.54 (sd 4.37); VAT mean 14.96 (sd 3.64) set variations I mean 25.88 (sd 3.52) set variations II mean 38.75 (sd 7.41) control group (n=23) organiser mean 8.39 (sd 4.29); VAT mean 13.13 (sd 3.05) set variations I mean 23.96 (sd 4.22) set variations II mean 33.78 (sd 8.22)

UNPRETESTED GROUP (n=53)

experimental group (n=28) mean 7.86 (sd 3.76); VAT mean 15.14 (sd 3.67) set variations I mean 24.68 (sd 4.13) set variations II mean 39.32 (sd 13.36) control group (n=25) mean 9.00 (sd 5.47); VAT mean 15.16 (sd 4.30) set variations I mean 23.28 (sd 4.97) set variations II mean 34.12 (sd 11.72).

Means and sd's of LPAD results by group (combined pretested and unpretested):

EXPERIMENTAL GROUP (n=52): organiser (a) mean 8.63 (Sd 4.10), VAT mean 15.06 (sd 3.62) set variations I mean 25.23 (sd 3.87) set variations II mean 39.06 (Sd 10.92)

CONTROL GROUP (n=48) organiser (a) 8.71 (sd 4.90) VAT mean 14.19 (sd 3.85), set variations I mean 23.60 (sd 4.59), set variations II mean 33.96 (sd 10.09). For the whole group (experimental and control groups) set variations yielded a significant difference between groups, $f(df=1, 898) = 5.85$ at $p < 0.05$ otherwise both groups not significant on any other task.

Means and sd's of transfer measures posttest results for pre and unpretested groups:

f score results for difference between experimental and control groups on transfer measures pretests:
N=47, organiser (b) 0.002; similarities 0.426 and RSPM 0.240.

Results on LPAD by group (experimental vs control groups):



PRETESTED GROUP (N=47)

experimental vs control group: organiser (a) ANCOVA 1.560, ANOVA 0.827; VAT ANCOVA 3.868, ANOVA 3.465. set variations I ANCOVA 3.228, ANOVA 2.879; set variations II ANCOVA 5.560* (p<0.05) ANOVA 4.741* (p<0.05)
 unpretested group exp vs control group: ANOVA: organiser (a) 0.800; VAT 0.000, set variations I 1.251 Set variations II 2.247.

f scores for results on the posttest transfer measures (exp vs con group)

PRETESTED experimental vs control Organiser (b) ancova 2.117 ANOVA 1.274; similarities ancova 0.380 ANOVA 0.006 RSPM ancova 1.540 ANOVA 1.440.

UNPRETESTED experimental vs control ANOVA: organiser (b) 3.247. similarities 1.475 RSPM 0.543.

f scores for results on posttest transfer measures (pretested vs unpretested)

EXPERIMENTAL pretested vs unpretested: ANOVA - organiser 1.328, similarities (WISC-r) 2.866 and WAIS - 2.592 and Raven's 0.008;
 CONTROL pretested vs unpretested: ANOVA - organiser 3.062, similarities (WISC-r) 0.232, WAIS 0.092 and Raven's 0.172.

f scores for results on LPAD by group (pretested and unpretested groups) n=100

EXPERIMENTAL (pretested vs unpretested) ancova organiser (a) 2.232 vat 0.033, set variations I 1.242, variations II 0.035
 CONTROL (pretested and unpretested) ANOVA organiser (a) 0.182, vat 3.503, set variations I 0.256 set variations II 0.013

means and sd's of transfer measures posttest results for pretested and unpretested groups:

PRETESTED experimental group (n=24), organiser (b) 10.17 (sd 4.69), similarities (WAIS and WISC averaged) 9.21 (sd 2.48), RSPM (percentile ranks) 76.46 (sd 22.91). Control group (n=23), organiser (b) 8.61 (sd 4.77), similarities (WAIS and WISC averaged) 9.18 9sd 3.72), RSPM (percentile) 72.61 (sd 27.55).

UNPRETESTED experimental group (n=28), organiser (b) 8.57 (sd 5.21), similarities (WAIS and WISC averaged) 8.36 (sd 3.52), RSPM (percentile) 77.14 (sd 26.61), control group (n=25), organiser (b) 11.24 (sd 5.57), similarities (WAIS and WISC averaged) 8.94 (sd 3.46) RSPM (percentiles) 72.20 (sd 24.37). No significant differences between groups.. Groups were then combined to form only experimental and control groups.

Means and sd's of transfer measures posttest results by group:

EXPERIMENTAL (n=52): organiser (b) 9.31 9sd 4.99), similarities (WAIS and WISC averaged) 8.75 (sd 3.17), RSPM (percentiles) 76.83 (24.73).

CONTROL (n=48), organiser (b) 9.98 (sd 5.32), similarities (WISC) 9.05 (sd 3.56) RSPM (percentile) 72.92 (sd 25.45): hence no significant diff between experimental and control groups.

Means and sd's of pre and posttest transfer measures:

EXPERIMENTAL (n=24) organiser (b) pretest 6 (sd 3.79), organiser (b) posttest 10.17 (sd 4.69), similarities (WAIS) pretest 8.69 (sd 2.76), posttest 9.21 (2.48), RSPM (percentiles) pretest 62.71 (sd 27.58) posttest 76.46 (22.91)

CONTROL (n=23): organiser (b) pretest 5.96 (sd 3.34), posttest 8.61 (4.77), similarities (WAIS) pretest 8.56 (3.55), posttest 9.18 (3.72), RSPM (percentiles) pretest 56.96 (sd 25.71) posttest 73.70 (27.10). Hence mediation on the LPAD did not appear to have a significant effect on the posttest transfer measure scores

(21) Only on one of the LPAD tasks, set variations II, was the performance of the experimental group significantly better than the control group. Trends in the expected direction were observed in the remaining tasks, but no significant differences were observed on the comparable measures. The first hypothesis was only partially supported while the second hypothesis did not receive any support

(22) No predictor variables as such; two groups were compared on findings of scores on LPAD and transfer measures

(23) The mediation offered was inadequate. Also a ceiling effect may have resulted in experimental subjects' scores not improving as much as expected.

(1) Zolezzi, S.A.

(2) 1995

(3) The effectiveness of dynamic assessment as an alternative aptitude testing strategy

(4) University of South Africa

(5) Unpublished doctoral thesis

(6) Pretoria

(7) The study sets out to evaluate the effectiveness of a dynamic approach to aptitude testing. It is believed that an alternative testing format could be facilitated by using a test-train-test procedure within a learning potential paradigm by making use of a "Newtest" battery (which consists of two tests given in both traditional and dynamic format) which will be able to identify students (particularly disadvantaged) who have the potential to succeed at university

(8) 50

(9) 2

(10) Mixed design incorporating both within and between-subject designs

(11) Students who were completing their Matriculation and who were prospective first year university students. n=28 female and n=22 male

(12) Age ranged between 16-20 median = 17-18

(13) Dynamic instruments included the Deductive reasoning test (including traditional and dynamic modes of assessments) (DRT/E) and the Pattern relations test (including traditional and dynamic modes of assessments) (PRT/E). Learning potential is operationalised as the differences between pre and post test results on both instruments (LR). Traditional tests included the High level battery (B/75) subtests, Mental Alertness test (MAT), Reading Comprehension (RC), Standard level arithmetic reasoning test (a/131) (AR) and the Raven's matrices (RM). Matriculation results also included as variable (MAT)

(14) As indicated by Feuerstein

(15) Two groups of students: 1992 group (n=18) who were assessed traditionally and the 1993 group (n=32) who were assessed dynamically. Both groups will be compared according to different predictor variables. EXPERIMENTAL group (n=18) 1992, advantaged (n=12) and disadvantaged groups (n=6), predictor variables MAT, MA, RC, AR, RM, DT/T and PR/T. Criterion variables: 1993 BSc marks, BCom and BA. DYNAMIC group (n=32), advantaged (n=22) disadvantaged (n=10), Predictor variables included the DR/E and PR/E, 1994 BSc, BCom and BA

marks. Group administered in groups of 7-8 students throughout the year but actual testing sessions at any one time lasted approximately 2 hours and 10 minutes for the traditional group. Smaller groups of 5-6 students were used for the dynamic sessions.

(16) Matriculation students who were prospective first year students for the 1992 and 1993 intake

(17) For the traditional group the entire testing session took 2-3 hours. For the dynamic group 15-20 minutes were used for mediation but the total testing lasted 3 hours and 45 minutes. Pretesting was followed by a short break then followed by mediation and then posttest

(18) HA1: it can be expected that prediction of university success will be significantly enhanced through a dynamic testing situation as operationalised for the purpose of the present study. HA2: Advantaged and disadvantaged students will have different predictors correlating significantly with the criterion of university success. HA3: no significant relationship exists between current ability and learning potential

(19) Although the procedure was standardised, interaction with subject did occur

(20) Hypothesis HA1:

Means and sd's of traditional test and November exam for the traditional 1992 group include: Matric mean 65.2 (sd11.6), MA mean 27.3 (sd 6.1), RC mean 12 (sd 3.6), AR mean 16.2 (sd 8.7), RM mean 33.1 (sd 3.2), November exam mean 58.1 (sd 8.9), BA mean 55.7 (sd9.9), BCom mean 54.4 (sd 3.9) BSc mean 64 (8.8).

The following traditional predictors and November exam correlations for the traditional group (1992) include : matric and November = 0.70* (sig at $p < 0.05$), MA and November = 0.26, RC and November = 0.41, AR November = -0.20 and RM and November = 0.09.

The following traditional and dynamic measures and November exam correlations for the dynamic group (1993) include: MAT and November = 0.69* (sig at $p < 0.05$), DR/T and November = 0.16, PR/T and November = 0.06, DR/E and November 0.36, PR/E and November 0.14, LP/DR and November 0.21 and LP/PR and November = 0.04.

Means and sd's of dynamic and traditional measures and November results for the dynamic group (1993) include: MAT mean 69.25 (sd 9.8), DR/T mean 22.9 (sd 5.1), PR/T mean 15.6 (sd 5.3), DR/E mean 25.8 (sd5.5), PR/E 22.7 (sd 2.9), LP/DR mean 2.9 (sd5.4), LP/PR mean 7.1 (sd 3.4), November exam mean 64 (sd 8.9), BA mean 64.7 (sd 6.9), B Com mean 58 (sd9) BSc mean 67 (sd 8.9).

LP score for dynamic group (Deductive reasoning enriched minus traditional score) for advantaged students include: mean 1.3 (sd 5.4) and disadvantaged group mean 6.3 (sd 3.8). Pattern relations enriched minus traditional score for advantaged students mean: 7 (sd 3.7) and disadvantaged students mean 7.3 (sd 2.9). However, even though predictability of the traditional test formats are enhanced through dynamic tests it is not significant. Deductive reasoning increases from 0.16-0.36 and pattern relations from 0.06-0.14.

Following are results for each faculty separately:

1992 traditional group BA - correlations between traditional measures and November: MAT- November = 0.76* (sig $p < 0.05$) , MA - November = -0.01, RC-NOV-0.21, AR-NOV = -0.13, RM-NOV 0.09. Correlations between dynamic and traditional measures for the dynamic group 1993: MAT- November 0.48* (sig $p < 0.05$), DR/T-NOV -0.16, PR/T-NOV -0.23, DR/E-NOV -0.19, PR/E-NOV -0.32, LP/DR NOV -0.01 LP/PR NOV -0.15. No correlations were conducted for the traditional group of 1992 for faculties of science and commerce as the groups were too small.

The dynamic group of both the commerce and science students (1993) (for the November exams): MAT -BSc 0.88 and BCom 0.7, DR/T -BSc 0.18 and BCom -0.03, PR/T-BSc -0.02 and BCom 0.12, DR/E - BSc 0.43 and BCom 0.26, PR/E -BSc 0.14 and BCom 0.32, LP/DR-BSc 0.26 and BCom 0.24, LP/PR BSc 0.26 and BCom 0.09. There was a sig increase for the DR/E measure for the BSc group (from 0.18 - 0.43 on the DR/T - DR/E).

Regarding HA2:

The following results will be assessed in terms of two groupings: advantaged and disadvantaged students:

Advantaged: traditional group correlations - MAT-NOV 0.47* (sig $p < 0.05$), MA-NOV 0.21 RC-NOV 0.33 AR-NOV -0.34 RM-NOV 0.29. Dynamic group correlations: MAT-NOV 0.64* (sig $p < 0.05$) , DR/T 0.09 PR/T -0.06, DR/E-NOV 0.27, PR/E -NOV -0.04, LP/DR-NOV 0.21 LP/PR-NOV 0.04. Only the matriculation result was significant for the advantaged group.

Disadvantaged group: traditional group: MAT-NOV 0.76* (sig $p < 0.05$), MA-NOV 0.24, RC-NOV 0.39 AR-NOV -0.64 RM-NOV -0.15. Dynamic group: MAT-NOV 0.86* (sig $p < 0.05$) DR/T -NOV 0.50 PR/T-NOV 0.38, DR/E-NOV 0.66* (sig $p < 0.05$) PR/E-NOV 0.61* ($p < 0.05$) LP/DR-NOV 0.23 LP/PR-NOV 0.04. Only the matriculation result was significant for the traditional group but although matric is still significant for dynamic group so too is DR/E and PR/E for the disadvantaged group.

Means and sd's for the disadvantaged and advantaged groups in the dynamic group 1993:

Advantaged group: DR/E mean 25 (sd 5.9) PR/E mean 22.2 (sd 3.6) LP/DR mean 1.3 (sd 5.4) LP/PR mean 7 (sd 3.7).

Disadvantaged group: DR/E mean 27.7 (sd 4) PR/E mean 23.9 (sd 3.3) LP/DR mean 6.3 (sd 3.8) LP/PR mean 7.3 (sd 2.9). Scores were higher for the disadvantaged group both on enriched and LP measures.

Correlations between traditional and dynamic measures for the advantaged students 1993:

MAT-BA 0.52 & BSc 0.86, DR/T -BA -0.26 & BSc 0.13, PR/T -BA -0.25 & BSc -0.38 DR/E - BA -0.21 & BSc 0.33, PR/E-BA -0.27 & BSc -0.08, LP/DR-BA 0.09 & BSc 0.22 LP/PR - BA 0.19 & BSc 0.54.

Correlations between dynamic and November measures for the disadvantaged group 1993:

DR/E-NOV 0.75, PR/E-NOV 0.66, LP/DR-NOV 0.22 LP/PR-NOV -0.11, hence the dynamic measures provide a better prediction of academic success for the disadvantaged students than the traditional tests.

Final grouping of sample into high and low learning potential:

The mean score for the learning potential in the DR/E =1.3 and the PR/E = 7. Those who score above these are high learning potential students and those who score below are low potential students.

(i)Correlations between the dynamic measures and November exams of high learning potential students in the 1993 group:

DR/E-NOV 0.58 PR/E-NOV 0.54 LP/DR-NOV 0.13 LP/PR-NOV -0.65.

(ii) Correlations between the dynamic measures and November exams of low learning potential students in the 1993 group:

DR/E-NOV 0.22 PR/E-NOV -0.01 LP/DR-NOV 0.22 LP/PR-NOV 0.30. Results indicate that the dynamic tests do distinguish between high and low potential students and that there is a significant correlation between university success for the high learning potential group.

(iii) Correlations between learning potential measure and the traditional ability measures for the full group 1993:

LP/DR-DR/T = -0.46, LP/PR-PR/T = -0.75, this hypothesis is confirmed as there is a negative relationship with LP and current ability.

(21) Traditional measures of aptitude were found to be invalid predictors of university success and matric results showed a relationship with academic success for both groups. The Newtest battery measures enhanced the prediction of academic success for both advantaged and disadvantaged groups. The deductive reasoning dynamic measure was found to be a valid predictor of university success for the disadvantaged students. HA1: correlations between predictor and criteria measures for the whole study MAT(1992 group)-NOV 0.70, MAT (1993 group) -NOV 0.69, DR/E-BSc exams 0.43, DR/E(DR/T 0.16) NOVEMBER 0.36, PR/E (PR/T 0.06)-NOV 0.14, hence there is an improvement but not at a significant level. HA2: advantaged group MAT (1992) - NOVEMBER 0.47, MAT (1993)-NOV 0.64, Disadvantaged students MAT (1992)-NOV 0.76, MAT (1993)-NOV 0.86, DR/E-NOV 0.66 AND BSc 0.75, PR/E-NOV 0.61 and BSC 0.66 thus HA2 confirmed although matric still best predictor for both groups. HA3: DR/T-LP/DR = -0.46 and PR/T-LP/DR = -0.75, hence a negative correlation thus HA3 is confirmed.

(22) The criterion measure included academic success based on the end of first year university success in the faculties of arts (BA), commerce (BCom) and science (BSc). Subject variables included level of educational disadvantage (advantaged and disadvantaged) and predictor variables included (i) traditional instruments such as the High level battery (B/75) subtests such as mental alertness (MA), reading comprehension (RC), arithmetic reasoning test (AR) (a/131) and the Raven's matrices (RM) and school results (MAT), deductive reasoning traditional (DR/T) and pattern relations traditional (PR/T) and (ii) dynamic instruments including the deductive reasoning test enriched (DR/E) and pattern relations enriched test (PR/E)

(23) 85% of the sample were volunteer students that were chosen for the study and is therefore biased in terms of the type of student represented (i.e. in terms of motivation and skill). Sample sizes were small thus limiting the generality of any findings.

(1) Van Aswegen, M.

(2) 1997

(3) The standardisation of a learning potential battery for the selection of poorly qualified employees

(4) University of Pretoria

(5) Unpublished Masters dissertation

(6) Pretoria

(7) The purpose of this study is to evaluate the validity and effectiveness of the Transfer, Automatisation, memory and understanding learning potential battery (TRAM-1) to predict future performance of poorly qualified employees on an accelerated adult development programme (a technical college programme)

(8) 101

(9) 1

(10) Within groups study design

(11) Black males employed by a consolidated mining organisation. Educational levels ranged from std two-std nine (grade four-grade eleven), median std 5 (grade 7). Sample chosen were those who were already on an accelerated programme and were chosen by supervisors based on work quality and performance on the Raven's progressive matrices as well as in-house maths and English tests and lastly performance on a structured interview

(12) Mean age 35.4 (sd 3.99)

(13) The TRAM-1 (nonverbal paper and pencil test) consists of the following: phase A test booklet, phase A dictionary, phase B test booklet, phase B dictionary and memory and understanding test booklet. Dimensions tested include speed of learning, accuracy of learning, transfer of learning, automatization of learning, memory and understanding as well as a composite test score

(14) As per T aylor's TRAM-1 instructions which advocate standardised group interaction including a test-teach-test format

(15) A non-experimental, predictive validity design to evaluate the criterion related validity of the TRAM-1. Nonprobability sample (purposive sampling) was used i.e. those subjects already involved in an accelerated programme within the mining company

(16) Disadvantaged black males in the process of improving their academic qualifications within an enriched environment

(17) The complete procedure took one year. The TRAM-1 was administered at the start of developmental programme which lasted for a year. During the administration of the TRAM-1 which takes one hour and 35 minutes, a test-teach-test approach is used

(18) The hypothesis is as follows: The TRAM-1 is a valid predictor of future performance on a development programme aimed at the advanced development of poorly educated employees in the mining industry environment

(19) Standardised group administration making use of a test-teach-test format within testing

(20) Univariate statistics of predictor and criterion variables were conducted. Order of variables: name, mean and sd in brackets, followed by minimum and maximum scores:

PREDICTORS: speed 121.41 (12.99) range 93.8-143.55; accuracy 62.97 (15.91) range 16.46-100; automatization 99.99 (20.36) range 21.09-151.30; transfer 99.99 (10.62) range 67.57-120.93; composite 50 (9.99) range 27.69-65.25.

CRITERIA (criterion tests were those standardised tests used during in-house evaluations within the mining company): English 63.08 (17.23), range 20-90; mathematics 60.63 (19.05) range 9-98; basic mining 58.43 (19.03) range 21-98; ventilation 58.16 (18.33) range 17-97

IN-HOUSE total scores 60.08 (15.33) range 30-95.75; mining 56.96 (20.62) range 17-99; communication 57.36 (17.52) range 17-90; ventilation 54.92 (20.69) range 30-99; mathematics 51.65 (17.46) range 12-93

COLLEGE TOTAL SCORE 55.18 (15.65) range 23.25-92; maths 56.14 (16.64) range 25-95.5; English/communication 60.22 (15.69), range 24.5-89; mining 57.69 (18.58) range 20-98.5; ventilation 56.54 (17.37) range 16.5-98

OVERALL SCORE: 57.63 (14.98) range 26.63-93.68. according to Van Aswegen the TRAM-1 subtests have the following reliabilities: speed of learning 0.81; accuracy of learning 0.83; automatization 0.78; transfer 0.86 and memory and understanding 0.94 with automatization being the only one below the Kuder-Richardson reliability of 0.80, and it was also the poorest predictor variable in the study.

Pearson correlations between dependent (in-house and college exams) and independent (TRAM1) variables (0.25* and above is significant at $p = 0.05$). In-house MATHS-speed 0.39*; - accuracy 0.46*; - auto 0.19; transfer 0.25*, memory 0.41*; composite scores 0.46*. In-house ENGLISH- speed 0.39*; - accuracy 0.23; - auto 0.06; transfer 0.27*, memory 0.31*; composite scores 0.35*. In-house VENTILATION- speed 0.33*; - accuracy 0.19; - auto 0.08; transfer 0.10, memory 0.33*; composite scores 0.30*. In-house BASIC MINING-speed 0.51*; - accuracy

0.27*; - auto 0.13; transfer 0.29*, memory 0.40*; composite scores 0.44*. In-house TOTAL-speed 0.49*; - accuracy 0.35*; - auto 0.14; transfer 0.27*, memory 0.44*; composite scores 0.47*. COLLEGE MATHS-speed 0.39*; - accuracy 0.43*; - auto 0.18; transfer 0.19, memory 0.45*; composite scores 0.46*. COLLEGE COMMUNICATION-speed 0.42*; - accuracy 0.24; - auto 0.16; transfer 0.34*, memory 0.31*; composite scores 0.39*. COLLEGE VENTILATION -speed 0.36*; - accuracy 0.23; - auto 0.09; transfer 0.26*, memory 0.27*; composite scores 0.33*. COLLEGE MINING-speed 0.36*; - accuracy 0.21; - auto 0.02; transfer 0.30*, memory 0.31*; composite scores 0.33*. COLLEGE TOTAL - speed 0.46*; - accuracy 0.33*; - auto 0.13; transfer 0.33*, memory 0.40*; composite scores 0.45*. COMBINED MATHS TOTAL- speed 0.43*; - accuracy 0.49*; - auto 0.20; transfer 0.25*, memory 0.47*; composite scores 0.51*. COMBINED ENGLISH/COMMUNICATION TOTAL-speed 0.45*; - accuracy 0.26*; - auto 0.12; transfer 0.34*, memory 0.34*; composite scores 0.41*. COMBINED VENTILATION-speed 0.39*; - accuracy 0.24; - auto 0.09; transfer 0.21, memory 0.34*; composite scores 0.35*. COMBINED MINING TOTAL-speed 0.46*; - accuracy 0.25*; - auto 0.07; transfer 0.31*, memory 0.38*; composite scores 0.41*. OVERALL CRITERION SCORE-speed 0.49*; - accuracy 0.35*; - auto 0.14; transfer 0.31*, memory 0.43*; composite scores 0.48*. The lack of a significant correlation for automatisisation may be due to the nature of automatisisation's rote learning as opposed to the other subjects learning for understanding. The composite scores emerged as the factor that correlates the highest with all criterion variables. In sum, the following relationships were identified between predictor variables (TRAM-1) and the total criterion score: speed of learning 0.49; accuracy 0.35, transfer 0.33, memory and understanding 0.43, TRAM-1 composite score 0.45. Evaluation of power of TRAM-1 for predicting performance: only the significant results are given -

Predictive power of speed of learning (percentages indicate variance explained followed by f ratio and p values): in house TOTAL 24.1%, F 31.39, P 0.002; college TOTAL 21.4%, F 26.93, P 0.034, OVERALL CRITERION SCORE 24.2% F 31.69 P 0.036

Predictive power of accuracy of learning: IN HOUSE MATHS 20.8%, F 26.04, P 0.001, COLLEGE AMTHS 20.9%, F 26.23, P 0.004, IN HOUSE TOTAL 14.3%, F 16.55, P 0.001, COLLEGE TOTAL 14.7%, F 17.09, P 0.001, OVERALL CRITERION SCORE 15.5%, F 18.15, P 0.001

Predictive power of automatisisation of learning: IN HOUSE VENTILATION 5.8%, F 0.58, P 0.048, COLLEGE VENTILATION 8%, F 0.798, P 0.037

Predictive power of transfer of learning COLLEGE COMMUNICATION 11.6% F 12.95, P 0.005, COLLEGE TOTAL 11.1%, F 12.302, P 0.007, OVERALL CRITERION SCORE 9.8%, F 10.799, P 0.001

Predictive power of memory and understanding IN HOUSE MATHS 16.4% F 19.434, P 0.001 COLLEGE MATHS 20.33% F 25.27, P 0.001, IN HOUSE TOTAL 19.29% F 23.66 P 0.0013, COLLEGE TOTAL 16.22, F 19.166, P 0.0053, OVERALL CRITERION SCORE 18.9%, F 23.08, P 0.0018

Predictive power of TRAM-1 composite score IN HOUSE MATHS 21.57%, F 27.238, P 0.0001, IN HOUSE MINING 19.55%, F 24.061, P 0.0004, COLLEGE MATHS 21.3% F 26.806, P 0.0001, COLLEGE MINING 19.79%, F 12.209, P=0.0007, IN HOUSE TOTAL 22.06%, F 28.03, P 0.0008, COLLEGE TOTAL 20.69%, F 25.841, P 0.007, OVERALL CRITERION SCORE 22.8%, F 29.276, P 0.0016

Intercorrelations with the TRAM-1 (multicollinearity): speed-accuracy 0.33, speed-auto 0.34, speed-transfer 0.45, speed-memory 0.65, speed-composite 0.76; accuracy-auto 0.25, accuracy-transfer 0.18, accuracy-memory 0.58, accuracy-composite 0.65; auto-transfer 0.36, auto-memory 0.45, auto-composite 0.63; transfer-memory 0.49, transfer-composite 0.66, memory-composite 0.92. Only the memory and TRAM-1 composite were correlated above 0.80. The observed low interrelationships between the independent predictor variables are encouraging as this gives assurance that no two predictors measure the same construct.

Influence of moderating variables:

INFLUENCE OF AGE: relationship between age and dependent and independent variables: only significant (0.25 and above p 0.05) correlations are included. Age and speed of learning -0.26*; age and college maths -0.26*, ventilation and age -0.26*, college total and age -0.26*. All observed correlations were negative

INFLUENCE OF EDUCATION: no significant correlations were found. Regarding the influence of the company there was a difference in scores between those based in city areas vs those in rural areas.

(21) Five of the six TRAM-1 dimensions have high correlations with the employees' performance on the accelerated development course. TRAM-1 is an effective predictor of learning performance of poorly qualified employees. Based on the results, the TRAM-1 is a valid predictor of academic performance on the advanced development course. Order of best predictors: speed of learning, TRAM-1 composite score, memory and understanding, accuracy of learning, transfer of learning, automatisisation (which was the poorest), TRAM-1 is also a valid predictor for the in-house course

(22) Independent predictor variable: TRAM1 (speed, accuracy, automatisisation, transfer and composite score) (non-verbal paper and pencil test). Dependent validation (criterion) variables: (1) technical college results such as: Maths, English, communication and basic mining (and an average of these four results) (2) in-house results: maths, English, mining ventilation and basic mining, (3) composites on the English and communication results from both the college results and in-house scores and (4) composite overall score combined scores (which was the combination of the in-house and technical college results)

(23) The study could have included additional independent predictor variables as well as increased of sample size and inclusion of females into sample

(1)Engelbrecht, M.

(2) 1999

(3) Leerpotensiaal as voorspeller Van akademiese sukses Van universiteitstudente (Learning potential as predictor of academic success of university students)

(4) Potchefstroom University

(5) Unpublished doctoral thesis

(6) Potchefstroom

(7) The aims of the investigation were to determine whether The Ability, Processing of Information and Learning Battery (APIL) is a good predictor of academic success of first year students; whether the APIL battery is a better predictor of the academic success of first year students than the General Scholastic Aptitude Test Senior (GSAT), and/or the Senior Aptitude Test (SAT), and to determine whether the APIL is a learning potential prediction instrument that provides fair assessment of the potential of learners from different cultures

- (8) Two samples were drawn from two different universities: Potchefstroom (Potch) n=656 and Rand Afrikaans University (RAU)¹ n=1200, Total N=1856
- (9) Successful (passed and partially passed) and unsuccessful groups
- (10) Within-groups designs (i.e. within each group from both universities)
- (11) First year university students from a variety of faculties
- (12) Not stated thus approximately 18 years
- (13) APIL, GSAT and SAT
- (14) As per the APIL manual (see Nel, 1997)
- (15) A one shot cross sectional design of first year students at two universities
- (16) All first year students at Potch and RAU, thus an availability sample
- (17) Immediate
- (18) Hypotheses as such not stated but rather in the form of questions: is the APIL as a learning potential test a good predictor of academic success for first year students? Is the APIL a better predictor of academic success than the GSAT and SAT? Finally, is the APIL a culturally fair assessment instrument?
- (19) As delineated within the APIL manual
- (20) For the Potch group no GSAT was used but the following is given for dispersion of marks for the total number of first year students. Order of variables: mark, number, sd, average - mark 1 n=683, sd 13.5 and average 59.56; mark2 n=658, sd 14.38 and average 58.08; mark 3 n=685, sd 14.11 and average 58.12; SAT 1 n=685, sd 4.34 and average 19.37; SAT 2 n=685, sd 6.005 and average 18.13; SAT 3 n=685, sd 6.748 and average 19.641; SAT 4 n=685, sd 3.819 and average 19.998; SAT 5 n=685, sd 6.026 and average 18.441; SAT 6 n=685, sd 6.025 and average 19.34; SAT 7 n=685, sd 6.380 and average 21.45; SAT 8 n=685, sd 5.951 and average 19.85; SAT 9 n=685, sd 4.403 and average 13.272; SAT 10 n=685, sd 6.149 and average 24.75; APIL (CFT) n=668, sd 4.969 and average 21.4; APIL (cl1) n=668, sd 10.831 and average 42.59; APIL (cl2) n=668, sd 21.78 and average 65.714; APIL (cl3) n=667, sd 31.94 and average 86.782; APIL (cl4) n=660, sd 39.884 and average 108.96; APIL (memory) n=668, sd 4.488 and average 17.617.

For the RAU group the GSAT was used and the following is given for dispersion of marks for the total number of first year students. Order of variables: mark, number, sd, average - mark 1 n=1200, sd 13.186 and average 56.45; mark2 n=1198, sd 13.71 and average 56.255; mark 3 n=1215, sd 13.18 and average 55.84; SAT 1 n=1217, sd 3.907 and average 18.871; SAT 2 n=1217, sd 6.739 and average 18.752; SAT 3 n=1217, sd 5.513 and average 19.075; SAT 4 n=1217, sd 4.450 and average 22.692; SAT 5 n=1217, sd 5.939 and average 20.512; SAT 6 n=1217, sd 5.866 and average 20.09; SAT 7 n=1217, sd 6.713 and average 18.48; SAT 8 n=1217, sd 5.698 and average 19.453; SAT 9 n=1195, sd 3.765 and average 13.437; SAT 10 n=1195, sd 4.365 and average 26.036; APIL (CFT) n=1217, sd 5.012 and average 22.855; APIL (cl1) n=1217, sd 10.312 and average 47.439; APIL (cl2) n=1217, sd 21.558 and average 72.901; APIL (cl3) n=1217, sd 31.124 and average 96.873; APIL (cl4) n=1217, sd 38.473 and average 118.632; APIL (memory) n=1217, sd 3.941 and average 18.525. GSAT 1 n=1217, sd 2.949 and average 21.0417; GSAT 2 n=1217, sd 3.596 and average 20.259; GSAT 3 n=1217, sd 3.293 and average 20.526; GSAT 4 n=1217, sd 3.074 and average 19.979; GSAT 5 n=1217, sd 2.916 and average 21.691; GSAT 6 n=1217, sd 3.419 and average 19.632; non verbal IQ n=1217, sd 13.698 and average 112.163; verbal IQ n=1217, sd 12.350 and average 110.715; total IQ n=1217, sd 12.905 and average 112.410.

Correlations between the predictor variables and the variety of criterion variables (*=average correlation; ** substantial) (different faculties) for the Potch group:

BA and BA languages n=111 for sat and 107 for APIL industrial communication n=53 for SAT and 52 for APIL; BCom and BCom law n=214 for SAT and 211 for APIL; BSc n=46 for SAT and 44 for APIL; BSc biology n=85; BSc maths n=32; engineering n=63 and HED n=81. Order of variables: BA/BA languages; b industrial, b com/b com law, BSc, BSc biology, BSc maths, b engineering and HED: SAT 1: 0.133, 0.37, 0.169, 0.0059, SAT 2: 0.079, 0.079, 0.211, 0.108, SAT 3: 0.102, 0.230, 0.119, 0.243, SAT 4: 0.150, 0.138, 0.201, -0.155, SAT 5: 0.159, 0.065, 0.249, -0.225, SAT 6: 0.121, 0.133, 0.269, -0.114, SAT 7: -0.019, -0.094, 0.018, -0.120, SAT 8: -0.026, 0.022, 0.077, -0.221, SAT 9: 0.105, 0.285, 0.193, 0.105; SAT 10: -0.012, 0.135, 0.18, 0.023, APIL (CFT), 0.154, 0.011, 0.267, -0.024, APIL (cl1) 0.175, 0.135, 0.154, 0.002, APIL (cl2) 0.146, 0.199, 0.228, -0.116, APIL (cl3) 0.179, 0.143, 0.237, -0.096, APIL (cl4) 0.096, 0.032, 0.23, -0.060, APIL (memory) -0.000, 0.0003, 0.084, -0.055. continuation: BSc biology n=85 for SAT and 84 for APIL, BSc maths n=32 for SAT and 31 for APIL, b engineering n=63 for SAT and n=62 for APIL, HED n=81 for SAT and 77 for APIL

order of variables: BSc biology, BSc maths, b engineering and HED: SAT 1 0.327, 0.573*, 0.382, 0.422*, SAT 2: 0.300, 0.319, 0.325, 0.116, SAT 3: 0.354, 0.452*, 0.298, 0.261, SAT 4: 0.178, 0.357, 0.525*, 0.419*, SAT 5: 0.177, 0.156, 0.348, 0.285, SAT 6: 0.149, 0.332, 0.463*, 0.353, SAT 7: 0.218, 0.179, 0.398, 0.226, SAT 8: 0.133, 0.224, 0.268, 0.159, SAT 9: 0.286, 0.425*, 0.256, 0.487*, SAT 10: 0.248, 0.372, 0.152, 0.333; APIL (cft) 0.214, 0.082, 0.290, 0.337, APIL (cl1) 0.147, -0.048, 0.490*, 0.371, APIL (cl2) 0.360, 0.193, 0.425*, 0.255, APIL (cl3) 0.383, 0.186, 0.477*, 0.232, APIL (cl4) 0.325, 0.247, 0.490*, 0.179, APIL (memory) 0.182, 0.273, 0.389, 0.136

correlations for the total RAU group per faculty order of variables: ba/ba lang n=259 for SAT and 259 for APIL and 259 for GSAT, law n=81 for SAT, 81 for APIL and 81 for GSAT, b com/b com law n=525 for SAT, 525 for APIL and 525 for GSAT, optometry n=38 for SAT, n=38 for APIL and n=38 for GSAT, BSc biology n=45 for SAT, n=45 for APIL and n=45 for GSAT order of variables: ba/ba lang, law, b com/b com law, optometry, BSc biology: SAT 1 0.243, 0.288, 0.191, 0.364, 0.212, SAT 2: 0.215, 0.064, 0.145, -0.225, 0.027, SAT 3 0.306, 0.133, 0.134, 0.221, 0.168, SAT 4: 0.165, 0.080, 0.077, -0.059, 0.462*, SAT 5: 0.164, -0.123, 0.145, 0.456*, 0.292, SAT 6: 0.148, 0.030, 0.104, 0.172, 0.161, SAT 7 0.059, -0.066, 0.028, -0.110, 0.138, SAT 8: 0.170, -0.029, -0.017, 0.003, 0.001, SAT 8: 0.170, -0.029, -0.017, 0.003, 0.001 SAT 9: 0.192, 0.085, 0.049, 0.552*, 0.220, SAT 10: 0.215, 0.222, 0.112, 0.465*, 0.331, SAT 10: 0.215, 0.222, 0.112, 0.465*, 0.331, APIL (cft) 0.137, 0.034, 0.149, 0.221, 0.433*, APIL (cl1) 0.156, 0.150, 0.108, 0.104, 0.320, APIL (cl2) 0.134, 0.136, 0.134, 0.282, 0.276, APIL (cl3) 0.136, 0.129, 0.140, 0.164, 0.223 APIL (cl4) 0.178, 0.079, 0.198, 0.256, 0.373, APIL (memory) 0.233, 0.022, 0.153, 0.515*, 0.453* GSAT 1 0.258, 0.052, 0.175, 0.524*, 0.142, GSAT 2: 0.246, 0.154, 0.170, 0.314, 0.272, GSAT 3: 0.229, 0.227, 0.257, 0.560*, 0.216, GSAT 4: 0.189, 0.008, 0.132, 0.006, 0.256, GSAT 5: 0.325, 0.114, 0.202, 0.458*, 0.138, GSAT 6: 0.238, 0.006, 0.129, 0.180, 0.121, non verbal IQ: 0.220, 0.026, 0.162, 0.152, 0.387, verbal IQ: 0.315, 0.195, 0.293, 0.581*, 0.374, total IQ: 0.293, 0.122, 0.252, 0.370, 0.416*. BSc maths n=105 for sat, n=106 for APIL n=106 for GSAT, b engineering n=119 for SAT, n=119 for APIL and n=119 for GSAT, nursing n=22 for SAT, n=22 for APIL and n=22 for GSAT, social work n=23 for sat, n=23 for APIL and n=23 for GSAT. order of variables BSc maths, b engineering, nursing, social work: SAT 1 - 0.058, 0.068, 0.665**, 0.623** SAT2: 0.138, 0.097, 0.284, 0.632**, SAT 3: 0.045, 0.208, 0.190, 0.474*, SAT 4: 0.301, 0.007, 0.139, 0.542* SAT 5: 0.093, 0.179, 0.184, 0.552*, SAT 6: 0.018, 0.3, 0.4*, 0.677** SAT 7: -0.012, 0.09, 0.194, 0.734**, SAT 8: 0.128, 0.254, 0.419*, 0.580*, SAT 9: -0.031, 0.184, 0.635**, 0.372, SAT 10: 0.00, 0.214, 0.020, 0.451*, APIL (cft) 0.121, 0.277, 0.337, 0.448* APIL (cl1) 0.136, 0.005, 0.249, 0.513* APIL (cl2) 0.009, 0.218, 0.168, 0.405* APIL (cl3) -0.025, 0.205, 0.151, 0.304, APIL (cl4) -0.012, 0.161, 0.115, 0.434*, APIL (memory) - 0.033, 0.116, 0.259, 0.385 GSAT 1 0.136, 0.146, 0.323, 0.774** GSAT 2: 0.106, 0.102, 0.603**, 0.678**, GSAT 3: 0.055, 0.139, 0.408*

¹ This university is now known formally as The University of Johannesburg but was at the time of this primary study known as RAU.



0.694**, GSAT 4: -0.023, 0.099, 0.346, 0.624**, GSAT 5: 0.106, 0.106, 0.450*, 0.652** GSAT 6: -0.021, 0.316, 0.484*, 0.703*, non verbal IQ: 0.034, 0.213, 0.490*, 0.701**, verbal IQ: 0.154, 0.171, 0.339, 0.721**, total IQ: 0.0699, 0.239, 0.450*, 0.728**

Only significant results for a variety of regression analyses will be given:

POTCH sample: r squared = 0.15 delivered by APIL (cft) for the black/Coloured group and the combination between the APIL and SAT delivered the following accounted for variance: 25.4% for the black/Coloured group, 17% b com, 21% BSc maths, 56% engineering, and 34.2% HED

RAU sample: the cft (APIL) r squared 0.054 explained variance for the b engineering. combined variance explained by APIL, SAT and GSAT for ba/ba languages 13.6%, b com/b com law 14.4%, BSc biology 49.9%, b sc maths 14% b eng 21% optometry 58.3%. There are many and varied statistics conducted within Engelbrecht's primary study, many of which are not pertinent to dynamic assessment per se. Only the above have to do with the predictive validity of the APIL

(21) The results of the investigation indicate that the various subtests of the APIL do not contribute significantly to the declared variance in the academic achievement of first year students. There is so much unexplained variance that the combination of these variables should only be used with great circumspection in predicting the academic achievement of first year students at Potch and RAU. Furthermore, the APIL cannot be considered as a fair assessment of potential of learners from different cultures. On the basis of the results of the regressions, there are no significant tendencies in terms of prediction of academic success of the two universities. In sum the combined APIL, SAT and GSAT account for 17% of variance for the Potch BCom, 56.2% for the Potch BSc maths, and 13.6% RAU BA/BA languages and 58.3% optometry

(22) Criterion variable: first year university results and predictors: GSAT, APIL and SAT

(23) Only cognitive tools were made use of whereas non-cognitive assessment may have added more comprehensive findings to the study

(1) Van Eeden, R., De Beer, M. & Coetzee, C.H.

(2) 2001

(3) Cognitive ability, learning potential and personality traits as predictors of academic achievement by engineering and other science and technology students

(4) South African Journal of Higher Education (University of South Africa)

(5) VOL 15 (1), 171-179

(6) Pretoria

(7) The aim of the study was to evaluate a battery of tests to be used as part of the process of selecting first year disadvantaged students for engineering and other science and technology courses at a Technikon in Natal and included in this battery is a dynamic assessment tool

(8) 224

(9) 1

(10) Within groups design

(11) First year students at a Technikon; majority were black and/or Indian with 58% male and 42% female

(12) Average age = 20

(13) The Learning Potential Computerised Adaptive Test (LPCAT), General Scholastic Aptitude Test Senior (GSAT), Senior Aptitude Test (SAT), Sixteen Personality Factor Questionnaire (16PF)

(14) As per the LPCAT manual. A computerised form of mediation (no human mediation)

(15) One sample group, as this study was merely evaluating certain tests for the selection of students

(16) First year Technikon students

(17) Immediate

(18) No hypothesis as such; merely establishing whether a set of tests will predict academic success

(19) LPCAT mediation: computerised adaptive testing using item response theory (IRT)

(20) Correlations between predictor variables with first year performance (**= sig 0.01 & *=sig 0.05): LPCAT pretest and first year results (n=119) = 0.18; LPCAT posttest and first year results (n=119) = 0.13; GSAT verbal and first year results (n=119) = 0.30**, GSAT nonverbal and first year results (n=119) = 0.14; GSAT full and first year results (n=119) = 0.24**, SAT calculations and first year results (n=145) = 0.21*; SAT spatial 3D and first year results (n=151) = 0.18*; SAT mechanical insight and first year results (n=153) = 0.10; maths and first year results (n=144) = 0.25**, science and first year results (n=140) = 0.28**, English and first year results (n=149) = 0.33**.

(21) The posttest scores of the LPCAT did not seem to predict academic success; school achievement was the best cognitive predictor of average first year performance

(22) Predictor variables included school marks for science, English and maths as well as SAT, GSAT and LPCAT scores. Criterion variables included: Technikon year marks and exam results of students

(23) Possible restriction of range due to the nature of volunteer students

(1) Lloyd, F. & Pidgeon, D.A**.

(2) 1961

(3) An investigation into the effects of coaching on non-verbal test material with European, Indian and African children

(4) British Journal of Educational Psychology

(5) VOL 31(2), 145-151

(6) Conducted in South Africa - Natal

(7) To compare the performance of children from different cultural groups on non-verbal tests, half the children were coached and the other half were not

(8) 900 (300 each for Indian, black and white)

(9) 2

(10) Mixed design, both within and between-groups design

(11) Primary school children - heterogeneous mix

(12) Age ranged between 10.6 - 12.6

(13) Non-verbal tests 1 and 2 of the National Foundation for Educational Research

(14) Coaching was used to make sure the children understood what was expected of them

(15) Pretest, mediation and posttest design. As two non-verbal tests were used, the experimental and control groups were themselves divided into two groups in order to ensure that the order of tests did not make a significant difference

(16) School children. The authors claim that both the Indian and European groups were fairly representative of a good cross-section of the socio-economic spectrum whereas the African group was more homogenous in terms of this spread

(17) Two sessions of coaching were included: one week between writing the pretest and then coaching (coaching lasted 30 minutes) and then a week later a second session of coaching before the posttest



(18) No hypothesis was stated but the emphasis was placed on investigating the differences between three groupings of children from different backgrounds

(19) Coaching sessions of 30 minutes to see if what was asked was understood

(20) It must be kept in mind that this study was conducted in 1961 and as such complete reporting of figures is not always satisfactorily delineated. Note also that group labels are those used by the authors in the primary paper

Means and sd's of initial and final test scores for experimental and control groups:

EUROPEANS - experimental group: (n=143) pretest 105.06(sd 10.45) and posttest 115.66(sd 12.85) control group: (n=133) pretest 101.34 (Sd 12) and posttest 108.73 (sd 12.6), experimental minus control group pretest = 3.72 and experimental minus control group posttest = 6.93

AFRICANS - experimental group: (n=136) pretest 88.75 (sd 8.15) and posttest 103.30 (sd 9.65) control group: (n=139) pretest 84.75 (Sd 6.80) and posttest 91.70 (sd 9.9), experimental minus control group pretest = 4 and experimental minus control group posttest = 11.6

INDIANS - experimental group: (n=133) pretest 86.40 (sd 7.45) and posttest 92.5 (sd 11.45) control group: (n=133) pretest 87.3 (Sd 10.7) and posttest 92.95 (sd 11.15), experimental minus control group pretest = -0.9 and experimental minus control group posttest = -0.45.

Gains made by the different groups:

EUROPEAN: experimental gain of 10.60 and control gain of 7.39 net gain (experimental - control) = 3.21*(sig p=0.05)

AFRICAN: experimental gain of 14.55 and control gain of 6.95 net gain (experimental - control) = 7.60* (sig p=0.05)

INDIAN: experimental gain of 6.10 and control gain of 5.65 net gain (experimental - control) = 0.45.

(21) Practice effect was similar in all experimental groups, but coaching differs significantly for each group. The superiority of the African children over the Europeans in responding to coaching together with the inability of the Indians to gain any benefit at all demonstrates clearly that the non verbal tests are far from being culture free

(22) Independent variable: performance on the pretest, dependent variable: performance on the posttest

(23) In some instances no reporting of t-test values for significant findings

(1)Skuy, M., Zolezzi, S., Mentis, M., Fridjhon, P.& Cockroft, K.

(2) 1996

(3) Selection of advantaged and disadvantaged South African students for university admission

(4) South African Journal of Higher Education (University of the Witwatersrand)

(5) VOL 10(1), 110-118

(6) Johannesburg

(7) This study aimed to explore the relative value of various predictors for advantaged and disadvantaged students

(8) 26

(9) One

(10) Within groups design

(11) Students enrolled in the pre-university bursary scheme (PBS) at the University of the Witwatersrand in 1991 within the faculty of Commerce. 54% male and 46% female. Advantaged students = 8 and disadvantaged students = 18

(12) Age ranged between 17-25 mean 21.9

(13) Static measures included the Biographical questionnaire (BQ), Interview measure (IM), mental alertness test (MAT), pattern relations test - traditional (PRT/T) and matriculation marks (MATRIC). Dynamic measures included the Pattern relations test-enriched (PRT/E), learning process measure (LSP), study process questionnaire (SPQ) and the learning and studies strategies inventory (LASSI). Criterion measures included the results of the mid-year university exams in accounting, maths, stats and business

(14) Feuersteinian mediated learning experience (MLE)

(15) Group administered

(16) First year university students both advantaged and disadvantaged

(17) Immediate (but the whole testing scenario took two days)

(18) None as such - exploring the value of new potential predictors

(19) Interaction between tester and testee and subject/learner which revolves around the meaning, applications and cognitive strategies underlying the tasks and carried out in terms of the principles of MLE

(20) Correlations of predictors and criterion variables for whole sample

(* P<0.05; ** P<0.005) and N=26: LSP-Maths 0.56**; LSP - JUNE 0.40*; LASSI-STATS 0.45*, SPQ (4) - BUSINESS STUDIES 0.42*, SPQ (b8) - business studies 0.43*, PRT/T - accounting 0.38*.

Correlations between predictors and criterion variables for the advantaged students only

(* p<0.01, ** p<0.01(??) ***p<0.005, **** p<0.0001) n=8, : IQ (MAT- BUSINESS 0.92****; LSP- maths 0.90***; MAT (Surface approach b7) - stats 0.95****; MAT (surface approach b7) - June average 0.88****; MAT (surface motive b1) -business studies 0.86***, IQ (Mat) -June average 0.75*; BQ - stats 0.75*, LSP- June average 0.72*; MAT (attitude 1) -business studies 0.81**, MAT (attitude 1) - accounting 0.76*; MAT (attitude 1) - June average 0.80**, MAT (surface approach b7) -accounting 0.82***; MAT (surface approach b7) - business studies 0.76* . The only significant correlation for the disadvantaged students was one subscale of the SPQ

(21) As expected the matric results did not predict for either the advantaged nor disadvantaged students, process measures also did not predict well (there was no correlation between PRT/E and criterion variables)

(22) Predictor variables (independent) static and dynamic tests - static: biographical questionnaire (BQ), Interview measure (IM), mental alertness test (MAT), pattern relations test - traditional (PRT/T) and matriculations marks (MATRIC). Dynamic measures - Pattern relations test - enriched (PRT/E), learning process measure (LSP), study process questionnaire (SPQ) and the learning and studies strategies inventory (LASSI) criterion variables: results of the mid-year university exams in accounting, maths, stats and business

(23) Small sample size restricted any generalisability of the results

(1)Lopes, A. , Roodt, G. & Mauer, R.

(2) 2001

(3) The predictive validity of the APIL-B in a financial institution

(4) Journal of industrial psychology



- (5) VOL 27(1), 61-69
- (6) Pretoria
- (7) The purpose of this study was to assess the predictive validity of the APIL test battery in a financial institution
- (8) 235
- (9) 1 (but further sub-divided into four)
- (10) One sample group
- (11) Job applicants' education ranged from standard seven (grade nine) - postgraduate
- (12) Age ranged between 16-58
- (13) The Ability, Processing of Information and Learning Battery (APIL-B)
- (14) Within the test as prescribed by the APIL-B manual. The authors do not explicitly state the type and manner of mediation used in their study but it is assumed that mediation takes place as this is a dynamic assessment instrument which used the posttest and pretest scores as indications of learning potential. Repeated exposure and instruction formed part of the administration of the dynamic sub-tests within the APIL-B, namely the Curve of Learning total (COL TOT) and Curve of Learning Difference (COL DIFF) and Memory (MEM)
- (15) Group administered
- (16) Successful job applicants at a large insurance company. 72 male and 162 female
- (17) Immediate
- (18) The APIL-B should be able to provide a more fair and accurate prediction of success on the job
- (19) Repeated exposure and instruction
- (20) Raw data from the six subtests of the APIL-B were available for a final sample of 235 subjects. The standard deviation and means of these raw scores were calculated and converted to z scores to facilitate comparisons. Subtests used within the APIL-B: curve of learning score (COL) is divided into two scores namely COL tot and COL diff and they are assigned half a weight, CFT (concept formation task), speed, accuracy and flexibility are static scores whilst COL and memory and knowledge transfer (KTT) are dynamic scores.

Means and sd's for ethnic groups on predictor scores (using z scores):

AFRICAN: CFT mean -0.88 (sd0.92), speed -1.13(sd0.84), accuracy -0.41 (sd0.37), flexibility -0.93(0.67), COL tot -0.52(0.32), COL diff -0.43 (0.29), memory -1.06(0.89)

INDIAN: CFT mean -0.05 (sd0.96), speed 0.09(sd0.81), accuracy 0.04 (sd0.38), flex -0.02(0.78), COL tot -0.03(0.38), COL diff -0.01 (0.45), memory (20)0.11(0.86)

COLOURED: CFT mean -0.33 (sd0.80), speed -0.3(sd0.78), accuracy -0.11 (sd0.28), flexibility -0.39(0.73), COL tot -0.16(0.34), COL diff -0.16 (0.37), memory -0.15(0.77)

WHITE: CFT mean 0.45 (sd0.82), speed 0.53(sd0.75), accuracy 0.19 (sd0.33), flexibility 0.52(0.93), COL tot 0.28(0.45), COL diff 0.23 (0.49), memory 0.44(0.82)

The predictive validity of the test was assessed by using a canonical discriminant analysis procedure and tests of equality of group means, however only the classification results are included here:

Original rating count (1-5) and predicted group membership, ratings 1-3 referred to as "poor to average" and ratings 4-5 referred to as "good to excellent". 66.7% of sample correctly classified in rating 1; 36.4% correctly classified in rating 2; 21.2% correctly classified in rating 3; 39.1% correctly classified in rating 4 and 62.9% correctly classified in rating 5. In sum, 36.6% of cases were correctly classified

(21) Although the APIL-B minimises any role that cultural variables play, the black group still scored lower in comparison to other groups. What has been shown is that despite concerns relating to the reliability of the criterion, the APIL-B is able to predict performance of employees in a financial institution

(22) Criterion variable included the manager ratings on a five point scale; the predictor variable: APIL-B

(23) If white and black testees were matched for language skill differences on the cognitive test may have disappeared

(1)Nunns, C. & Ortlepp, K.

(2) 1994

(3) Exploring predictors of academic success in psychology 1 at WITS university as an important component of fair student selection

(4) South African journal of psychology

(5) VOL 24 (4), 201-208

(6) Johannesburg

(7) The aim of the study is to explore empirical predictors of students' performance in the psychology 1 course at the university of the Witwatersrand and is comprised of two studies: (1) an archival study and (2) a predictive validity study involving the Arts faculty rating programme

(8) Sample size for first archival study = 1101 of whom advantaged n= 1048 and disadvantaged n=53. For second study: N=133 of whom advantaged n = 107 and disadvantaged students n=26

(9) 2 groups each in both studies

(10) Between groups design

(11) First year university students who came from educationally advantaged as well as educationally disadvantaged schooling backgrounds

(12) For first study not stated therefore approximately 18. For second study, the advantaged students average age 18.25 and disadvantaged students average age 23.6.

(13) In the second predictive study: the Conceptual reasoning test (CRT); mental alertness test (MA), reading comprehension (RC) was used

(14) Incorporated within the CRT but the learning potential measure used within the test if not explicitly stated in the primary study

(15) Two separate studies: (1) an archival study in which information is gathered on first year arts faculty students registered for psychology 1 over five years (from the university records), this period was chosen to attempt to establish whether past academic performance would predict subsequent academic performance and (2) predictive study assessing the validity of the arts faculty ratings as well as various psychometric instruments

(16) First year advantaged and disadvantaged students

(17) Immediate

(18) Investigating new predictors of academic success making use of a dynamic assessment instrument

(19) The CRT is said to assess potential which is presumably incorporated into the test

(20) Archival study:



The arts faculty rating, English, Afrikaans, biology, science, history and accounting correlated significantly (at the 0.01 level) with psychology 1 across all five years. But for the disadvantaged group, only the arts faculty ratings programme was significant (also at 0.01)

Predictive study: correlations were conducted for the faculty ratings and other test scores with psychology 1 scores. For the advantaged group the arts faculty rating correlated ($r=0.58$ at $p<0.0001$ with psychology 1. The CRT also sig 0.44 at $p<0.0001$, also MA 0.37 at $p<0.000$; RC 0.31 at $p<0.001$. However for the disadvantaged group: only the CRT was sig at 0.48 $p<0.01$

(21) The CRT at least predicts for the disadvantaged group in terms of psychology 1 performance and is significant as the test assesses potential and this measure is also independent of matric results. The more traditional marks (matric and university marks) predict for the advantaged group

(22) For this archival study data comprised biographical information, matric results, matric average and university academic record

(23) Small sample sizes used for the disadvantaged group

(1) Shochet, I.M.

(2) 1994

(3) The moderator effect of cognitive modifiability on a traditional undergraduate admissions test for disadvantaged black students in South Africa

(4) South African journal of psychology

(5) VOL 24(4), 208-215

(6) Johannesburg

(7) In this study, a measure of students' cognitive modifiability, assessed by means of an interactive assessment model was added as a moderator of the traditional intellectual assessment in predicting first year university success

(8) 52

(9) 1 (further subdivided into two)

(10) Both within and between-groups design

(11) First year university students

(12) Not stated approximately 18 years

(13) The traditional intellectual predictor: Deductive reasoning test used in static and dynamic form (DRT); moderator variable - cognitive modifiability (CM): to assess modifiability students were administered two conventional cognitive measures which yielded their baseline functioning, after which coaching was given followed by another test. Gain scores made after the mediation were added together to yield a cumulative gain score known as cognitive modifiability. Pattern relations tests (PRT) used in traditional and dynamic form

(14) Coaching on tasks which mediated the necessary cognitive strategies needed to complete the tasks

(15) Group administered

(16) First year university students

(17) 20 minute break between baseline measure and mediation on both the DRT and PRT

(18) Learning potential tests will yield better predictors of academic success for disadvantaged students

(19) As prescribed by Feuerstein

(20) t test for paired differences for dependent samples used to compare the difference in scores between the traditional and dynamic administrations for both the DRT and the PRT. The DRT had a gain of 4.23 ($t=6.62$) and PRT 5.19 ($t=7.78$), thus the mean cognitive modifiability scores was 9.42.

Hierarchical moderated multiple regression to determine if cognitive modifiability (CM) moderates DRT for disadvantaged students ($N=49$). Step 1: criteria are credits: predictor variable entered is DRT: beta 0.25, t 1.77, sig t 0.0820, multiple r 0.25, $f=2.02$ sig f 0.0820; and average: predictor entered is DRT: beta 0.26, t 1.85, sig t 0.0702, multiple r 0.26, $f=3.43$ sig f 0.0702. step 2: adding CM: criteria are credits: predictor variable entered is DRT: beta 0.28, t 1.93 sig t 0.0594, cm: beta -0.16, t -1.13, sig t 0.2653 multiple r 0.30, $f=2.22$ sig f 0.1197; f change step 1 to step 2 = 1.27 sig f change 0.2653. average: predictor DRT: beta 0.29, t 2.07 sig t 0.0402* ($P<0.05$), cm: beta -0.20, t -1.43, sig t 0.1600 multiple r 0.33, $f=2.77$ sig f 0.0729; f change step 1 to step 2 = 2.04 sig f change 0.1600. step 3: adding CM X DRT: criteria are credits: predictor variable entered is DRT: beta 0.77, t 3.34 sig t 0.0017* cm: beta 0.70, t 1.98, sig t 0.0539 CM X DRT beta -1.10, t -2.63, sig t 0.0115* multiple r 0.46, $f=3.99$ sig f 0.0133*; f change step 2 to step 3 = 6.93 sig f change 0.0115*. average: predictor DRT: beta 0.62, t 2.65 sig t 0.0112* ($P<0.05$), cm: beta 0.38, t 1.05, sig t 0.2981; CM X DRT beta -0.75, t -1.74, sig t 0.0889 multiple r 0.40, $f=2.94$ sig f 0.0432*; f change step 2 to step 3 = 3.02 sig f change 0.0889.

Correlations between DRT and the criteria for the more and less modifiable disadvantaged students sub-grouped on the CM score: less modifiable ($n=26$) 0.59** (0.01) for DRT X credits and 0.61** (0.01) for DRT X average; more modifiable ($n=23$) 0.00 and 0.02 for the above (note less modifiable means that higher baselines were achieved)

(21) Cognitive modifiability significantly moderated the predictive validity of the traditional intellectual assessment for the sample i.e. the higher the level of modifiability, the less effective were traditional measures for predicting academic success. Modifiability serves as a moderator between the DRT and academic success for disadvantaged students. Less modifiable students (those with higher baseline scores) were shown to have significant correlations with the static tests as opposed to the more modifiable students (those with lower baseline scores)

(22) Criterion measures: end year academic results - namely (1) number of credits obtained ranging from 0-4 and average percentage achieved at the end of the first year. Predictor variables: both the static and dynamic forms of the DRT and PRT

(23) Small sample size used which limited generalisability

Appendix 2 Content analysis of returned questionnaires

1. Introduction

Appendix 2 details the content analysis of eleven completed and useable questionnaires that were originally emailed to one hundred dynamic assessment researchers and practitioners across the globe. Follow-up emails were sent via internet four months later to invite further responses as well as to invite responses from those who had initially accepted participation. The covering letter is shown followed by the questionnaire that was sent. A brief introduction to content analysis is given which fleetingly focuses on issues of reliability and validity, qualitative and quantitative content analysis as well as manifest and latent meaning within analysed text. The method of content then follows and details the coding of the content, the process involved and the coding frame utilised. The next section looks at the results of the analysis by summarising the responses and extracting themes and then coding the extracted themes followed by a brief excursion into what was not mentioned in the responses as further source of information. This appendix concludes with a discussion of the content analysis. The questionnaire which was sent to researchers and practitioners will now be discussed.

1.1 Questionnaire

One hundred questionnaires were emailed to dynamic assessment/learning potential practitioners across the globe. Contact details were obtained from published books and articles as well as web pages on the internet. Of the original one hundred respondents to whom the questionnaire was sent, thirty-one initial responses were received, of which eight practitioners kindly refrained from answering due mainly to two reasons: lack of time and unfamiliarity with the field (those who had made contributions to the field were no longer active within this area). Follow-up emails were sent to practitioners who had originally expressed interest in the study. Eleven completed questionnaires were returned and the final analysis was conducted on these eleven contributions. Of the eleven respondents two indicated their choice of anonymity. Seven respondents (over and above the eleven respondents who completed the questionnaire) indicated their interest in completing the questionnaire but these were unfortunately never received. As only eleven questionnaires were returned as well as the fact that most responses were quite short (less than five sentences on average per question) this technique was considered most suitable for a thematic analysis.

Of the hundred contacts to whom questionnaires were sent, only four were South African researchers, only one of whom initially responded to the email request for participation but failed to return a completed questionnaire. It is perhaps telling of South African dynamic assessment research then, that if this method of assessment is to find itself more firmly ensconced within the broader field of assessment, more care needs to be taken in theoretical issues surrounding the method especially within the South African context. Why is it that, during much of the research pertaining to dynamic assessment in South Africa, it is the international community which seems more interested in the goings-on in the local context or at least more willing to participate in research? A curiosity not evident from local researchers. It is conjectured that had the author formed more personal contacts with those on the list of potential contacts that a greater number of respondents would have participated but as 96% of the contacts emanated from overseas this proved difficult.

That the content analysis of the questionnaires appears as an addendum to this study attests to the small role this section plays within this larger theoretical study and is possibly an avenue for further research. Had the sample been larger it is assumed that a more balanced view would have been expressed and as the current analysis stands it may well be one-sided and thus biased. The eleven completed questionnaires were completed and returned within just over one month which in itself is informative of the nature of the sample. Perhaps the views expressed are similar due to the similar nature of the sample. These are of course mere speculations. No major conclusions can thus be drawn from this analysis due to the sample size, although on its own, salient themes do emerge. At times the answers to some questions did not really address the question as posed but due to the nature of the lengthy questions it is understandable that some answers might not have answered what was asked. In the analysis, answers given under each question are kept as such even though it may be construed as ill-fitting to the question posed. The covering letter and questionnaire is available below as are the results of the content analysis of the returned questionnaires which is preceded by a brief exposition of content analysis.



The covering letter was as follows:

Dear Professor/Dr

My name is Raegan Murphy and I am currently exploring a number of aspects within the field of dynamic assessment for my PhD, at the University of Pretoria, South Africa. However, it would not be complete without your views on certain aspects pertaining to a few issues. I realise time is of great importance to you, but I would be greatly honoured if you would kindly consider looking at the attached document. It would be exceedingly generous of you if you were to answer the few questions as you see fit as this will greatly enhance my discussion of the topic.

If you have any queries regarding the document (MS word 2000) or if you experience any difficulty in opening it, please do not hesitate to contact me. Further details are included in the document.

I trust the questions will not take up too much of your time.

In the event that I am unsure and although unlikely at this stage I might, if necessary, email you to seek further clarification to your answers.

Many thanks in anticipation for your assistance.

Sincerely

Ms Raegan Murphy

PO Box 27846

Sunnyside

0132

South Africa

raegan@mweb.co.za

+27 12 386 9298

Promoter Professor David J.F. Maree

Department of Psychology

Faculty of Humanities

University of Pretoria

Pretoria 0002

South Africa

djfmaree@postino.up.ac.za

+27 12 420 2916

<http://www.up.ac.za/beta/academic/humanities/eng/eng/psyc/eng/index.htm>



The questionnaire was as follows:

The theoretical underpinnings and fundamental assumptions within dynamic assessment

Good day and thank you for opening up this document! The following brief questionnaire seeks to understand more fully the underlying affinities that your particular theoretical rendition of dynamic assessment encapsulates within the broader field of intelligence and change assessment. As part of my theoretical doctoral study, philosophical underpinnings as well as a myriad other issues pertaining to this sub-field within the intelligence arena are of particular interest to me and although disagreements may be rife within certain areas in dynamic assessment, the aim of the questionnaire is not to further incite argumentation but to enlighten the field as to its current and future directions within the discipline of psychology. This undertaking is more modest than is presented here! I am not aiming to make major structural changes within the field, but to perhaps stress and highlight certain issues which in my opinion are either missing or over-emphasised.

Any and all written communication within this document will be strictly cited as personal communications as on the day of delivery and any possible error in interpretation will be attributed to me alone. I hope that you will find the questions quick and easy to answer even though such an area of investigation is inherently difficult to summarise and cognisance is taken of this. If you have any queries at all about answering these questions or feel that they should be otherwise stated, please do not hesitate to communicate this to me, Ms Raegan Murphy at: raegan@mweb.co.za , tel. +27 12 386 9298. My supervisor is Professor DJF Maree at the University of Pretoria, South Africa and Professor Maree can be contacted at david.maree@up.ac.za , tel. +27 12 420 2916. If you prefer to remain anonymous yet have your views aired, please indicate this with an X in your chosen box and your replies will be treated anonymously.

I would prefer to remain anonymous

I do not mind being cited by name



There are eight (8) questions in total, one per page, please scroll down for question 1. Details are given after question 8.

(a text box has been provided, if there is not enough space please merely expand the text box as you see fit and simply enter the text down as it suits you).¹

Kindly note that each question is preceded by a statement reflecting the focus of that particular question.

Question 1 - The status of your particular view of dynamic assessment within intelligence assessment

What would you consider to be the nature of your view of dynamic assessment? Do you consider it to be a conceptual scheme, model or theory? (taking cognisance of the differences inherent within these various modes of views). Some may consider their views as merely tentative schemes whereas others may perceive their ideas to be fully-fledged theoretical stances hence moulding and structuring their practical endeavours accordingly.

Question 2 - Philosophical bent or underpinning of your theoretical take on dynamic assessment

Deeply entrenched within any idea/model/theory or simply a view on life, are allied philosophical understandings of how things ought to work. Whether tacitly supporting this notion or taking a dislike to its deterministic way of looking at the world, makes no difference for the moment. On each of the following views or issues listed below, what are your ideas concerning dynamic assessment within the broader field of intelligence assessment? For instance, based on your views concerning mind/brain, it may be that in downplaying the role of neuroscience, one concentrates on the behavioural level only, thus hypothesising that behaviour is indeed malleable at a level not synonymous with neural architecture (or at any other level). This will of course play out in your fundamental beliefs and hence theory(ies) behind your views on dynamic assessment.

- *On nature and nurture*: From nativists, empiricists to selectionists (how the brain/mind develops along purely genetic lines to environmental impingements),
- *On mind/brain*: from Cartesian dualism to succinct mind/brain identity theory,
- *On agendas*: from pragmatic to Socratic ideas as to the role of your view and the resulting influences in practice and on the 'science' of the field²
- *On historicity*: the direct/indirect (or total lack thereof) impact of historicity within the current understandings within intelligence assessment (how varied and indeed colourful is dynamic assessment's historical and geographical vistas!)

Question 3 - Your developmental model followed or model most adhered to within the field of child development/educational development as well as adult growth and maturation

Which developmental model within child/adult development and maturation do you most closely follow when working within dynamic assessment?

For instance, biologically driven theories of development, environmentally aligned theories of development and theories which challenge both extremes by meeting midway are offered as tentative guides as to how you might want to answer this particular question. Any well-known theory may equally be appended with the prefix of "neo" seeing as older theories are or will have to be continually assessed in terms of their fundamentals (neo-Piagetian, neo-Vygotskian, neo-anything you think would be appropriate here). Any inclusive model be it ethologically driven, social-learning theory driven or information-processing driven as well as any hybrid theories on the horizon can be utilised to explain your thoughts.

Question 4 - The ensconcement of your theoretical take on dynamic assessment within the broader fields of intelligence assessment

Where do you think your view/theory of dynamic assessment should be placed within the broader framework of intelligence assessment? Factors to consider when answering this question include (but is not an exhaustive list, you may add more factors which you find important):

- The intelligence models you most closely follow when placing your dynamic assessment view within it
- Your views concerning dynamic assessment and intelligence; are they divorced from any such particular intelligence model or are they firmly embedded within two or more models?
- Your views concerning dynamic assessment and your chosen model(s) of intelligence: are they linked in anyway or do you perceive them to be ill-at-ease conceptually?

¹ The text boxes have been removed in order to conserve space in the thesis.

² I use 'science' here very cautiously, as I do not want to bog this particular question down in the philosophy surrounding what is and is not construed as science within the behavioural and social sciences, although this is addressed in the dissertation.



- Your views on dynamic assessment and how they fit in within the various competing views of intelligence. Where would you place your views? For instance, you might classify your views as nesting within a psychometric model itself housed within an intelligence model. The tenets inherent within a psychometric view as well as those inherent within your chosen model of intelligence will impinge on your view of dynamic assessment

Question 5 - The affinities your particular theory of dynamic assessment has with aligned fields of neuroscience, neuropsychology and computational intelligence

Current findings in the popular science and psychology literature as well as the increasing findings within academic literature (or at least the reporting of such findings) at times leads one to conclude that this new century could well be cited as the century of the “physical” (brain, genes, proteins etc).

Keeping this in mind, consider the following:

- How have these fields of enquiry been built into your view/model/theory of dynamic assessment within intelligence research?
- Do you think such findings should/should not play a role in further defining how your view should or should not be adapted? If yes, how do you think this should proceed?

Question 6 - The historical development of this sub-field of enquiry and its potential future within the realm of psychology (itself moving towards a more integrated field comprising natural science and behavioural sciences methodologies)

Having emanated from a natural philosophical background, allied to the natural sciences, finding favour with various movements within psychology through the century and having traversed a large field of enquiry, psychology and in particular intelligence (and dynamic assessment) is the proud bearer of a rich history, albeit a brief one.

Will dynamic assessment as a movement/model/theory simply die a death due to various factors or, will it in your opinion, forge ahead making strides unbeknownst to practitioners today?

Humans are not terribly successful in determining what will and will not make an impact, even though an impact may not be construed as such for a long time to come. On the other hand, pursuing avenues with no definitive profit in terms of theory development may hinder development in other realms with resources better spent in these other realms.

- What are your thoughts on this matter?

Question 7 - The quantitative imperative

The role of statistics within psychology has been questioned and even criticised (the APA's Task Force on Statistical Inference, 1996) and the works of Joel Michell and others give a voice to the critical philosophy of mathematics and measurement within the social sciences. Would this perhaps add fuel to the fire as far as your view on dynamic assessment within intelligence is concerned or would this add support and buffer your views in terms of how dynamic assessment and intelligence should in fact proceed?

Question 8 - The meta-theoretical solution or pie in the sky

Some regard meta-theories as too reductionistic and their practitioners as naïve in attempting to simplify too complex an area of research within intelligence research. What potential lies within such an endeavour for dynamic assessment in your opinion? Would it help to stabilise the field or merely contribute towards confusion?

Due thanks is hereby extended to you for your willingness to take the time to answer these questions! I understand the time limitations under which we all work these days. Once again, it is reiterated that any and all interpretation of your answers to my queries will be understood to be mine alone, although clarification may be sought if it is considered that such interpretations may have the potential to be misconstrued. As soon as the study is completed I will notify you via email as to the URL needed to access the document should you so wish.

Kind regards,
Ms Raegan Murphy
PO Box 27846
Sunnyside
0132
South Africa

1.2. Brief introduction to content analysis

In order to objectively analyse the narrative content within the questionnaires, content analysis as a chosen qualitative technique was decided upon, although it typically employs quantitative techniques as well (Carney, 1972; Holsti, 1969; Lindkvist, 1981; Weber, 1986). Various definitions have been delineated throughout the nineteenth and twentieth centuries of what exactly content analysis is. Classical content analysis was already being practised by German classicists in the nineteenth century albeit in a less rigorous fashion that is current today (Carney, 1972) and was utilised on Swedish material as far back as the eighteenth century (Krippendorff, 1980; Rosengren, 1981). It has as primary roots journalistic ventures into the analyses of newspapers (Payne & Payne, 2004). Interestingly enough Carney draws an apt analogy (seeing as this study deals with dynamic assessment and intelligence) by comparing what the first World War did for intelligence assessment to that of the influence of second World War on the use and development of content analysis – pertaining here to propaganda and its requisite use in military intelligence (Andr n, 1981; Carney, 1972; Krippendorff, 1980). Sociology, anthropology, psychology and communications studies later became users of this approach to textual analyses.

An all-round definition of content analysis which suits the purposes of this study's analysis can be stated as follows: a technique for drawing objective inferences from text³ by categorising latent and manifest themes into quantifiable sections or in this case themes (Berg, 2001; Carney, 1972; Krippendorff, 1980; Payne and Payne, 2004; Ryan & Bernard, 2000; Silverman, 2005; Weber, 1986; Whitley, 2002). Current content analysis techniques also allow for the analysis of latent themes (Graneheim & Lundman, 2004). The main rationale behind utilising content analysis for questionnaire responses, as Dumont and Frindte (2005) illustrated with their research, is to convert the raw data into categories of meaning (Henning, Van Rensburg & Smit, 2004). Latent or implicit themes may arise which necessitate inference from the analyser and as such this brief analysis will be mostly descriptive and only cursorily inferential towards the conclusion after all responses have been discussed. Graneheim and Lundman (2004) highlight the process when selecting text to code. Firstly, a unit of analysis has to be established followed by a meaning unit which is systematically condensed or attenuated so as to allow for coding. Once codes have been allocated, themes or content areas are derived from the aggregated codes. Hierarchically, sub-categories are placed in overarching main categories referring to the layering of themes (Creswell, 2002). The mode of operation will be thematic analysis as unit meaning (Carney, 1972; Marais & Bondesio, 1996), which, as Creswell (2002) states, is really just coding. Specific non-overlapping themes (dichotomised coding) were originally chosen for this study so as to make the task of counting responses easier (Ryan & Bernard, 2000; Silverman, 2000) as well as to help keep subjective inferences to a minimum (Graneheim & Lundman, 2004). The actual words used by respondents and the author's conceptualisation of the meanings may at times be given to debate as "meaning is inherently ambivalent and context dependent" (Henning, Van Rensburg & Smit, 2004, p.128; Strauss & Corbin, 1998). However, the focused nature of the questions within a niche area of research was predicated on common understandings of terms.

In addition to thematic analysis as primary meaning unit (Graneheim & Lundman, 2004), frequency counts will also be employed as secondary technique, which literally counts the number of times specific words are used. These counts serve only as descriptors of information more so than inference-bearing markers in the data and as such play less of a role in the analysis and interpretation of the findings (Berg, 2001; Holsti, 1969; Silverman, 2005; Weber, 1986). The basic assumption in content analysis is the relevancy of frequency of unit meanings as opposed to infrequent⁴ units of meaning (Lindkvist, 1981) and semantic analysis will also be a procedure utilised (Andr n, 1981; Berg, 2001). Classification will be conducted on the basis of meaning, how the words relate to the "overall sentiment of the sentence" (Berg, 2001, p.247). Counting can be misleading in terms of extrapolating meaning from frequency. These counts are not employed to cast a reductionistic blanket over the themes but to offer a more rounded quantifiable description of findings which will serve only to highlight main trends. The dichotomously phrased questions and their subsequent responses are tabulated in order to illustrate the number of agreements and disagreements that were made on certain issues.

Silverman (2005) highlights a negative aspect of categorisation by stating that although quantification via such categories is useful in providing a "powerful conceptual grid" (p.123), it can often be constraining as it may deflect away from categories that remain uncategorised. However, sorting texts into categories does assist in reducing heavy loads of information (Arredondo, Rosen, Rice, Perez & Tovar-Gamero, 2005; Creswell, 1998). Table 1 details verbatim text examples of how certain sections of text were analysed and coded for both dichotomous and thematic text. The meaning units or actual responses are condensed into statements and then coded or thematicised (Graneheim & Lundman, 2004; Henning, Van Rensburg & Smit, 2004; Maree & Maree, 2005). Categories are pre-formed as the questions themselves function as categories. Respondents had to answer within the confines of the questions and as such the content to be analysed is already known in general. The aim of this

³Need not of course be limited to textual analysis (Lindkvist, 1981; Neuman, 1997) but for the present circumstance this definition will suffice. Semiotic structure of meaning pertains to areas as diverse as music and architecture for instance and involves the signs applicable to the inherent meanings conveyed to the recipient of the message (Cobley & Jansz, 1999; Lindkvist, 1981).

⁴Carney (1972) highlights the fact that what is missing in texts can be just as informative as what is present.

analysis, then, was not to extract exploratory ideas but to establish the degree of support for each of the views expressed within each question via thematic analysis and can thus be classified as non-formal content analysis (Carney, 1972).

Appendix 2 Table 1 Example of selective verbatim response text coding for dichotomous text (question 1, 2, nature vs. nurture debate) and thematic text (question 3)

Meaning unit / actual response	Condensed meaning unit / statement	Code / theme (partially predetermined by the questions posed)
Question 1		
<ul style="list-style-type: none"> ▪ I consider DA to be more of a model than a theory ▪ I consider it to be a model ▪ I would consider my approach as a Paradigm ▪ It represents different paradigmatic views ▪ DA is a model ▪ It is mainly a measuring procedure 	DA is a model, paradigm or measuring procedure	Model Paradigm Measuring procedure
Question 2		
<ul style="list-style-type: none"> ▪ The environment plays an important role in the development of children. However the genes also play a role ▪ More toward nurture ▪ Nature-nurture arguments are of little interest for my work in dynamic assessment ▪ The limits of intellectual capacity are set by the “hardware” ▪ Nurture is unquestionably more important ▪ There is an interaction between nature and nurture ▪ The theory assumes a set of interactions almost genetically driven [and the] nature of that mediation is strongly influenced by linguistic and culturally based worldviews, values and interactions 	Nature and nurture both play roles. The distinction is moot. Nature is overriding. Nurture is the more important of the two. There is an interaction between the two	Nature Nurture Interaction Neither is important
Question 3		
<ul style="list-style-type: none"> ▪ I don't think that any specific theoretical model is sufficiently comprehensive ▪ Ecological view of intelligence ▪ Environmentally aligned theories of development ▪ Necessarily eclectic ▪ Bio-ecological model of Bronfenbrenner and Ceci (1994) ▪ A mixture of classical Vygotskyan theory, more modern representations of this work such as that of Kozulin, Haywood and Karpov, Grigorenko and Sternberg, information processing theory and intelligence theory ▪ Based primarily on Feuerstein's work as it informs Vygotsky's view of the zone of proximal development ▪ Barbara Rogoff's work on cultural models of development. 	None in particular. Ecological, environmental, bio-ecological models. Models emphasising contextual factors. Eclectic mixes of various models are employed. Modern renditions of older classical theories and models are utilised. Vygotsky, Feuerstein and Piaget are notable researchers within this area of concern. Mentions of other researchers are made such Rogoff, Kozulin, Haywood, Karpov as well as Sternberg and Grigorenko	Eclectic mixes Contextual models Old and new models along similar lines Vygotsky Feuerstein Piaget Other

1.2.1 Reliability and validity

Issues of reliability and validity play no less a role in qualitative research (Andrén, 1981; Bryder, 1981; Silverman, 2005) and terms such as credibility, dependability and transferability have been utilised within the qualitative domain to reflect similar meanings inherent in reliability and validity, for instance, trustworthiness (Graneheim & Lundman, 2004; Henning, Van Rensburg & Smit, 2004). An aspect of concern which directly impinges on the reliability of the coding of the text is the internal consistency with which the exercise is carried out (Weber, 1986; Whitely, 2002). The below-mentioned coding process employed fewer rather than more categories, had a broad focus (although specified questions), pre-defined response units, “after the fact” coding (coding which took place upon receipt of responses) and did not infer from text resulting in higher reliability of coding (Whitley, 2002).

Three types of reliability pertain to content analysis namely, stability, reproducibility and accuracy (Krippendorff, 1980; Weber, 1985). Stability is ensured in this study as only one coder (the author) was used and coding was invariant over time. However the same cannot be said of reproducibility as no other coder coded the text and thus no conclusions can be deduced which might portray similar findings across coders, although Andrén (1981) and Krippendorff (1980) point out that such coding reproducibility is not necessarily tenable when applied to semantic content analysis. Accuracy relies heavily on category reliability (Holsti, 1969). Four types of diagnostic devices are put forward by Krippendorff (1980, p.149) as part of his discussion on reliability-enhancing strategies within content analysis and consist of the following: unit reliability (unreliability in the material), individual reliability (coders used for the exercise), single-category reliability and conditional reliability (unreliability with the recording instructions). Holsti (1969) admits it is often the case that in order to improve upon reliability within content analysis, coders and categories are the dual aspects in need of attention and are subsequently the limits of improvement achievable. No standard codings (for instance coding dictionaries) were used for this study as it was deemed too small.

Validity issues also cannot be definitely ascertained as there exists no other known measure similar to this one and hence similar substantive conclusions will be needed in order to conclude that the analyses is a true and accurate picture of what was written. Carney (1972) adds that as much of the total context as possible, including communicator, message and audience needs to be explored in order for the message contained in the text to be accurately understood. However the simpler the strategy in terms of counts the less likelihood of invalid inferences as opposed to semantic content analysis (Andrén, 1981). Andrén (1981), Holsti (1969) and Krippendorff (1980) list a typology of validity-bearing concerns pertinent to content analysis. Validity concerns itself with data-related validity and encompasses semantic and sampling validity. Pragmatic or product-oriented validity includes correlational and predictive validity as well as process-oriented validity such as construct validity. Data-related validity assesses how well the method of analysis agrees with the information contained in the content and relies upon the sensitivity given to symbolic meanings in the text as well as sampling validity which reflects the degree of bias of the sample of texts chosen for the exercise. For the purposes of this study, the sampling is indeed biased towards practitioners of dynamic assessment and moreover is a reflection of researchers who champion the basic ideals of dynamic assessment. Pragmatic validity concerns itself with how well the chosen method of analysis “works” under various circumstances (Krippendorff, 1980). Construct validity looks at the degree of consensus between the process of analysis and the context from which the data derives.

1.2.2 Qualitative vs. quantitative content analysis

This study seeks to confine itself to a less rigorous and informal analysis of content aimed at analysing thematic content as well as word usage. Notwithstanding the efficacy of these more quantitative methods of analysis, Berg (2001) and Holsti (1969) caution the researcher about the danger inherent in arbitrarily limiting the content to that which is quantifiable only and so a blend of both quantitative and qualitative techniques is offered as guidance. The growth of content analysis from crude quantitative measures to a method including qualitative methods (Payne & Payne, 2004) is indicative of the wide field to which this technique can be applied. It need not be constrained by the quantitative imperative, even though this is often viewed as the more objective of the two (Holsti, 1969). Qualitative analysis sheds light on patterns of content (less positivistic in overtone) whereas quantitative analysis sheds light on the form and duration of such patterns (more reductionistic) and each gives to the results something which the other lacks (Berg, 2001).

Casual inferential modelling may also shed light on various relationships between and within themes and specifies the relationship between latent variables and brings to light causal effects of latent and observed variables (Weber, 1986). Originally frequency counts of word occurrences (Krippendorff, 1981) construed as quantitative would suffice as measure of the importance or lack thereof of certain issues within texts, however, the more qualitative avenue of analysis comes to the fore when the meanings behind these textual counts becomes manifest (and is considered as anti-quantitative) (Payne & Payne, 2004). This extrapolation between counts and interpretation is not yet an issue entirely resolved within content analysis (Berg, 2001) but is the blend of method utilised in this small study.



It is imperative that text, theory and content analysis results be related (Lindkvist; 1981; Weber, 1986) and it is for this reason that the content analysis is located at the end of this theoretical study which aided in its analysis, especially the analysis of content that was not mentioned (Holsti, 1969). The need for analysis of both manifest and latent meanings within the responses warrants both types of quantitative (frequency counts) and qualitative (thematic identification) research methods (Payne & Payne, 2004). Rosengren (1981) conceives of the differences between quantitative and qualitative analyses as the latter ranging from impressionistic, interpretative and intuitive accounts of the data to systematic analysis carried out on the nominal scale level. The former being representative of nominal scale measures which can then be aggregated at higher scale levels. A range of techniques spanning a range of application areas is thus reflective of this multitudinous method of content analysis. There is thus, as far as this author is concerned, no reason to avoid either approach as both, inherent in their separate methodologies, are able to extract information from text in different ways and in so doing compliment each other.

1.2.3 Manifest and latent meaning

An issue often debated within the methodology of content analysis is the limit of inference that should be drawn (Berg, 2001; Graneheim & Lundman, 2004; Holsti, 1969 Payne & Payne, 2004). This issue has bearing on interpretative meanings assigned to the units of analyses. Inferential descriptions beyond manifest content must be supported and validated (Krippendorff, 1981). This has been partially solved by theory-guidance of responses (Whitley, 2002). Weber (1986) advises researchers to validate quantitative findings by revisiting the text in order to highlight hypotheses originally stated.

2. Method

The sample, procedure and analysis will now be looked at.

2.1 Sample

Respondents who indicated their willingness to participate consisted of geographically spread and internationally established authors in the field of dynamic assessment. Respondents were clinicians in private practice as well as professors and doctors within schools of psychology, counselling psychology, education, pedagogical psychology, teaching, learning and instruction, learning disabilities, biological psychology as well as psychoeducational consultation and training. Respondents currently teach and/or co-ordinate programmes within tertiary institutions in the United States, Canada, Britain, Israel and the Netherlands and are also involved in private-practice consultations.

2.2 Procedure

Whitley (2002) states four main steps along which content analysis should proceed which reflects the process of qualitative research in general (Creswell, 2002) namely, the sources of data to be utilised, the sampling of respondents, the development of a coding scheme and the measurement of the content. Electronic mail was delivered to one hundred dynamic assessment practitioners and researchers whose contact information was gleaned from accredited literature, textbooks, chapters in edited texts as well as articles. Undoubtedly there remained many practitioners and researchers whom the author did not contact. However, those chosen were most visible in terms of research output. Thirty-one individuals initially responded making their intentions clear as to possible participation in the study. Follow-up emails were sent via Internet four months later to invite responses from those who had initially accepted participation but who had not yet participated. A covering letter was also included with the questionnaire.

2.3.1 Analysis

The analysis of the responses is now delineated by discussing the coding of the content, the process as well as the development and deployment of the coding frame.

2.3.1.1 Coding the content

The author followed Creswell's (1998) general and overall mode of enquiry which is viewed as a spiral of research emphasising the description, classification and interpretation of text. Once the pooled views were grouped under each question a thematic analysis was conducted in order to fully explore underlying themes within the summarised responses. Once the themes had been highlighted it was considered prudent to investigate the responses for aspects not mentioned as this is also considered of importance within content analysis. Certain issues that were not mentioned are indicative either of the irrelevance of the topic or the lack of knowledge surrounding the particular issue of concern. "The idea is to ignore what a writer says a theme means, and instead to focus on what it is that he mentions when he talks about it and exactly what he *isn't* mentioning can [then] be seen" (original emphasis) (Carney, 1972, p.162). The process and coding frame follows after which a discussion of the findings is included. Conclusions are drawn from the themes highlighted and suggestions proffered in light of these findings.

2.3.1.2 Process

Weber (1986) delineates seven steps towards creating and testing a coding scheme and this consists of defining the recoding units, defining the categories, testing of the coding on a sample of text, assessing the accuracy or reliability, revising the coding rules, returning to sample coding, coding of all the text and lastly assessing achieved reliability or accuracy. Closely allied to this scheme is the one proffered by Strauss and Corbin (1998) which has as its point of departure the conceptualisation or breaking down of the data, the defining of terms and lastly the categorisation and sub-categorisation of data.

Themes are utilised as meaning unit and the categories of the coding frame (Berg, 2001) will consist of axes of meaning which will either support an issue or not, or else will be regarded as neutral on an issues. The first three responses to each question was investigated for suitability after which the aforementioned axes were routinely applied across the questions except for question 3 which did not avail itself of such coding. The coding of the text proceeded line-by-line (Strauss & Corbin, 1998) and followed Whitley's (2002) recommendations in terms of the characteristics of coding systems. Briefly, these characteristics dealt with the nature of the coding scheme and are discussed below:

- Coding schemes can be theory-based or ad hoc. This study dealt with very specific issues which were directed at learned practitioners and academics within the field. Thus the questions were inherently theory-based at the outset. In other words the author was expecting most of the responses to fall into predetermined categories. Emerging themes from the responses were thus pre-empted by the foregoing pre-established categories. Due to possible sample bias⁵ certain themes may have arisen which were not originally considered. This was not, however, the case
- The focus of the coding scheme is reflected in the breadth of scope. A broadly focused scheme was employed seeing as the study sought to extract counts of themes running through the responses. The nature of the focus of the coding scheme is dependent on the amount of information available about the responses as well as the degree of detail necessitated by the research. The main motivation for conducting the research was to determine the extent to which practitioners and academics agreed or disagreed with certain sentiments but simultaneously allowing them the scope to support their contentions as well as to add any information they sought fit to include. Unfortunately, most responses were not very detailed primarily due to the overly complex nature of the questions
- There were eight general coding categories, one for each question. Within each question a number of sub-coding categories were utilised based on the nature of the question
- Inferences from responses was kept to a minimum as the author and respondents conversed about a topic which was known to both parties. In other words the respondents' intended meaning was less likely to have been misperceived (Strauss & Corbin, 1998)

2.3.1.3 Coding frame

All eleven returned questionnaires were read for similarities and differences in viewpoints to each question. The coding frame was worked out based on the responses to the various questions and was not a fully developed framework developed before receipt of the questionnaires. However, due to the focused nature of the questions which already juxtaposed certain viewpoints, a vague coding frame had already been established. Questions 1 to 4 elicited a range of views on a number of issues whereas questions 5 to 8 were dichotomously phrased. The coding frame employed for each question follows below.

Question 1 The status of dynamic assessment as scheme, model or theory

In this question

- A tentative scheme would be considered as being least developed conceptually, while
- A model is considered more specific yet less explicit than
- A theory which would be most conceptually developed.

Question 2 Dynamic assessment regarding the following core issues:

- Nature vs. nurture
- Mind and brain vs. Mind is brain
- Pragmatic agenda vs. Socratic agenda
- Direct role of history vs. Indirect role of history

Question 3 Developmental model followed within dynamic assessment

⁵Interestingly, of the eleven completed responses which were returned, ten completed questionnaires were returned within one month of being sent. This observation may or may not indicate some information about the respondents. Whether this can be called a response bias though is debatable.



Although a few developmental theories were offered for illustrative purposes most of the respondents aligned themselves with these theories even though the question was an open-ended one. This question will be analysed in an open-ended fashion and not as categorised as the first two questions.

Question 4 Dynamic assessment within the broader field of intelligence assessment

This question was concerned with

- Intelligence models and
- Mutually inclusive and mutually exclusive views as well as theories on dynamic assessment within intelligence

Question 5 Dynamic assessment and neuroscience

This question revolved around the relevant issue of neuroscience research within an area such as dynamic assessment and sought to locate responses on the following issues:

- Findings in neuroscience have played a role in dynamic assessment vs. Findings in neuroscience have not played a role dynamic assessment
- Findings in neuroscience should play a role dynamic assessment vs. Findings in neuroscience should not play a role dynamic assessment

Question 6 The future of dynamic assessment

This question elicited views concerning the future trend of dynamic assessment given its current status.

- It will cease to exist in its current form vs. It will continue to flourish if changes are made

Question 7 The quantitative imperative

This question focused on the perennial debate of quantification within psychology and how it may or may not relate to dynamic assessment in particular.

- Measurement is a necessary part of dynamic assessment vs. Measurement has resulted in dynamic assessment moving away from its original ideals

Question 8 Meta-theory

This question only superficially probed the significance of deploying meta-theory within this field as framework for guiding further theory at its current point of development.

- Meta-theory will only serve to add confusion to the field vs. Meta-theory is a welcome addition to this field

3. Results

This section details the results of the summarised responses as well as the content analysis of identified themes as per the coding frame. Responses to each question were pooled and analysed in terms of the coding frame employed above. Themes were extracted and summarised. The results contained in this section are the author's summarised analysis of respondents' results and are not the respondents' verbatim responses to the questions. For clarity and ease of use, only the statement preceding each question will be repeated. Each question will be discussed, firstly, in terms of the summarised theme that was extracted and secondly, in terms of the coding frame, the results of which are tabulated.

Question 1 - The status of your particular view of dynamic assessment within intelligence assessment

Extracted summarised theme

Dynamic assessment is currently considered as more of a model than a fully-fledged theory and consists of hybridised approaches towards the assessment of the whole individual within varied contexts. There appears to be lack of consensus surrounding its status. It is informed from a variety of implicit assumptions about learnability, the learning experience, the potential to learn and the modifiability of individuals. Dynamic assessment undergirds the assessment of cognitive functioning and includes affective and non-cognitive aspects over and above intelligence traits. Dynamic assessment is seen as relevant in culturally diverse situations where the whole person and context is taken into account and where reality is socially constructed. It is an holistic and ethical approach towards understanding cognitive functioning. Dynamic assessment, like intelligence, is a construct and not a reified thing, it is a method of assessment.



Theory does inform dynamic assessment but there are many such theoretical underpinnings. Dynamic assessment cannot yet be truly classified as a theory but rather a conceptual scheme and model. Some practitioners prefer to house dynamic assessment under the overarching theory advocated by Feuerstein and model advocated by Vygotsky as well as the structuralist approach of Piaget. However, there are contrasting views which espouse that dynamic assessment is in fact a paradigm and is firmly entrenched in theory whereas others have indicated that dynamic assessment is a philosophical stance in which there simply is not enough research to support dynamic assessment as a theory. Clearly there is diversity in what dynamic assessment is. Some practitioners view dynamic assessment purely as a convenient measure of potential. There is difficulty in placing dynamic assessment within the larger arena of assessment as it seeks to measure things which are themselves ill-defined. There is difficulty in deriving a tool or instrument sufficient and worthy enough to measure something which in some ways defies measurement, as operationalising the concept is problematic. At this stage there are still too many questions to be answered before dynamic assessment as a tool, view, scheme, model, construct or theory can be digested into the greater realm of assessment within psychology. There is, however, empirical evidence to suggest construct validity and this is seen as a tentative beginning to its growth as a method within the scientific discipline of assessment.

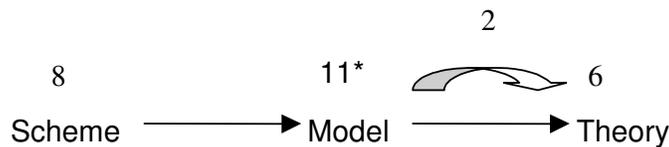
Coding analysis and tabulation of extracted themes

Analysing the results from both a mutually exclusive and mutually inclusive point of view shows a greater tendency to view dynamic assessment as a model. Seven mentions of dynamic assessment as model were counted. Dynamic assessment's nature also wavered between model and theory status with it being viewed as a scheme or sorts. Six mentions of dynamic assessment as theory were counted. Although the perception surrounding the nature of dynamic assessment varies from mentions of schemes, models and theories, it is perceived by more practitioners as a model than a theory. Table 2 illustrates these trends. Figure 1 illustrates the continuum of increasing conceptualisation and where respondents indicated dynamic assessment could be placed.

Appendix 2 Table 2 Dynamic assessment's status as scheme, model or theory

Mutually exclusive categories	Frequency count	Mutually inclusive categories	Frequency count	Opposing view	Frequency count
Model	7	Theory/model hybrid	2	Not a theory	2
Theory	6	Model with concept as subservient	3		
Assessment/procedure/approach	3	Model with theory as subservient	1		
Concept	3				
Paradigm	1				
Philosophical stance	1				

Appendix 2 Figure 1 View of dynamic assessment along a continuum of increasing conceptualisation



*when including the mutually inclusive category. Seven exclusive mentions

Question 2 - Philosophical bent or underpinning of your theoretical take on dynamic assessment

Extracted summarised theme

Dynamic assessment is firmly anchored in socio-cultural and bio-ecocultural models of a socially constructed reality. It places greater emphasis on how the environment influences change although cognisance is taken of the increasingly important role of heredity. The potentially rich field of neuroscience is acknowledged as method of potential future utility only. The processes involved in learning are socially constructed and hence many views of dynamic assessment are rooted in social constructivism. The main emphasis within dynamic assessment is placed on learning as opposed to thinking although this is not always explicit. There is a problem surrounding the notion of intelligence which is most often considered socially constructed, intelligence being



a hypothetical construct. Intelligence should always be considered within an ecological context. It is ironic that, although dynamic assessment's emphasis on culture is great, within practice the goal of enhancing cognitive functions pays little attention to culture. If there are problems concerning what intelligence is then dynamic assessment itself will face this same problem if its aim is to assess intelligence. The main aim is not to assess intelligence but rather cognitive functioning within learning. Learning as a concept should be included in the definition of intelligence. Dynamic assessment seeks to investigate the learning process and how better to make manifest potential resident within individuals and as such is not aligned with intelligence measures per se as is conventionally understood.

Genetic heritage places constraints on the limits and parameters of intelligence and this type of intelligence can be assessed to a greater extent within static assessment as opposed to dynamic assessment. Dynamic assessment is more concerned with the role of nurture and how the socio-cultural environment influences outcomes. It concerns itself with cultural communities' influence on thinking and knowledge and how mediators can facilitate the process of reaching resident potential within individuals. Knowledge is also socially constructed. Nature and nurture are considered equal role-players but there is a distinct emphasis on the role of nurture. Static and dynamic assessment are at once viewed as antithetical as well as complementary although it is recognised that the two differ in approach as well as differing in basic implicit and explicit assumptions. The two methods attend to and answer different questions. Dynamic assessment does not really sit well within psychometric approaches but is more suited to educational, counselling and clinical models and is referred to as an anomaly within the intelligence field. Brain malleability and plasticity within neural substrates and the results emanating from neuroscience are starting to play a more important role within intelligence assessment and dynamic assessment takes cognisance of this fact. Dynamic assessment has always had as its basic premise a malleable brain. Learning is itself malleable as is thinking, actions and human potential. In this manner dynamic assessment seems to have foreshadowed the role of the plastic brain in intelligence assessment, although not in neurological detail.

Dynamic assessment is also viewed as both quite an established philosophy but also as a new method. Historically, dynamic assessment is an ideology associated with Vygotsky whereas some view the beginnings of dynamic assessment with the work of Feuerstein and Rey. Dynamic assessment's history is referred to as quiet and not having "caught on" in mainstream assessment, it being simultaneously ignored, supported and discredited. In order for dynamic assessment to become accepted within mainstream intelligence assessment, if this is where it wishes to lodge itself, it needs to become standardised which is antithetical to the notion underpinning it.

Coding analysis and tabulation of extracted themes

There is a singular effort to ensure understanding of the dual role of interacting nature and nurture within dynamic assessment. Nine mentions of this interacting role were counted. Cognisance is taken of the "hardware" or genetic constraints within which the individual operates and eleven mentions of regarding this view was tabulated. There is, however, a strong and almost overriding concern with the nurturing component within dynamic assessment indicating the role this plays within the model, evidencing thirteen such mentions. That the debate is obsolete is validated by the responses given, illustrating an appreciation for both nature and nurture. Not much mention was made of the roles of mind/brain; the (in)direct role of history or the agendas behind dynamic assessment as an assessment initiative. Table 3 shows the frequency counts for the responses to this question.

Appendix 2 Table 3 Dynamic assessment philosophy

Aspect	Frequency count	Aspect	Frequency count
Nature	11	Nurture	13
Nature/nurture interaction	9		
Mind and brain	0	Mind is brain	1
Pragmatic agenda	2	Socratic agenda	0
Direct role of history	2	Indirect role of history	0

Question 3 - Your developmental model followed or model most adhered to within the field of child development/educational development as well as adult growth and maturation

Extracted summarised theme

No one particular developmental model is adhered to within dynamic assessment and of those mentioned most are ecologically aligned theories evidencing eclectic influences. Theorists such as Feuerstein, Vygotsky and Piaget are mentioned as being influential in dynamic assessment's progress since inception. Whether a narrowing in scope of chosen developmental models will occur is speculative as best. Multiple theoretical models are utilised as no one specific model encompasses all that is necessary to explain the global functioning of the learner. Hybrid and eclectic models are preferred over-and-above any one



particular model. Cultural, bio-ecocultural, developmental and environmental models which take cognisance of community, person-in-context, biological predispositions, religious contexts, social contexts, ecological influences and individual differences are commonly espoused as the necessary models when tackling issues such as assessment. It is clear from these choices that no one specific model will do. More importantly, models which emphasise environmental influences are considered more potent in terms of explanatory power than those which are more narrowly focused on intelligence as a biological derivative only. More modern approaches are fused with older models of intelligence and information processing and thus incorporate the influencing aspects of the environment, socio-economic status of family and quality of learning environment at home and school. How these factors help in co-constructing knowledge acquisition is important in these constructivist approaches.

Coding analysis and tabulation of extracted themes

As stated above this question was analysed in an open-ended manner and no table exists for this question. There is clearly no one particular developmental model adhered to throughout the developmental life-span of individuals. Changes occur within individuals, so models are accommodated to suit these changes and dynamic assessment approaches are considered hybridised and eclectic. As biological predispositions are said to play a greater role during childhood and that of the social environment too, later on, development models with varying focuses are employed to aid in the description and explanation of change. Ecological, environmentally aligned and bio-ecocultural models are most often cited as those followed as the socio-cultural environment including home and work are emphasised as impinging upon development. Individual theorists are mentioned such as Piaget, including neo-Piagetian models, Vygotsky, including neo-Vygotskian models, Feuerstein and the Mediated Learning Environment and Structured Cognitive Modifiability theories; Ceci, Bronfenbrenner, representative of the bio-ecocultural and ecological models respectively, Kozulin, Haywood, Karpov, as modern representatives of classical Vygotskian theory; social-constructionist views on development (Gergen); Sternberg and Grigorenko as well as information processing and intelligence theories as well as Rogoff's work on cultural models of development. Greater importance is attributed to the role of the environment with all its associated models.

Question 4 - The ensconcement of your theoretical take on dynamic assessment within the broader fields of intelligence assessment

Extracted summarised theme

Dynamic assessment makes use of intelligence tests which function in a role considered complementary. Intelligence is an ill-defined notion and as such is difficult to define especially its role within dynamic assessment. Dynamic assessment sees itself as encompassing intelligence and not as intelligence encompassing it. Dynamic assessment was therefore not placed within any intelligence theory or model. Intelligence as conventionally understood and measured is not the target of dynamic assessment intervention. The focus is more on the process of cognitive functioning and how best to address these issues if improvement is to be the final goal. The emphasis is on the remediation of problems within cognitive functioning and very often these problems are a culmination of environmental, ecological, socio-cultural, community and family concerns in which the developing child is situated. Adult intervention is also of concern to dynamic assessment. Standardised intelligence assessments are used complementarily and are useful to a point within dynamic assessment interventions. They are utilised mainly as estimates of current levels of functioning, levels which are improved on through the use of dynamic assessment and often serve as the pretest during a dynamic assessment intervention. Standardised test results for individuals from different cultural contexts typically reflects a very narrow construal of what intelligence is purported to be. Also problematic is the ill-defined concept of intelligence. Some consider intelligence not to exist as it is currently accepted. Intelligence is viewed as that which an individual can do without aid from a mediator. Some view it as higher mental functioning referring to the conventional and mainstream concepts of speed of information processing for instance. Of note is the role of executive functioning as a definition of intelligence. Dynamic assessment research is related to intelligence research but cannot be equated with it.

Models of assessing intelligence are derived from the assumption of what intelligence is, namely, thought processes which are easily penetrable and easy to access en masse. Contrary to this is dynamic assessment's assumption that each individual is so unique that in order to properly address cognitive skills, knowledge of individual contextual factors is necessary. This is usually not an assumption included in mainstream standardised testing. Taking cognisance of contextual factors results in a time-consuming process in comparison to mainstream assessment. Mainstream intelligence testing is product-bound whereas dynamic assessment is process-based. Issues such as testee morale and ethical responsibility are more closely scrutinised within dynamic assessment as opposed to conventional intelligence assessment. Greater responsibility is placed on the dynamic assessment practitioner when assessment and intervention programmes are discussed. Some do view dynamic assessment as fitting in within views on intelligence but view the technique as an alternative to mainstream testing. Perhaps intelligence as a concept should be elaborated to include aspects not currently included.

Others view intelligence and dynamic assessment as distinct and not embedded as dynamic assessment does not sit comfortably within traditional models of intelligence. Intelligence, if it does exist as a unitary concept, cannot be divorced from its



non-cognitive counterparts, an aspect not dealt with adequately within mainstream intelligence testing. Intelligence as a concept can be further refined by adding that it is something that develops and represents developing expertise; it is an executor of cognitive functions; it is a metacognitive concept; or is a socially constructed concept and should encompass change as a basic premise. Problems arise with dynamic assessment in the typical intelligence assessment environment and these issues usually pervade the area of psychometrics which is the typical set of tools used to extract information from intelligence assessment. If dynamic assessment does fit into the boarder field of intelligence assessment then it has, by definition, to lend itself to psychometric scrutiny. Issues concerning test items, changeability and testee malleability come to the fore. Psychometric models do not fit very well with dynamic assessment models as the two are opposed in terms of how reliability and validity is to be supported.

Coding analysis and tabulation of extracted themes

Intelligence is understood in various ways. It is seen as measuring a single entity and providing only single scores, for instance when total IQ scores are utilised. It is viewed as fairly inconsequential and thus very rarely necessitated, utilised only as complementary in the assessment procedure. It is also seen to function as an estimate of current functioning and a control variable within the dynamic assessment process. As a construct it refers to higher mental functions (Vygotsky) and includes abstract thinking, strategic problem solving, speed of processing and retrieval ability. It is related to but not equivalent to dynamic assessment and functions as the pretest part of a dynamic assessment procedure. It is viewed as the ability a child before intervention. It is seen to fit in with dynamic assessment views but as alternative measure. It has been referred to as a notion, as something in need of elaboration (and thus being more inclusive of aspects relating to ZPD) as well as a repertoire notion which can benefit from dynamic assessment.

Simultaneously it is seen as distinct from dynamic assessment and not embedded within it; a verb and adjective more so than as a noun; as Sternberg's notion of developing expertise; of being applicable, for instance through the use of executive functions and lastly as a socially constructed notion as opposed to a definitive reality. It is also viewed as not necessarily being divorced from change-based assessments. Five mentions were made of intelligence as serving no direct purpose for the furthering of dynamic assessment in contrast to eight mentions of intelligence as an indirect adjunct to dynamic assessment. Two mentions were made as to its placement within and on the foundation of dynamic assessment. Table 4 illustrates the frequency count of the responses to this question.

Appendix 2 Table 4 The role played by intelligence within dynamic assessment procedures

Aspect	Frequency count
No direct use for intelligence assessment	5
Use for intelligence as an indirect adjunct	8
Intelligence assessment can be built upon by dynamic assessment	2

Question 5 - The affinities your particular theory of dynamic assessment has with aligned fields of neuroscience, neuropsychology and computational intelligence

Extracted summarised theme

There is fairly wide-spread consensus about the increasingly important role of neuroscience within dynamic assessment and how it may inform the process but this is tempered by the fact that such consilience will occur later rather than sooner. At present such findings do not play a very prominent role within dynamic assessment. Findings from neuroscience are increasingly playing a major role in psychological literature as it pertains to cognitive functioning. Practitioners are starting to take a closer look at what the neuroscience field has to offer dynamic assessment as there is a move towards greater understanding of brain functioning. Information gleaned from neuroscience is important in some areas of dynamic assessment interventions such as reading disabilities and areas of individual differences. It has been the case that theory within psychology has been validated with techniques from neuroscience and perhaps multidisciplinary research efforts will be the order of the day in the future. Neuroscience can, among other things, help in the understanding of the cognitive functions involved in learning and instruction as these neurological processes and other reciprocal processes are related to factors that are assessed during dynamic assessment although it is not necessarily a main focus for many educationalists.

To date, neuroscience-type results do not form part of results and recommendations emanating from dynamic assessment intervention programmes (nor should they as some have stated) and is often seen as futuristic in approach, even though current studies have in fact evidenced that neuroscientific findings can currently be of use. The perception still lingers that neuroscience has far to go in mapping a one-to-one relationship between behaviour and brain. Dynamic assessment is a more practically-based approach to helping individuals in their day-to-day functioning and as such neuroscience results have less of a role to play. Some view the role as alien and simply not befitting to the task assigned and taken up by dynamic assessment. There are



competing views on this topic. If neuroscience views are to be utilised within dynamic assessment interventions then they will have to be tailored to suit the approach and inherent assumptions and are seen as a means towards dynamic assessment's goals. Dynamic assessment's main focus should rather be attuned towards developing better theory and hence better practice as opposed to focusing on the integration of neuroscience in the field. There is consensus that although neuroscience is obviously a major development within psychological science, its role is not paramount to the future development of dynamic assessment as a model.

Coding analysis and tabulation of extracted themes

There is considerable support for the use of neuroscience findings within dynamic assessment and how dynamic assessment can benefit from such knowledge. Thirteen mentions were made of the prospective role that neuroscience should play within dynamic assessment. Diagnosticians should avail of themselves the time needed to understand more fully the field of neuroscience, however it is still considered an approach whose findings will be more easily incorporated into the model in the future. Neuroscience explanations of developmental functioning is not yet widely reported within dynamic assessment procedures. There is strong vocal support for the complete separation of the two distinct fields as they may "have nothing to say to each other" with six mentions made of the separation of neuroscience from dynamic assessment. In this regard neuroscience as a field is viewed as having less to say about how to help individuals learn on a behavioral level. Mention was made of one dynamic assessment approach which does in fact make use of neuroscience leanings. There is a strong sense of the weaving together of both approaches but that neuroscience as method with its results should serve only in the capacity of "consultant" in the dynamic assessment procedure and not vice versa; it should not be the main focus. Table 5 illustrates the frequency count of the responses to this question.

Appendix 2 Table 5 The role of neuroscience within dynamic assessment

Aspect	Frequency count
Findings in neuroscience have played a role in dynamic assessment and these findings in neuroscience should play a role in dynamic assessment	13
Findings in neuroscience have not played a role in dynamic assessment and findings in neuroscience should not play a role dynamic assessment	6

Question 6 - The historical development of this sub-field of enquiry and its potential future within the realm of psychology

Extracted summarised theme

Dynamic assessment will cease to exist in its current form if changes are not initiated at very core levels. However, it is these very core issues which currently define the manner of assessment so there seems to be a tension of sorts, which will need to be addressed if it is to remain extant although no viable solutions were offered as tentative solutions to these issues. Most practitioners fervently hope that the field of dynamic assessment remains a field of research and also wish for greater acceptance among mainstream assessment. What is perhaps the most implicitly impassioned complaint is the fundamental philosophy underlining dynamic assessment and how this will be lost if the method which it underpins ceases to be practiced. Problems with standardisation as well as issues of costs and time seem to mitigate against the further acceptance of the approach within a wider setting and it is ironic that at times those most opposed to the approach are very often psychologists. Unbeknownst to some mainstream educationalists, assessors and researchers, are the basic philosophies behind dynamic assessment. These philosophies lend themselves to current trends of "responses to treatment" and "evidenced based treatment". As is commonly lamented, cost and value are often opposed in the process of intervention.

Although the hey-day of purely qualitative dynamic assessment is over (assuming that it ever had one), what is of more concern currently is whether there is a critical mass of younger generation practitioners who wish to engage in this method of assessment. As some have stated, dynamic assessment as such will most likely die. What may be holding back the development of dynamic assessment is its need for strong theory and solid research and taken together with high costs and amount of time needed to use the method, the feasibility of the method as a whole is questionable. This is especially so as it is often perceived as an experimental technique and not an established one. Removing the persistent question of changeability and modifiability of the individual and the subsequent need to address this will be challenging. Some view the future to be less gloomy as it is stated that there is yet enough interest in the field to sustain its future growth. The influence of tangible quantifiable results, however, mar the progress of the field especially in countries which emphasise the need for speedy and accurate predictive results. This is in stark contrast (and is in some ways ironic) to those who view the natural science model of intelligence assessment as too rigid. Collaboration with other social scientists may be one avenue of hope for this method and seeking across-the-board collaboration fits well with the holistic trends which pervade the method, in terms of assessing the whole person by considering context in all its forms.

Coding analysis and tabulation of extracted themes

There is indeed a heartfelt expressed concern that dynamic assessment as a manner of assessment should not die out and that it should continue to flourish but this is only possible if a number of core issues are solved. Issues of concern are its usurpation of time, its incompatibility with standardised psychometrics, lack of funding for expert training and its, at times, under appreciated status by psychologists, the very professionals who should be seen to be supporting it. These core issues, if not adequately addressed, will ineluctably bring this approach to its knees (where it apparently is according to some views). Only one mention was made of dynamic assessments' inevitable demise. Far more respondents were in agreement as to its future only if changes were instituted. Twelve mentions of its future existence is testament to this positive sentiment. It offers rich and tantalising assumptions which are difficult to criticise but the fact that these assumptions are not firmly expressed in terms of theory is seen as a fault in need of remedying. Certain types of dynamic assessment are considered to be in grave danger of simply dying out as there are fewer trained professionals utilising such approaches, however this is not to say that the entire suite of dynamic assessment will cease to exist. Table 6 shows that only one mention was made of a particular view on dynamic assessment's total cessation and the remaining mentions concluded that, if issues were adequately addressed, there remained hope for its continuance.

Appendix 2 Table 6 Dynamic assessment's future existence

Aspect	Frequency count
It will cease to exist in its current form	1
It will continue to flourish if changes are made	12

Question 7 - The quantitative imperative

Extracted summarised theme

The move away from quantification has stemmed from a backlash of criticism aimed at standardised methods of assessment which have had some unfortunate consequences for some groups of individuals. Dynamic assessment's predicate of qualitative intervention aimed specifically at change through assessment views quantification as anathema yet a necessary part of the assessment process. Historical relevance in terms of these issues was however not mentioned. As highlighted in previous questions dynamic assessment's core problem within the standardised arena is one of psychometric validity and reliability. It would seem that this issue will forever plague this approach towards assessment. It is commonly felt that unless this issue is adequately addressed (something it seems not to have been to date) dynamic assessment will remain a fringe movement. However the basics of this approach advocate a distinct turn away from the very core of what underpins mainstream assessment. Issues of great import to conventional intelligence assessment are simply not considered that important even if relevant to dynamic assessment. Adding fuel to the fire is the degree of acceptance of qualitative techniques of research which meet the standards laid down for the more quantitative methods. There is an underlying view of quantitative scores being rendered meaningless unless they inform the remediation of the person being assessed. On the other hand, it is perceived that statistical techniques will most likely be the saving grace for this movement as it might progress with the falsification of certain hypotheses. Both quantitative and qualitative are seen as necessary data analytic techniques within dynamic assessment although critical views on the topic are needed if scientific advancement is to be the next step within the field. The analogous pendulum-swing is cited as evidence of movements within educational research and practice, where, what is considered priority in one time and place, is no longer considered as such in another time and place. The need for empirical verification of educational interventions is stressed. Conversely the unquestioning use of statistics and psychometrics within dynamic assessment intervention is viewed as anathema to the further development of the field and also as prohibitive in the growth of dynamic assessment. In order to establish validity for dynamic assessment, new techniques need to be brought to the fore. This is seen as problematic by some in certain countries such as North America where the reliance on mainstream testing norms is prolific and may be damaging to those individuals who need to be assessed through alternative manners.

Coding analysis and tabulation of extracted themes

A firm stance was taken on the role of measurement within dynamic assessment and seven mentions of its necessity within dynamic assessment were made. Some advocated that there really is no role for statistics within this approach as it has in the past resulted in mismeasures, narrow-minded conclusions and social injustices and to this effect five mentions were evidenced. A middle-of-the-road view considers a place for both quantitative and qualitative research methods including the role of statistics, but only in so far as they can be used for the further remediation of the individual. Dynamic assessment's approach towards the role of assessment and the rightings of social injustices is reminiscent of a critical psychology approach and is quite evident in the responses to this question. The role of quantification in general and statistics in particular are rendered useless,



unless something positive can come from these results and aid in any future changes within the individual. Table 7 illustrates the frequency count of the responses to this question.

Appendix 2 Table 7 The role of quantification and measurement within dynamic assessment

Aspect	Frequency count
Measurement is a necessary part of dynamic assessment	2
Measurement has resulted in dynamic assessment moving away from its original ideals	5

Question 8 - The meta-theoretical solution or pie in the sky

Extracted summarised theme

Meta-theory as approach towards finding possible solutions is considered premature and perhaps even confusing for some, yet there are opinions that just such a method of analysis may well yield fruitful information, unobtainable via other means. Examples of such a technique were not highlighted. There is a complement of support for the notion of tackling the issue of meta-theory as it pertains to dynamic assessment and fervent pleading for such meta-theorising to cease as it will merely add chaos to a field already deeply entrenched in controversy. Meta-theorising, as understood to be theory about theory within intelligence and dynamic assessment may be fruitful in the long-run. More brain-storming and think-tank sessions are required to creatively grapple with cores issues that still remains unsolved. Issues such as the definition of intelligence and what it entails and the need to currently consider eclectic views within dynamic assessment are broached. Some view the entry of meta-theory within this field as a welcome addition whereas others voice their concern at the confusion that may ensue if it is introduced too early. Some consider dynamic assessment as yet too young a field theoretically-speaking and meta-theorising is often a task undertaken after established solid theorising has taken place. There is also a need to bridge or link up the various models within the broader dynamic assessment field, as its current state is still too disparate. As part of this meta-theorising initiative, thinkers from across disciplines need to be brought on board and this will help alleviate the reductionist trends now common within dynamic assessment research literature.

Coding analysis and tabulation of extracted themes

The fact that there is a call for a more integrated approach towards the assessment of the total individual from as many perspectives as possible helps make the need for a meta-theory understandable. This view is tempered with cautionary warnings about interfering too soon within the eclectic mix that is currently the situation within dynamic assessment. Most responses considered the need for meta-theory as useful but not essential to its development. Six mentions about meta-theory's confusion-causing role were voiced and indicated that the role of meta-theorising would not aid in the future development of the field. Four mentions were made in support of a meta-theory as framework for the development of dynamic assessment. There is a noted concern for the need to study the finer theories and models within the field first before heading off on grander meta-theorising. Table 8 illustrates the frequency count of the responses to this question.

Appendix 2 Table 8 The role of meta-theory within dynamic assessment

Aspect	Frequency count
Meta-theory will only serve to add confusion to the field	6
Meta-theory is a welcome addition to this field	4

4. Discussion

Qualitative research is by nature a probing exercise delving more deeply into issues than is usually the case with more quantitative research. Richness of responses as opposed to the number of responses plays a greater role in this type of research. Although only eleven completed questionnaires were analysed for this study, this in no manner detracts from the depth and richness of the results. The nature of the sample has been detailed and it is due to the cumulated expertise of respondents that the study's results can be considered as worthy contributions to the field of dynamic assessment. The author felt secure that a number of issues had been adequately addressed in terms of the patterning of responses. For instance, it was very evident that almost all responses to the nature-nurture question were unanimous in terms of agreement that both play a role in assessment. Other issues such as the quantitative imperative did not yield a similar pattern of responses and further answers from more respondents would have greatly enhanced the study's utility value. Some responses were thus more saturated than others. Specific issues which were not mentioned will now be discussed.

4.1 What was not mentioned - a further source of information

Certain issues were only fleetingly discussed and if expanded on may well have proved fruitful if taken to their logical conclusions within the discussions. Lack of considering certain points within each question may well prove insightful as these neglected areas are possible aspects that may be perceived to be either non-essential or as having already been solved.

Regarding the first question, the lack of consensus regarding the status of dynamic assessment as theory, model, conceptual scheme, philosophy or even construct brings into question the degree of explanatory power of a theory versus that of a model. Broader based conceptual schemes may encompass more variables and may be lodged within larger spectrums which are unable to explain with greater clarity, issues that are more easily explicated in more narrowly confined conceptual schemes which utilise more accurate terminology. The precision and power paradoxes may be useful in furthering this notion. It is possible for theories and models to achieve high precision in predictions without recourse to explanations as to how the predicted outcome was achieved whereas it is also likely for theories and models to achieve high explanatory power without predicting precise outcomes.⁶ The need for precision and prediction as espoused by mainstream assessment often goes against the philosophy of dynamic assessment, yet the lack of these seemingly essential ingredients may prove the demise of this approach.

Of interest within the responses to the second question is the mere passing familiarity with the literature emanating from neuroscience studies. This might indicate that neuroscience as a potentially rich source of information is not necessarily the area in which dynamic assessment practitioners should focus attention. Likewise, it has not yet become feasible to more fully integrate neuroscience given the already limited resources available to practitioners, educationalists and educational psychologists. Only one respondent indicated that findings in the neuroscience field have indeed proved useful in the psychological context. It is ironic then that more practitioners have not followed the neuroscience route, if as it is maintained, dynamic assessment is predicated on a plastic and malleable brain. Perhaps this view of neuroscience smacks of intensive reductionism and there is concern that the whole person-in-context will be neglected. Although alluded to cursorily in some answers the various sub-headings in question two were glossed over most likely due to the complexities inherent within the question.

The future trajectory of dynamic assessment's developmentally aligned theories will remain eclectic for now and no new developmental models were proffered. Regarding the fourth question, it would appear from the responses that intelligence as understood at its broadest level plays a role of relegated support although of importance in the determination of current levels of functioning. The specificities that the question sought were not given, most likely due to time limitations. No particular theory of intelligence was discussed as pertaining to dynamic assessment. No models of intelligence were addressed either but reference to intelligence in its general format was made.

Responses to question five were not specified in terms of particular findings within neuroscience and no mention was made of any one specific neurocomputational model as fitting in with the assumptions inherent within dynamic assessment. As this view of assessment does not currently subsume neuroscience as methodology within the approach on a large scale, the details concerning its efficacy were not discussed. Only in its generality was it considered of future benefit. This leaves one wondering whether psychologists within this assessment arena understand the fuller implications of future neuroscience modelling but perhaps, as mentioned, this is simply not yet a feasible priority. Specificities surrounding the potential resolutions of a number of issues in question six was side-stepped. No mention was made, for instance, of the use of item responses theory as measurement theory in helping to abate the increasing flood of criticism leveled at dynamic assessment's lack of robust measurement technique.

Although not directly stated within the seventh question, no mention was made of the basic philosophy underpinning the very need to utilise quantitative measures within psychological assessment. The routes travelled by mainstream intelligence assessment (leaning heavily upon psychometrics and factor analytic statistical foundations emanating from pragmatic American psychology and British empirical psychology) and that of dynamic assessment (leaning towards open-ended change and informed more from continental European considerations of assessments) was not highlighted as playing potential roles in the quantitative debate. Perhaps if dynamic assessment's origins are contrasted to those of mainstream intelligence assessment's origins, another light may be cast upon the method of change assessment and help in arguing for or against the need for quantitative measures of inference. This might not necessarily solve any immediate pressing issues but can aid in setting the scene for debate and in so doing setting an agenda for dynamic assessment in the twenty-first century.

⁶ Cf. Fawcett & Downs, 1992.



Other than a brief reference to meta-theory in question eight no mention was made of any potential developed frameworks which could be employed in a discussion of critical analysis of the direction in which dynamic assessment is to take. No mention was made of the principles informing meta-theory and how theoretical psychology could facilitate deployment of such a framework, and no specificities were delineated beyond general statements considering the utility of such an approach.

5. Conclusion

Dynamic assessment currently finds itself at a distinct cross-roads, a juncture in its history where fateful decisions will have to be made to ensure the future survival of a method of assessment which is intuitively appealing but scientifically lacking. Such broad-brush statements are bound to incite fevered retorts and it is for this reason that questionnaires were sent to one hundred of the foremost dynamic assessment researchers and practitioners, hoping to elicit responses on these very issues. Had the sample been larger it is assumed that a more balanced view would have been expressed and as the current analysis stands it may well be one-sided and thus biased. The eleven completed questionnaires were completed and returned within just over one month. At times the answers to some questions did not really address the question as posed but due to the nature of the lengthy questions it is understandable that some answers might not have been fully explored.

Although an initial response rate of 31% seemed encouraging, a number of individuals declined for various reasons thus leaving the author to analyse content of only eleven completed and useable questionnaires. Eight questions dealing with core issues within dynamic assessment were put forward and content analysis as technique for response analysis was chosen. Issues pertinent to content analysis such as the rationale for employing the technique, reliability and validity, the quantitative and qualitative nature of content analysis as well as the manifest and latent meanings within texts were briefly highlighted as they impinged on this study. The questions were discussed and the process and coding frame to each response was delineated. The small percentage of responses and answers which were kept to a minimum may have been partially due to the complexity of the questions. The results discussed the coded responses according to the coding scheme utilised for each question. Further information pertaining to issues not discussed by the respondents was also looked at. The analysis concerned itself mainly with two modes of operation; thematic patterns which were discussed at length and tabulated frequency counts of mentions for or against certain issues. The emphasis on interpretation of themes played a greater role in this analysis as opposed to the frequency counts which were purely descriptive summaries.

Concluding remarks highlighted the main trends within responses. These trends evidenced the predominating model-like status of dynamic assessment and its primary emphasis on contextual factors as major influencing variable in the testing situation. Dynamic assessment is predicated on hybridised and eclectic development models and views intelligence assessment as complementary to its main goal of mediatory intervention strategy. The utility of neuroscience and meta-analytic approaches as intertwined with dynamic assessment methods is noted but most consider it of future value only. The method of dynamic assessment as it currently stands will cease to exist if urgent attention is not given to core issues such as standardisation and quantification.