

## 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

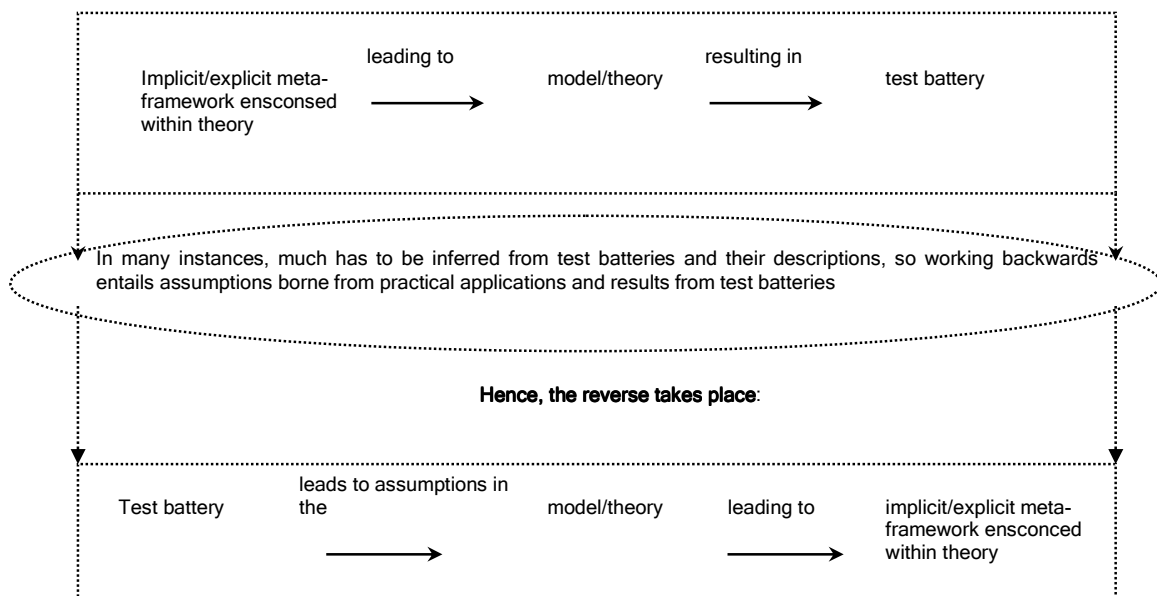
devised which can partial out certain negative features currently resident within some techniques attempting to explore and “prove” that theoretical constructs are synonymous with empirical constructs.

4. Vygotsky is perhaps the most often cited psychologist within Western educational and psychological theory and perhaps comes a close second to Pavlov.<sup>1</sup> 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:



<sup>1</sup> 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 possibly<sup>2</sup> 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)

<sup>2</sup> 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

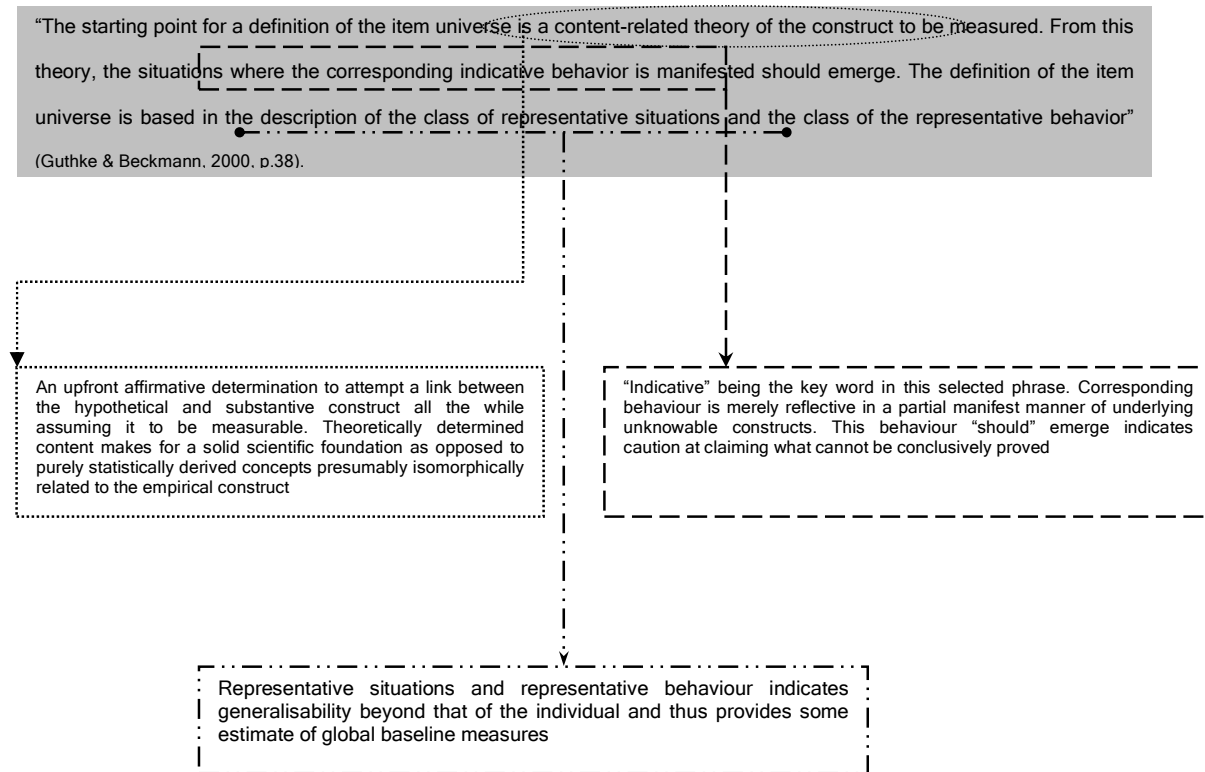
This level or stratum is really the most concrete in terms of the entire framework and concerns itself with obviously manifest and observable behaviour either behaviorally or in terms of scores on tests items. For instance, working at a slower working pace illustrates less impulsive reactions to test 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



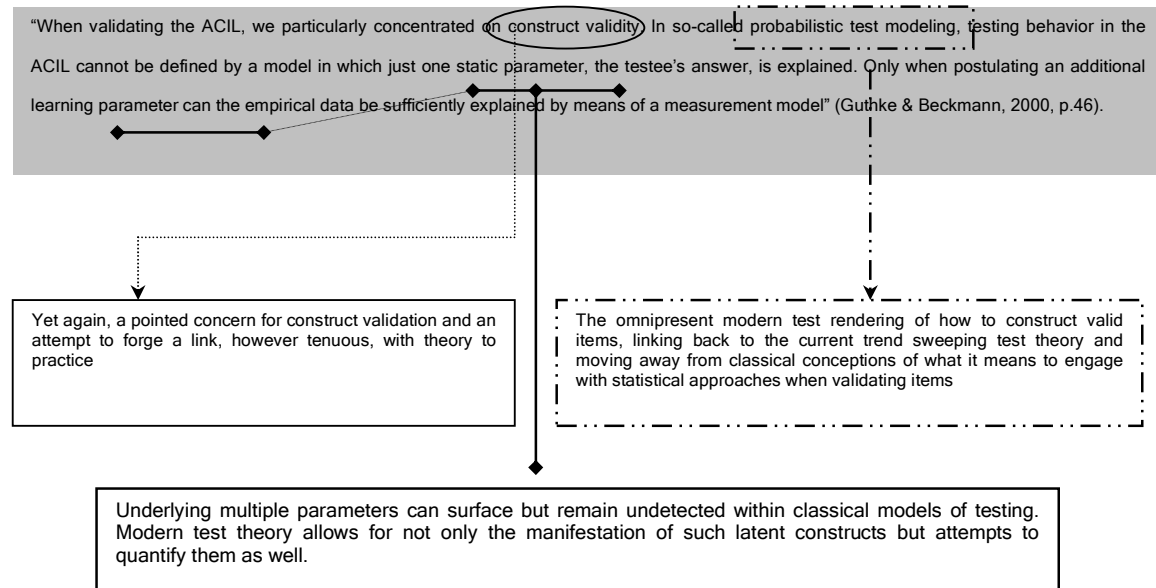
- Various dimensions which are assessed in terms of answering questions; time, latency, number of steps and so on

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)<sup>3</sup> 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:

---

<sup>3</sup> 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 numerically 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.<sup>4</sup>

#### *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 hypotheses<sup>5</sup>*

- H (Learning evidenced through verbal analogies) - S (encoding)
- H (Learning evidenced through verbal analogies) - S (inference)

<sup>4</sup> 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".

<sup>5</sup> 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.

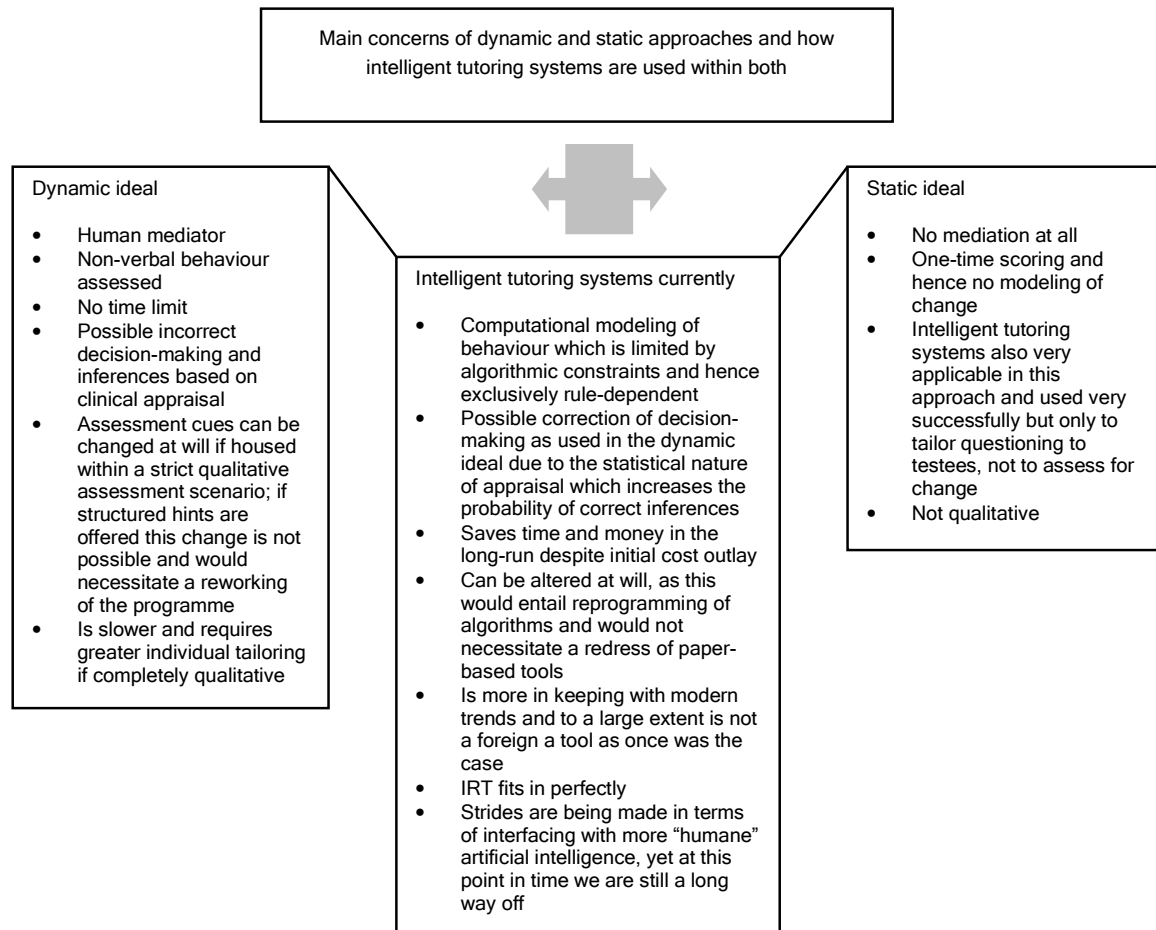
### 5.2.2.8 C(iii) Prime considerations

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

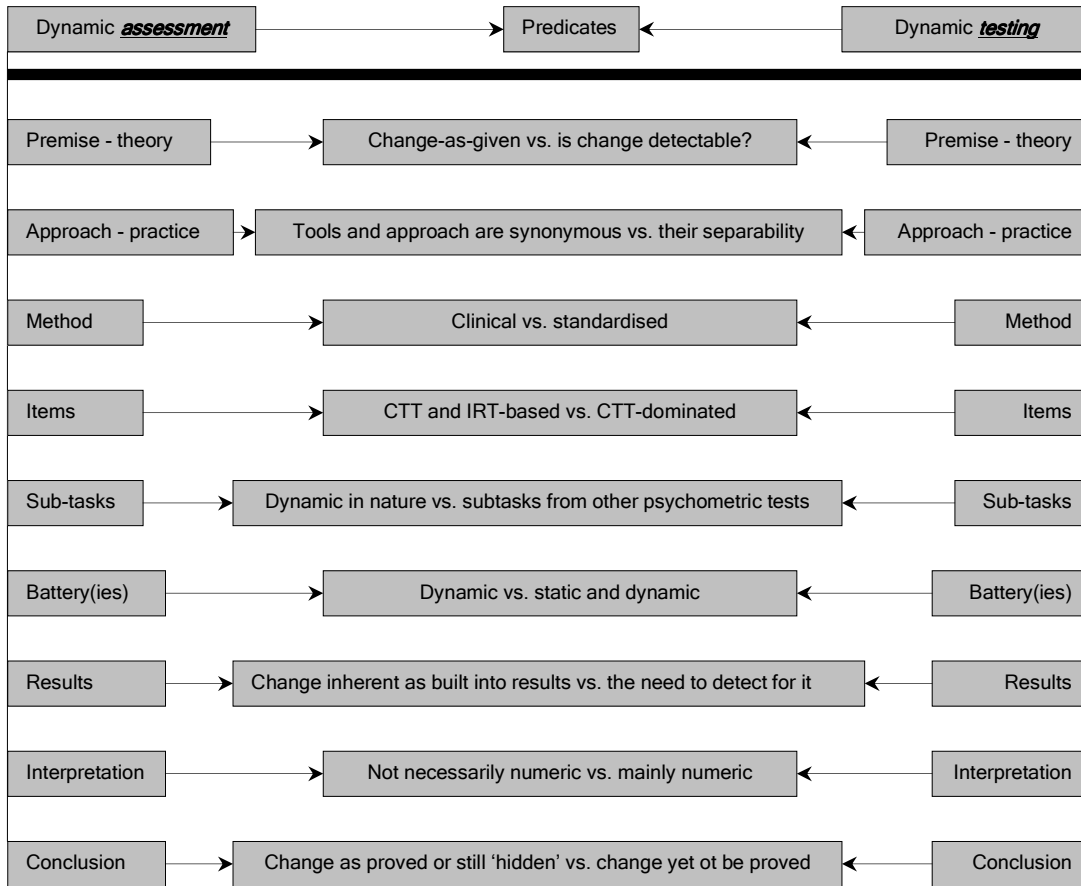
The meta-stratum ontology advocates the possibility and need to study cognitive universals which are not dependent on either cultural or social contexts. This immediately points to a standardised manner of assessment, which can prove both valid and reliable. Scores are ordinal-level representations of 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.

"... evidence must be gathered about what a given test score *means* and what conclusions and inferences can *properly* be made from it. It is at the point of conclusions and *inferences* that the issue of validity is critical" (own emphasis) (Carlson & Wiedl, 2000, p.681).

Chapter 4 discussed this issue. It is highlighted in this model's concern for what a score means; although not argued from a statistical point of view, the meaning of a result is dependent on the nature of the feedback conditions

Statistical or clinical inferential decision-making? Inference based on a sound bedrock of quantifiable measures or case-study approaches? Should the word even be used within clinical set-ups?

'Properly' in terms of 'scientific'? When will we ever know if the construct is being measured 'properly'?

Comparing techniques within the feedback methodology allows researchers yet another angle at probing the elusive multifaceted diamond<sup>6</sup> 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 (C<sub>1</sub>-C<sub>6</sub>). The modes of assessment feedback are detailed below (Carlson & Wiedl, 1978, p.560; 1992, pp.170-171).

- C<sub>1</sub> Standard procedure
- C<sub>2</sub> 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
- C<sub>3</sub> Verbalisation after solution: involves the child describing the reasons for his/her choice after the choice is made
- C<sub>4</sub> Simple feedback: where the child is informed after the choice has been made whether or not it was correct
- C<sub>5</sub> 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
- C<sub>6</sub> Elaborated feedback plus verbalisation during and after solution: is a combination of conditions C<sub>2</sub> and C<sub>5</sub>; 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

---

<sup>6</sup> 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"> <li>tentative evidence of high level of observable construct</li> <li>possible false construal of low level of observable construct (correct answer but wrong method used to reach answer)</li> <li>no explanation as to why this is the case</li> </ul>	<p><i>High change scores:</i></p> <ul style="list-style-type: none"> <li>tentative evidence of learning having taken place</li> <li>transfer accounted for by change in observable construct</li> <li>no explanation as to why this is the case</li> </ul>	<p><i>High change scores:</i></p> <ul style="list-style-type: none"> <li>evidence of learning having taken place - more assured than with generic dynamic assessment</li> <li>more accurate identification of which intervention process was more efficacious in bringing about change and transfer (near and far)</li> <li>explanation more forthcoming as to why this may be the case</li> </ul>
<p><i>Low scores:</i></p> <ul style="list-style-type: none"> <li>tentative evidence of low level of observable construct</li> <li>possible false construal of high level of observable construct (incorrect answer but partially correct method used to reach answer)</li> <li>no explanation as to why this is the case</li> </ul>	<p><i>Low change scores:</i></p> <ul style="list-style-type: none"> <li>tentative evidence of little or no learning having taken place</li> <li>little or no transfer accounted for by little or no change in observable construct</li> <li>no explanation as to why this is the case</li> </ul>	<p><i>Low change scores:</i></p> <ul style="list-style-type: none"> <li>evidence of little or no learning having taken place - more assured than with generic dynamic assessment</li> <li>more accurate identification as to why certain interventions were not as efficacious as others in various sample groups</li> </ul>
<p><i>Learning:</i></p> <ul style="list-style-type: none"> <li>not applicable</li> </ul>	<p><i>Learning:</i></p> <ul style="list-style-type: none"> <li>equated with 'modification'</li> </ul>	<p><i>Learning:</i></p> <ul style="list-style-type: none"> <li>identified as resulting from genuine modification</li> <li>or resulting from context-dependent intervention referred to as 'compensation'</li> </ul>

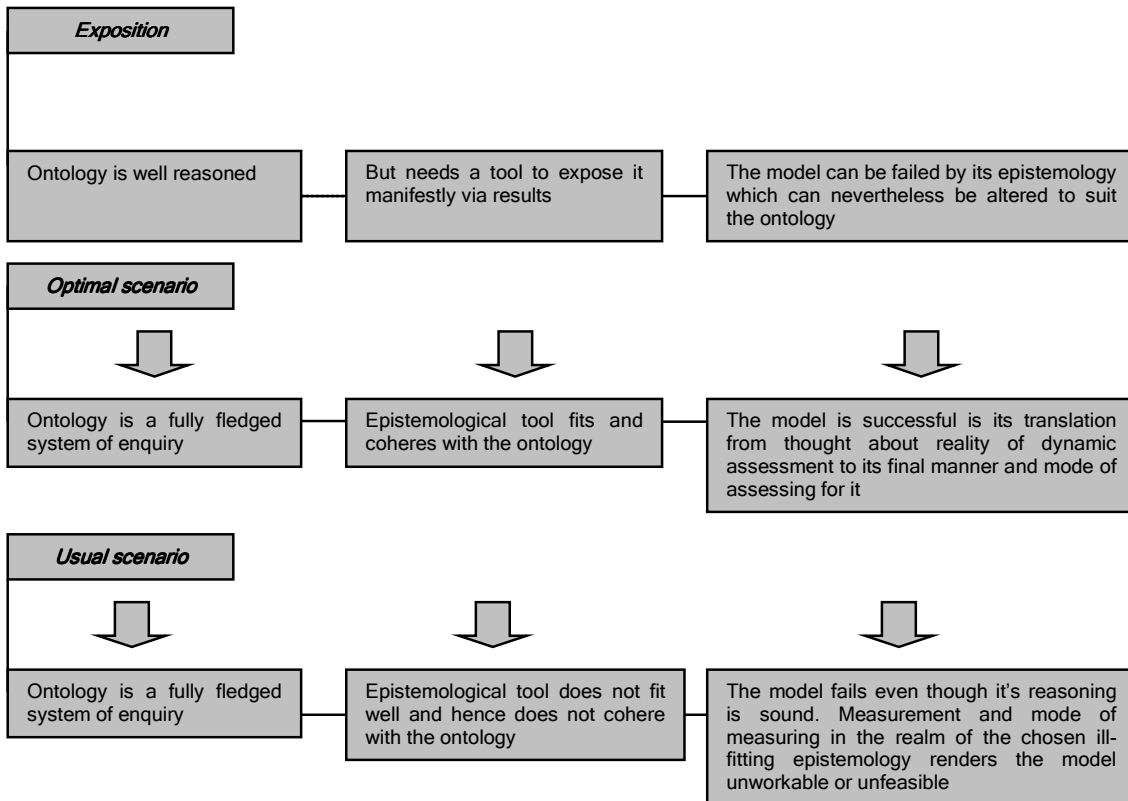


*Compensation:*

- 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 compensation<sup>7</sup>
- 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:



<sup>7</sup> 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)

<b>I Task characteristics</b>	
<i>Levels</i>	<ul style="list-style-type: none"> <li>• practical activity</li> <li>• direct perception</li> <li>• indirect perception</li> <li>• linguistic conceptual knowledge</li> </ul>
<b>II Personal factors</b>	
<i>Structures</i>	<ul style="list-style-type: none"> <li>• epistemic</li> <li>• heuristic</li> </ul>
<i>Components</i>	<ul style="list-style-type: none"> <li>• basic abilities (relationships, seriation, classification)</li> <li>• procedural characteristics (planning, flexibility, self-regulation, activation)</li> <li>• orientation variables (task-specific orientation, self-concept, situation conceptualisation)</li> </ul>
<i>Processes</i>	<ul style="list-style-type: none"> <li>• analysis</li> <li>• synthesis</li> </ul>
<b>III Diagnostic approaches</b>	
<i>Modification</i>	<ul style="list-style-type: none"> <li>• implicit/explicit</li> <li>• predetermined /self-determined</li> </ul>
<i>Compensation</i>	<ul style="list-style-type: none"> <li>• prosthetic</li> <li>• catalytic</li> </ul>
<i>Inhibition</i>	<ul style="list-style-type: none"> <li>• complicating factors</li> </ul>

#### 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 problematic<sup>1</sup> as a variety of operational definitions are often used for the same or very similar theoretical construct<sup>2</sup>. Although ambiguity can be reduced when theoretical constructs are operationally defined<sup>3</sup>, measurement alone leads to its own ambiguities of explanation and prediction if it is independent of theoretical context<sup>4</sup> (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 enterprise<sup>8</sup>
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

<sup>8</sup> 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)

$$\Sigma (H-S) = 2; \Sigma (H-R) = 2; \Sigma (H-H) = 7.5$$

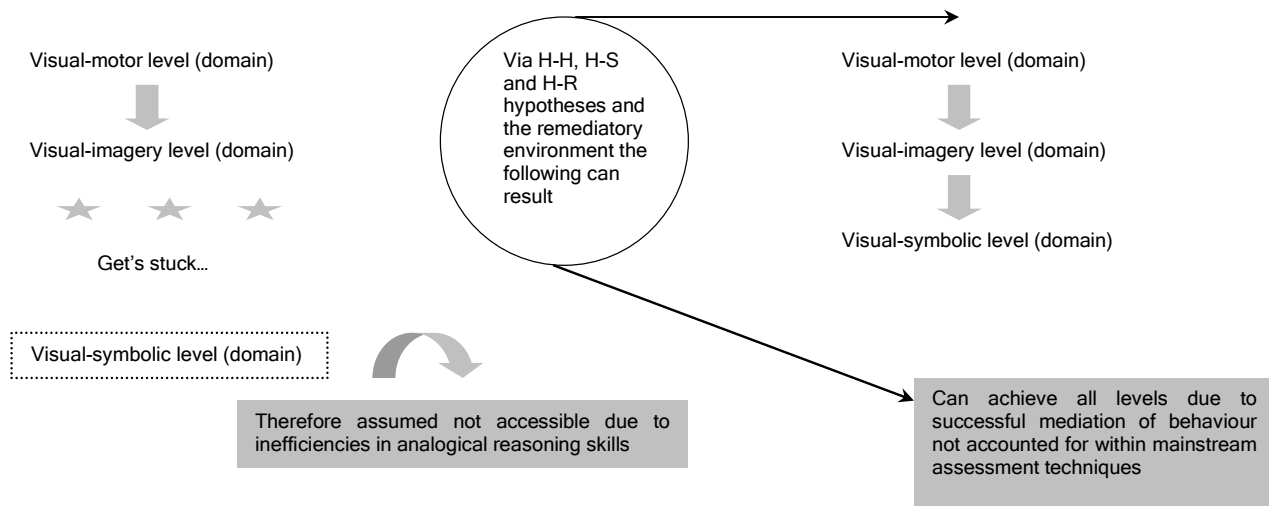
Hence HQ for the approach towards assessing for analogical reasoning =  $7.5/(2+2) = 1.88$  evidencing that the further one moves away from conventional assessment practices the less likely the model is testable. Compare this HQ result to those HQ results emanating from empirically based groundings. 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.<sup>9</sup> 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

<sup>9</sup> 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

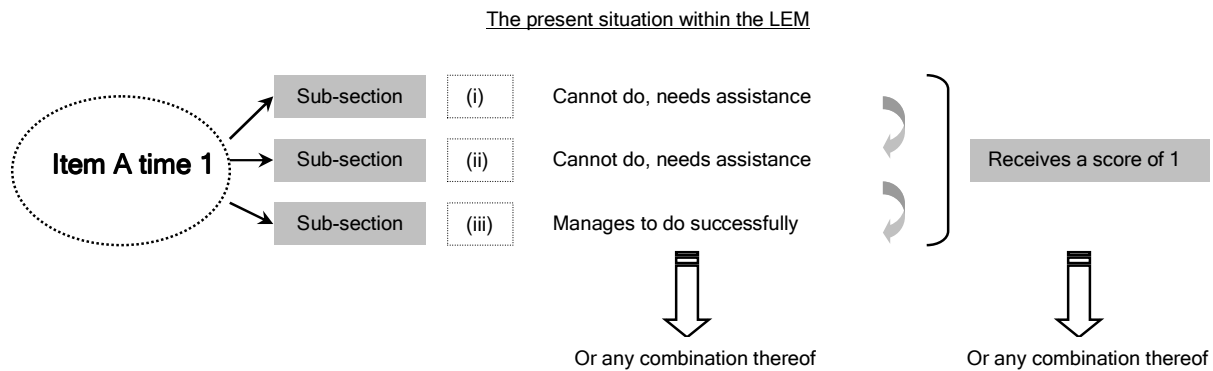
The abstract data theses are already accounted for in the model's description. Recall the abstract data's relevance in terms of its enmeshment with empirical relations.

### 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.

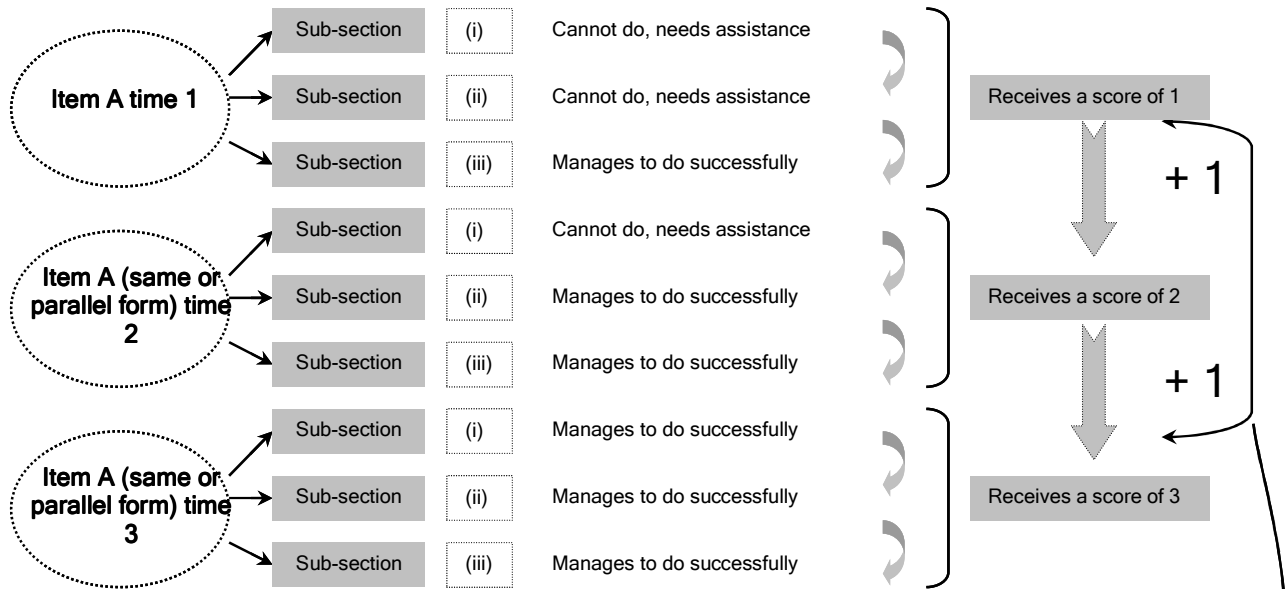
### 5.2.6.8 C(iii) Prime considerations

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.

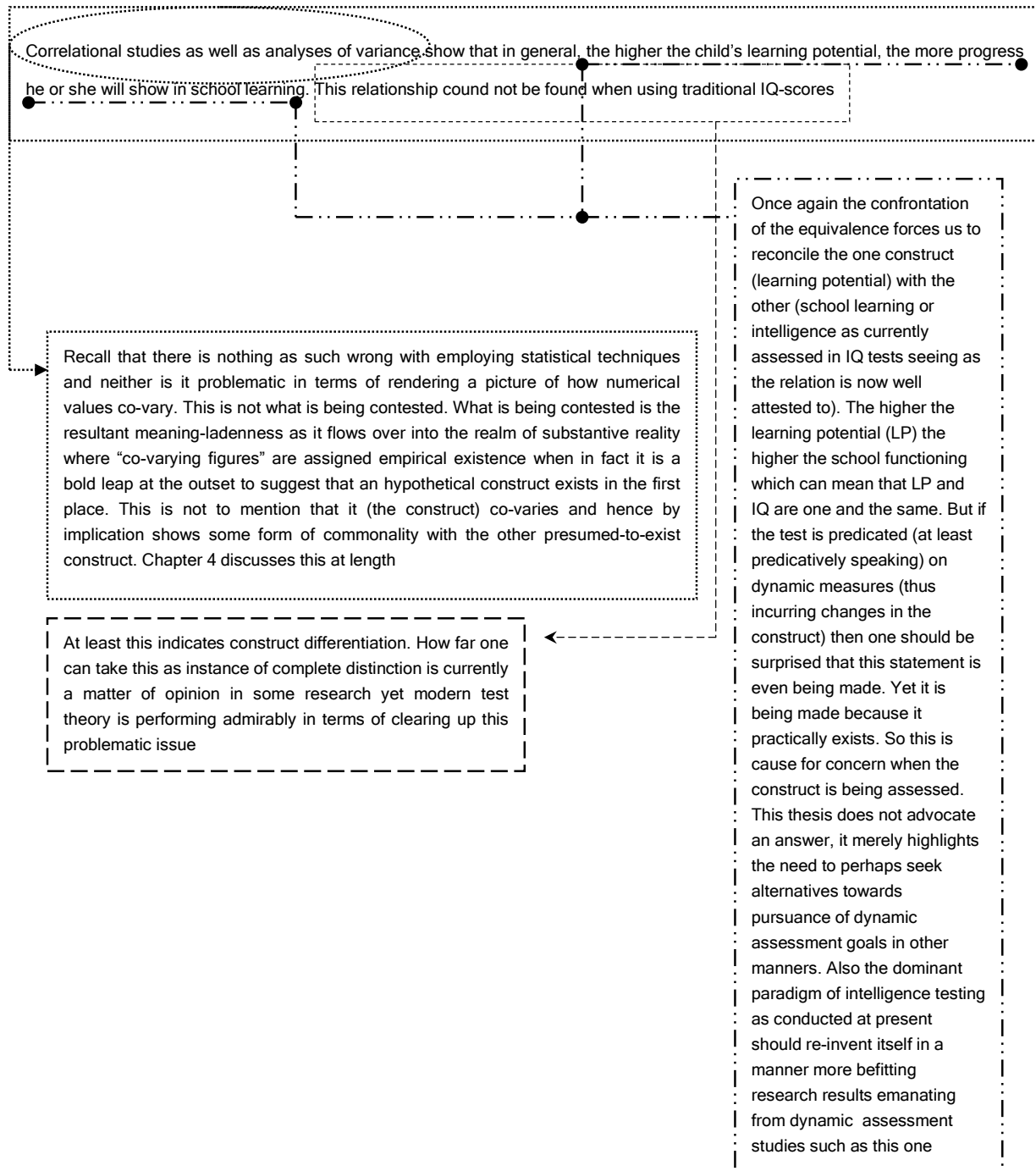
A possible situation within the LEM



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  $b = c$ <sup>10</sup>? Something to ponder.

---

<sup>10</sup> 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)

$$\Sigma (H-S) = 11; \Sigma (H-R) = 7; \Sigma (H-H) = 3.$$

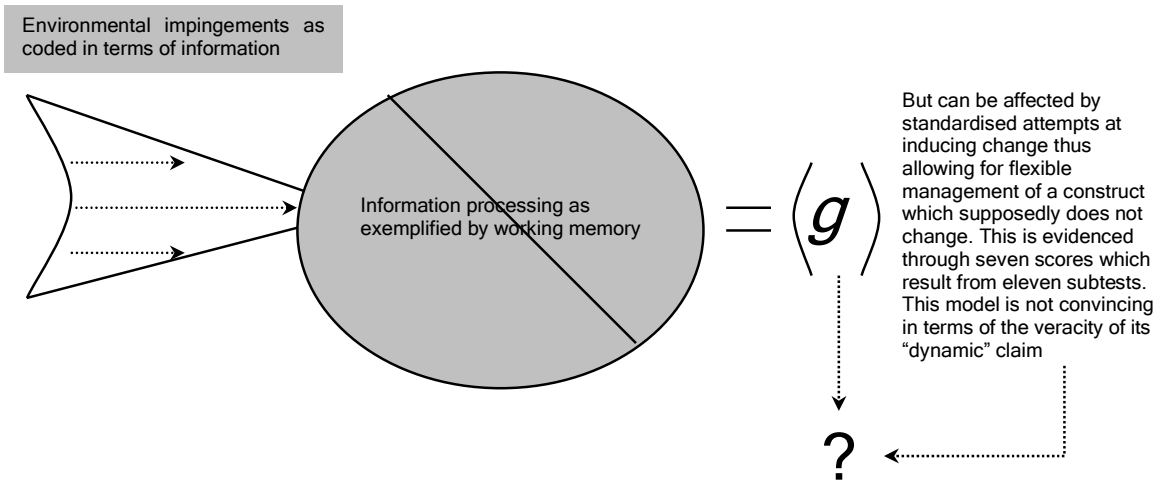
Hence HQ for the S-CPT =  $3/(11+7) = 0.16$  which is not surprising. It is remarkable how such a simplistic rendering of a model by means of the HQ can be predicted by the types of terminology utilised by the authors in their models. The S-CPT is indeed a testable model.

#### 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.<sup>11</sup> 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.

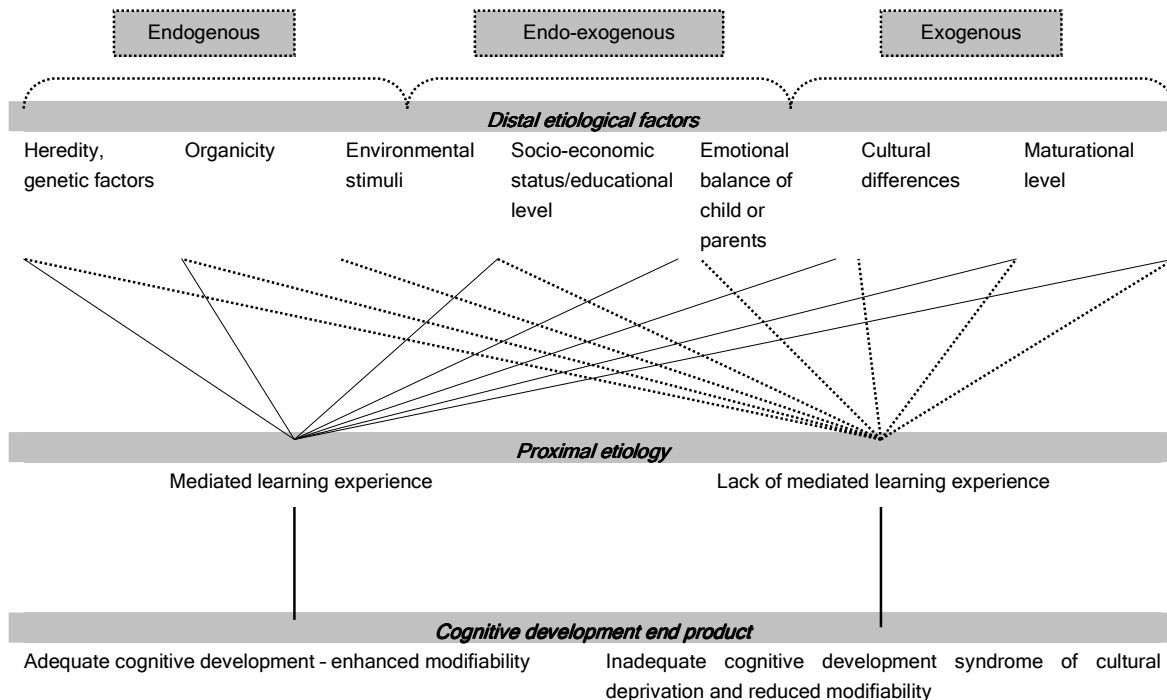
#### 5.2.8.1 A(i) Ontology

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.

---

<sup>11</sup> 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 other 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

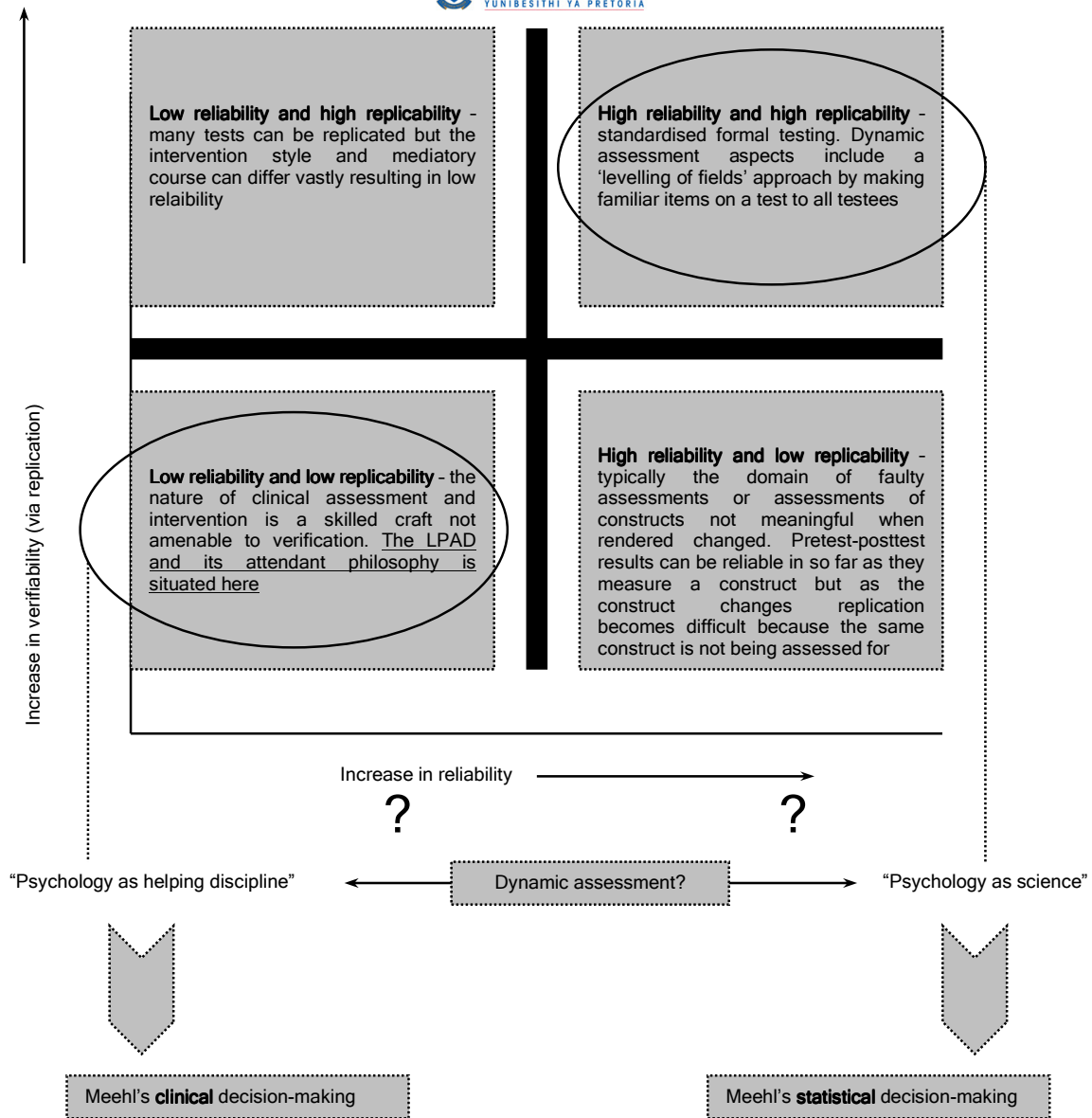


cognitive modifiability (SCM) and mediated learning experience (MLE) and both theory and tool co-ordinate the effort to make change not only theoretically feasible but practicable. The goal for Feuersteinian assessment is to mediate and remediate and not necessarily to assess. Assessment is a core function 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<sub>m</sub> - 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<sub>o</sub> - 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<sub>c</sub> - 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)

$$\Sigma (H-S) = 11; \Sigma (H-R) = 4 ; \Sigma (H-H) = 1$$

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,<sup>12</sup> 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.

<sup>12</sup> But can be found in Feuerstein et al. (2002, p.164).

### 5.2.8.8 C(iii) Prime considerations

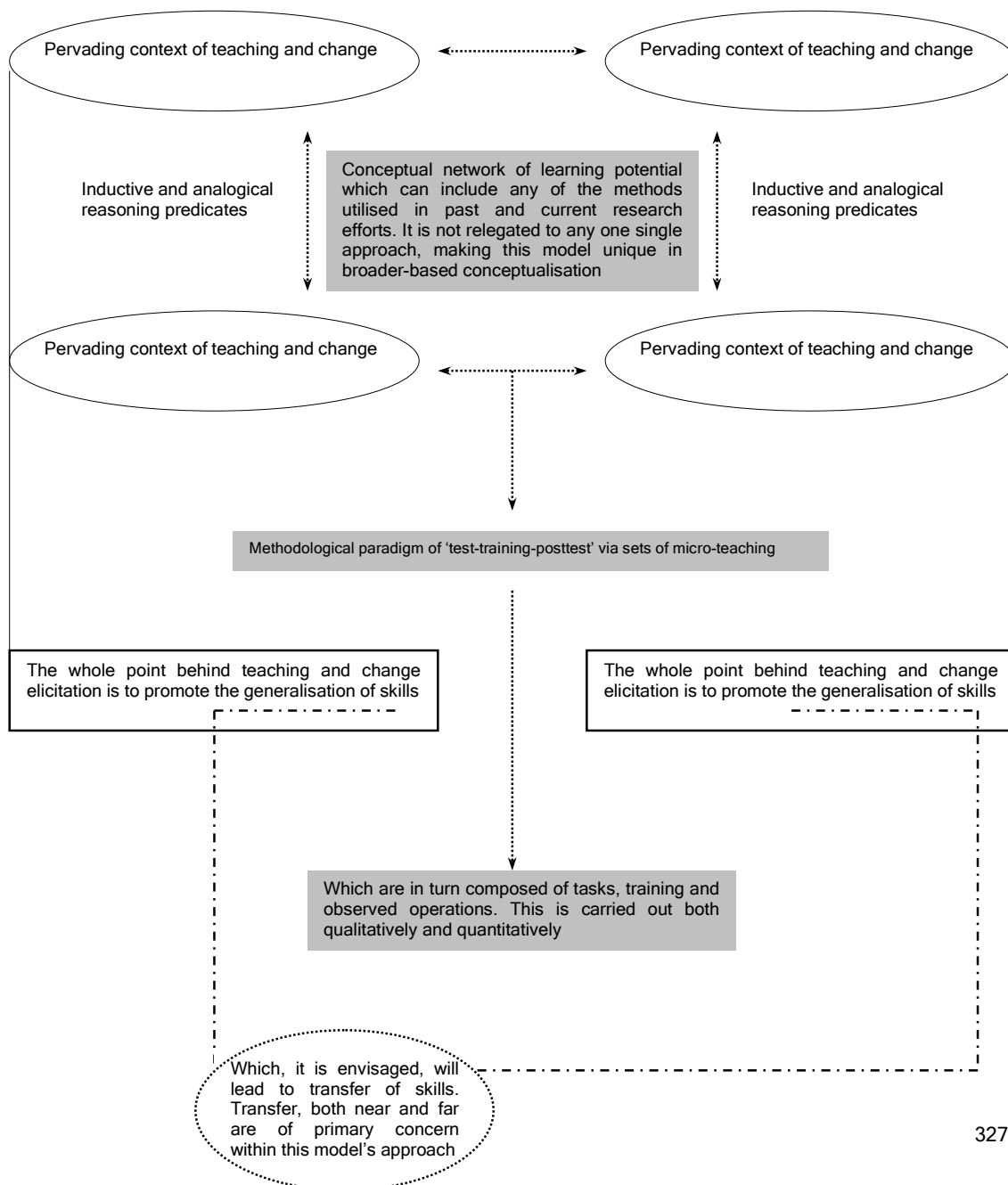
Much of what has been said above in terms of concrete and abstract data can be included in this section.

### 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)?

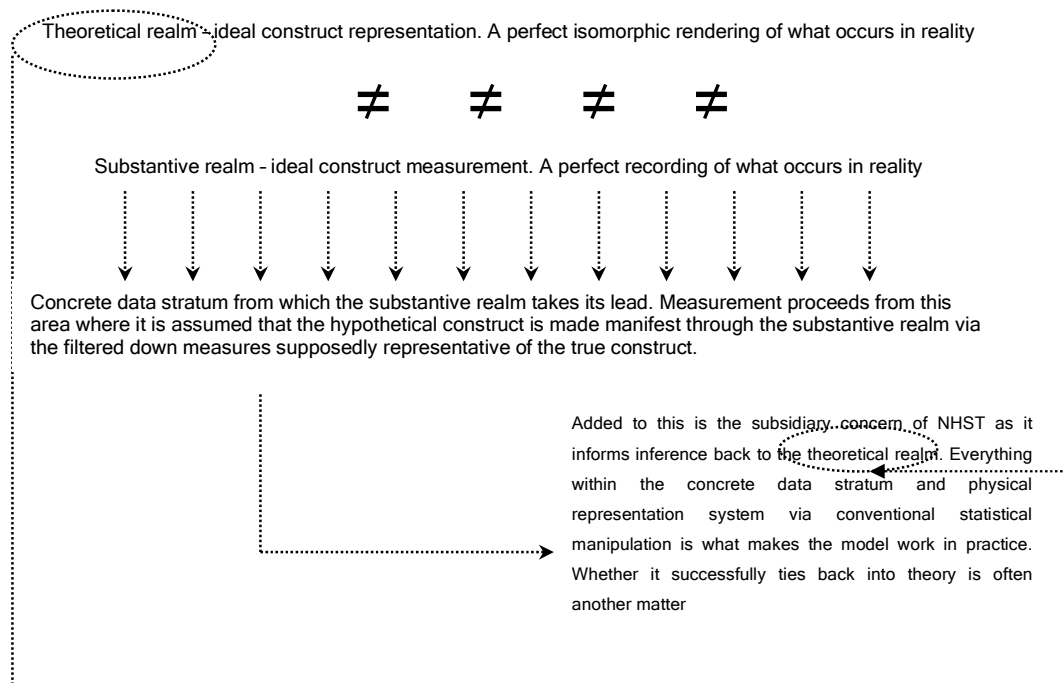
### 5.2.9.6 C(i) Abstract data

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 4.4.2.2.2) 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

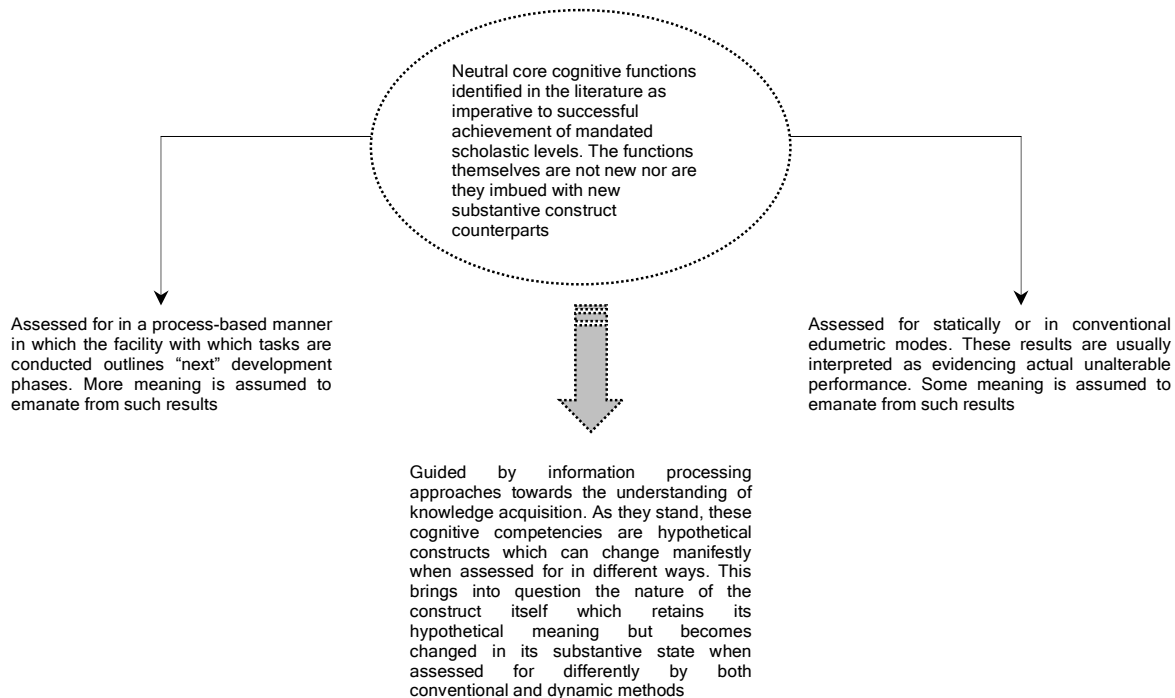
*Conclusion:* the nature of societal concern for school-specific constructs results in many individuals falling behind, not all of whom can be assisted. Dynamically assessing for such specific skill can aid in better educational attainment at school but the generalisability beyond the curriculum-context 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.

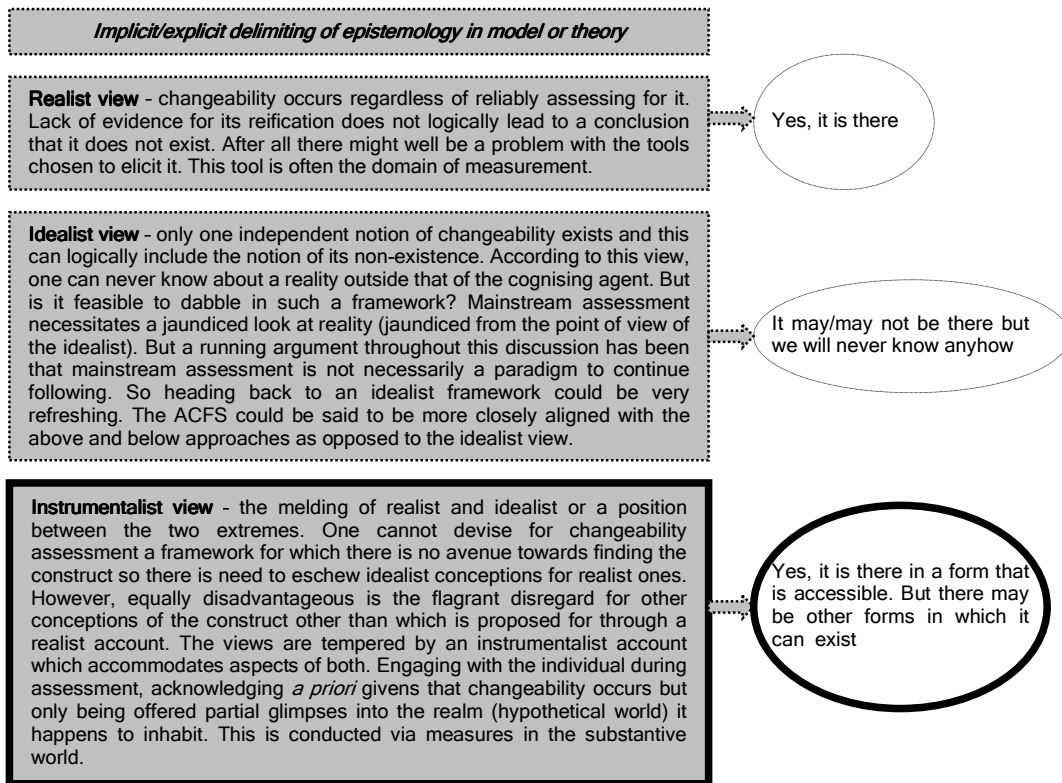


### 5.2.10.2 A(ii) Philosophy

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.

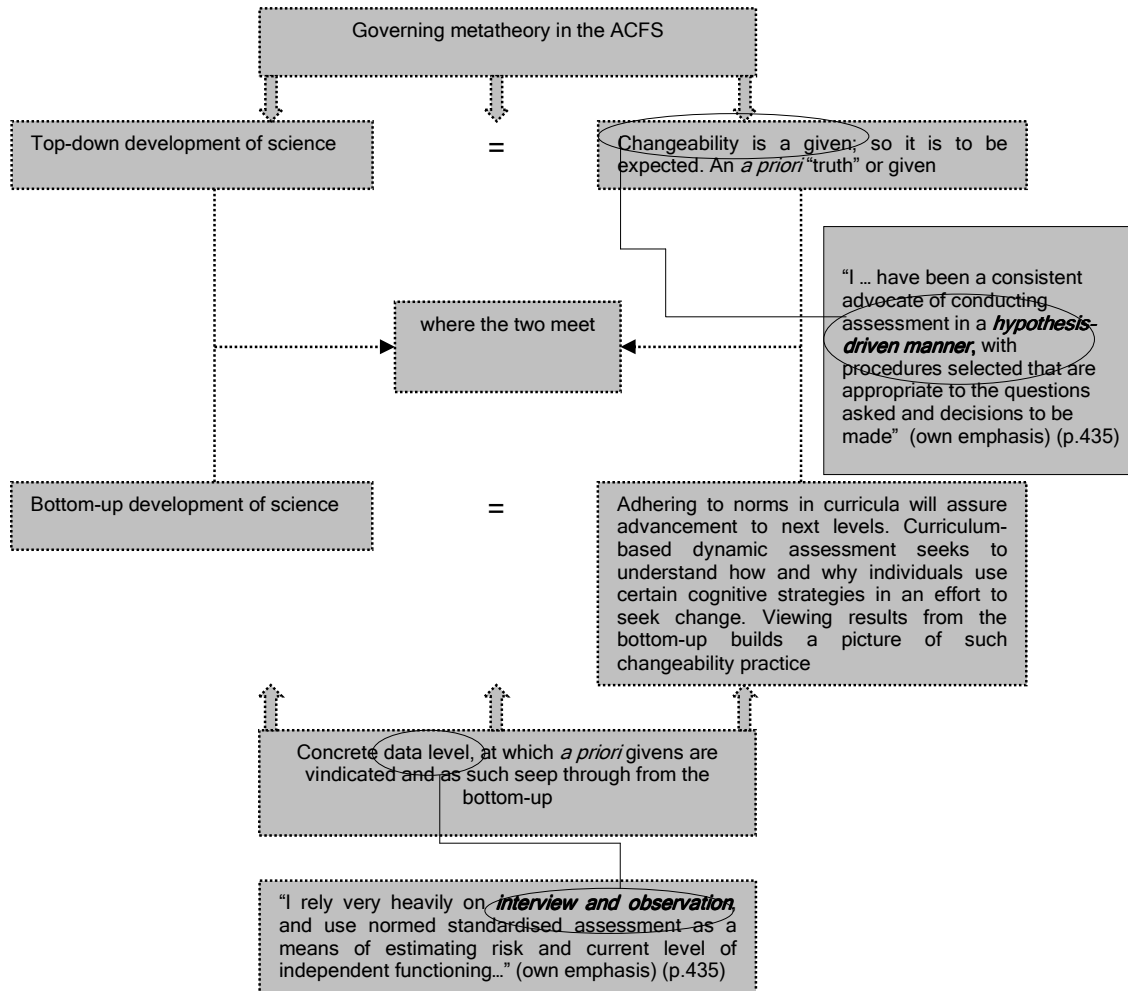
#### 5.2.10.6 C(i) Abstract data

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.



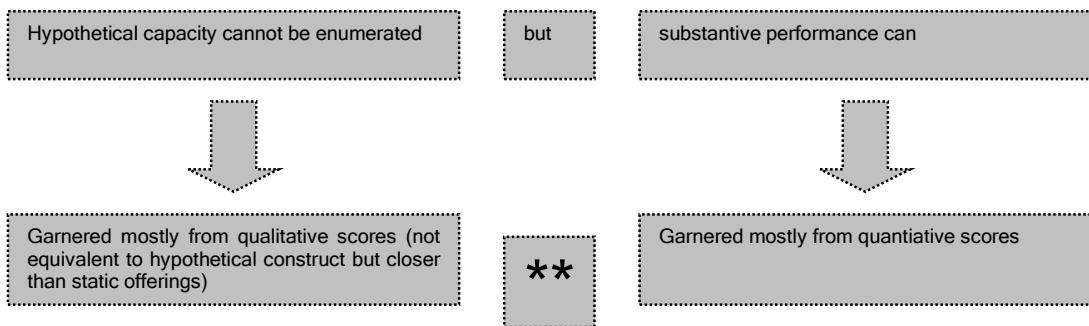
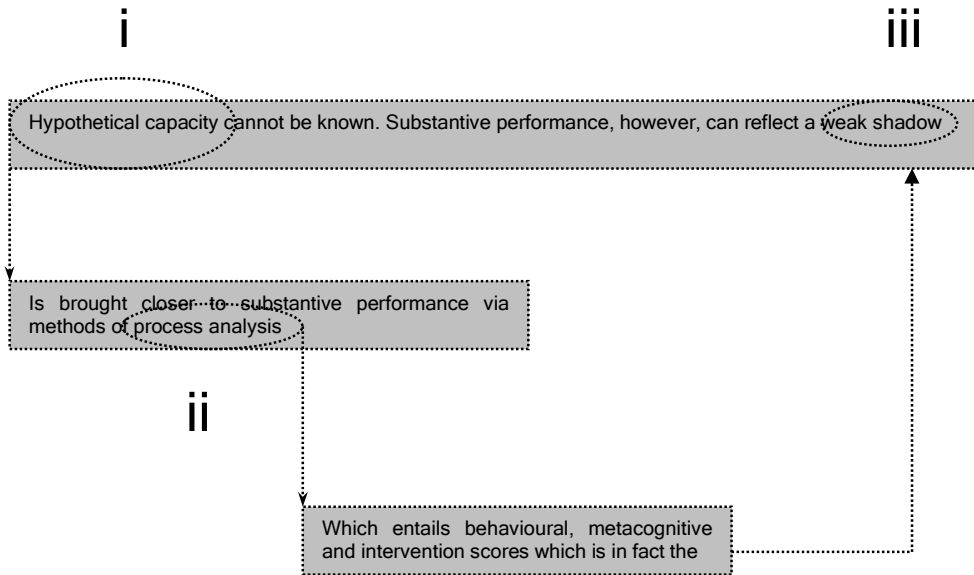
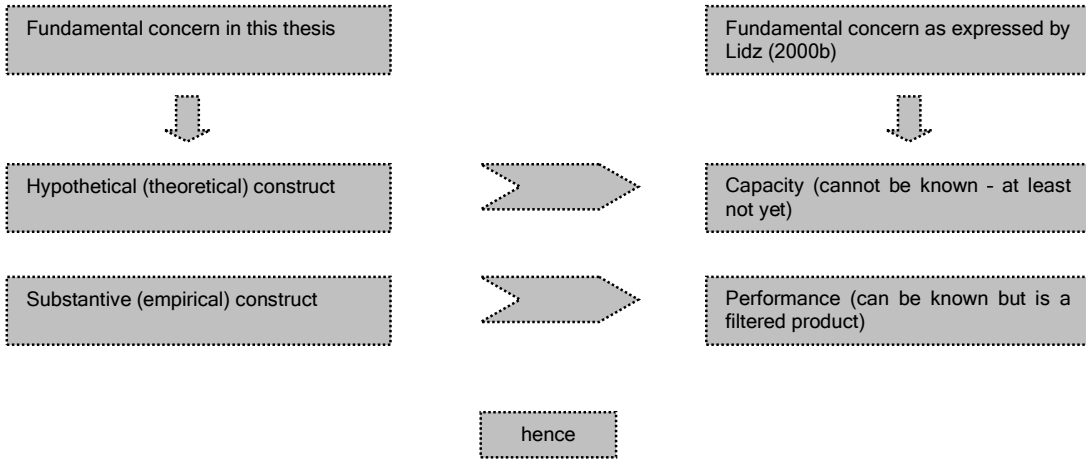
### 5.2.10.8 C(iii) Prime considerations

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"> <li>• Use of raw scores and percent mastery scores on both pretest and posttest items across six tasks</li> <li>• Decision as to what to measure based on “rational task analysis” and observations</li> <li>• Rational task analysis is based on the requisite skill for each task in order to correctly solve it</li> <li>• Error analysis</li> <li>• Cross-task comparison</li> <li>• Approach towards task</li> <li>• Reaction to assessor intervention</li> <li>• More diagnostic approach allows for deviation from semi-scripted intervention</li> <li>• Scores for both task completion and ability to respond successfully to intervention if required</li> <li>• Testing the limits can also be utilised as method if so chosen</li> <li>• Ceiling level is inherent but ACFS it not a battery to be utilised by high functioning performers</li> <li>• Scoring               <ul style="list-style-type: none"> <li>○ Score for accomplishment of each task</li> <li>○ Pretest score (6 scores)</li> <li>○ Posttest score (6 scores)</li> <li>○ Gain score (near transfer) score calculated in the traditional manner of pretest subtracted from posttest (6 scores)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Initially based on observations of children interacting with tasks and agreed upon after inter-rater analyses</li> <li>• Score apportioned to utilisation of metacognition</li> <li>• 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</li> <li>• Summary scores for all pretests and posttests, total scores for each behaviour across each task and total scores for the behaviour rating scale</li> <li>• 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)</li> </ul>

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)

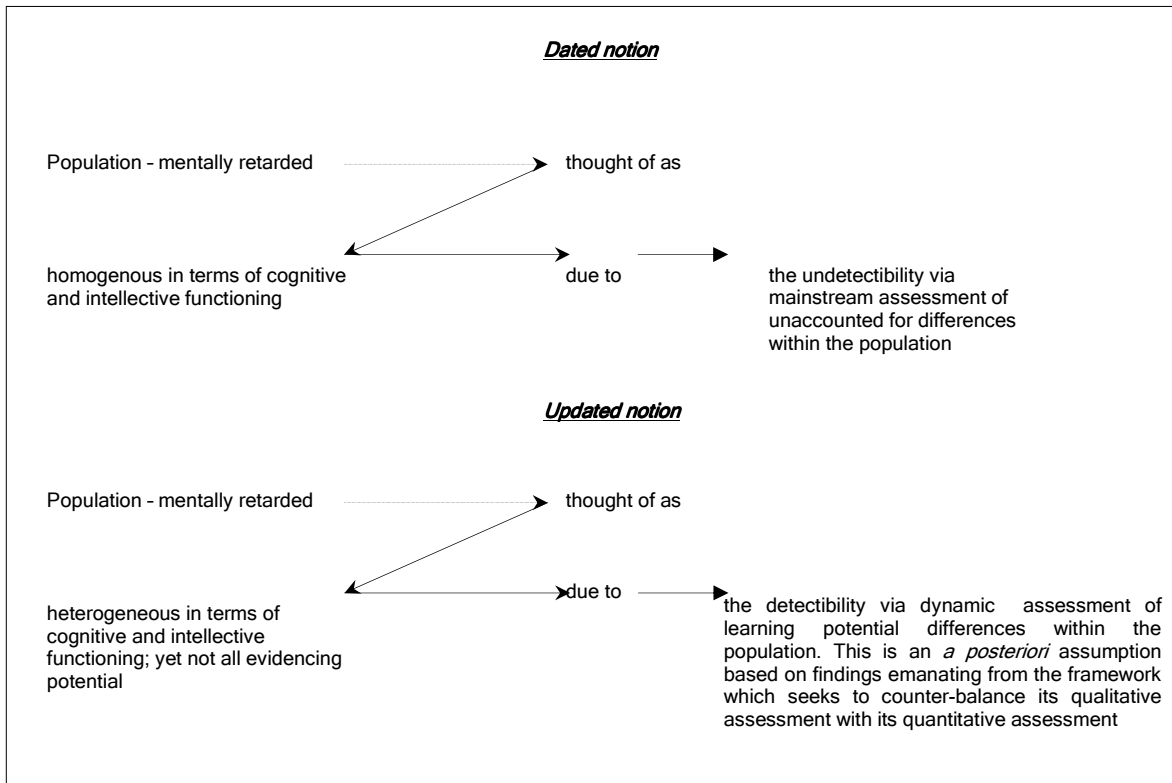
$$\Sigma (H-S) = 5; \Sigma (H-R) = 2; \Sigma (H-H) = 6$$

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

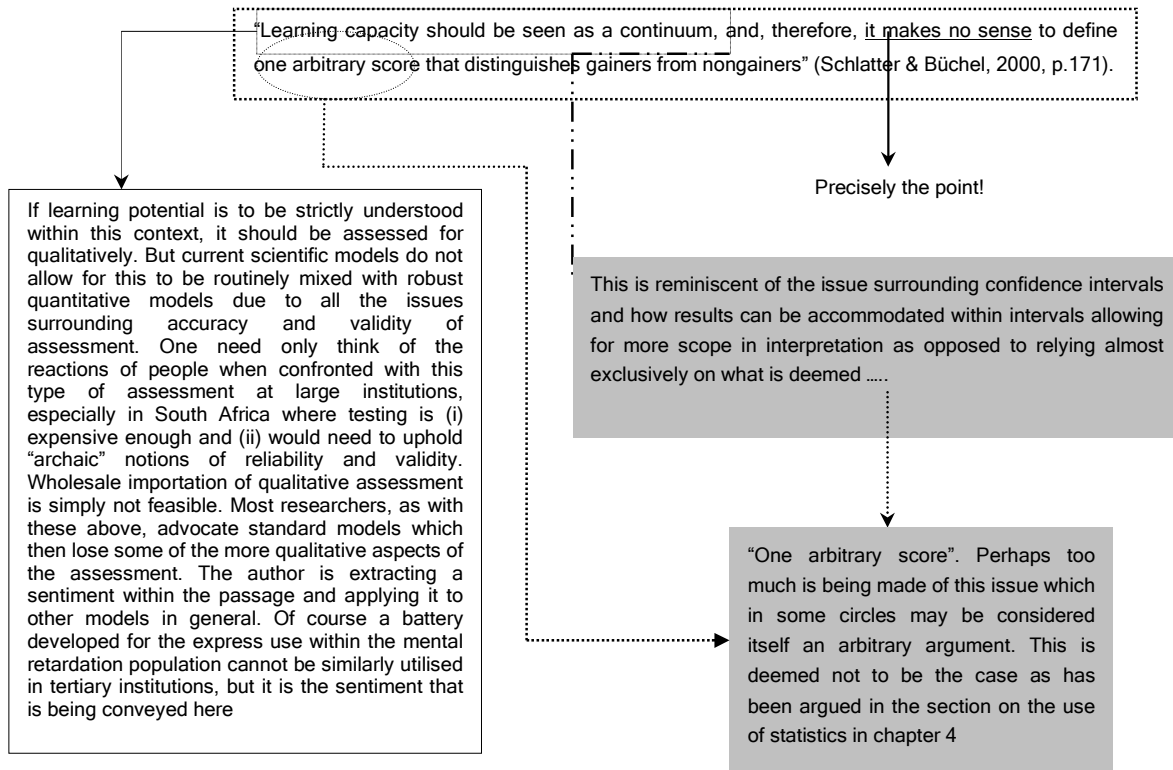
The data stratum houses functional relations between variables more so than correlational variables. The model highlights both observable and non-observable behavioural characteristics especially behavioural traits such as metacognitive aspects. Most of the data stratum variables are accommodated in 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):

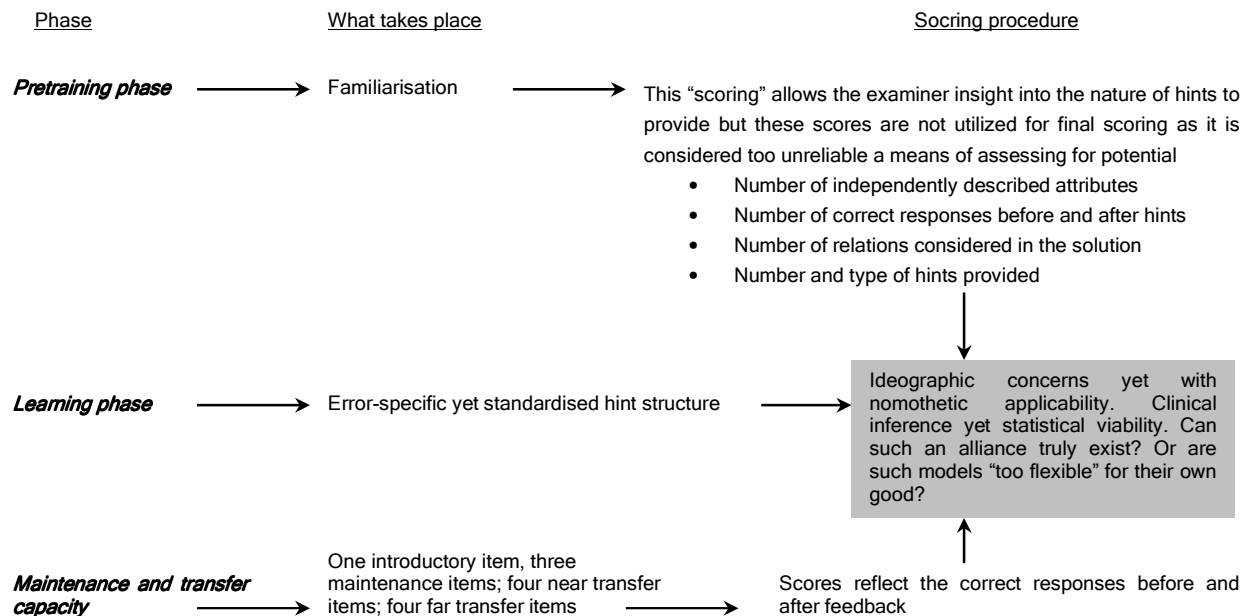
- i. It offers deeper insight into the nature of the problem and allows for individualised and in some instances tailor-made assessment strategies
- ii. Provides information that would have otherwise have been missed by statistical derivation over a number of cases and added to this is the central concern 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:



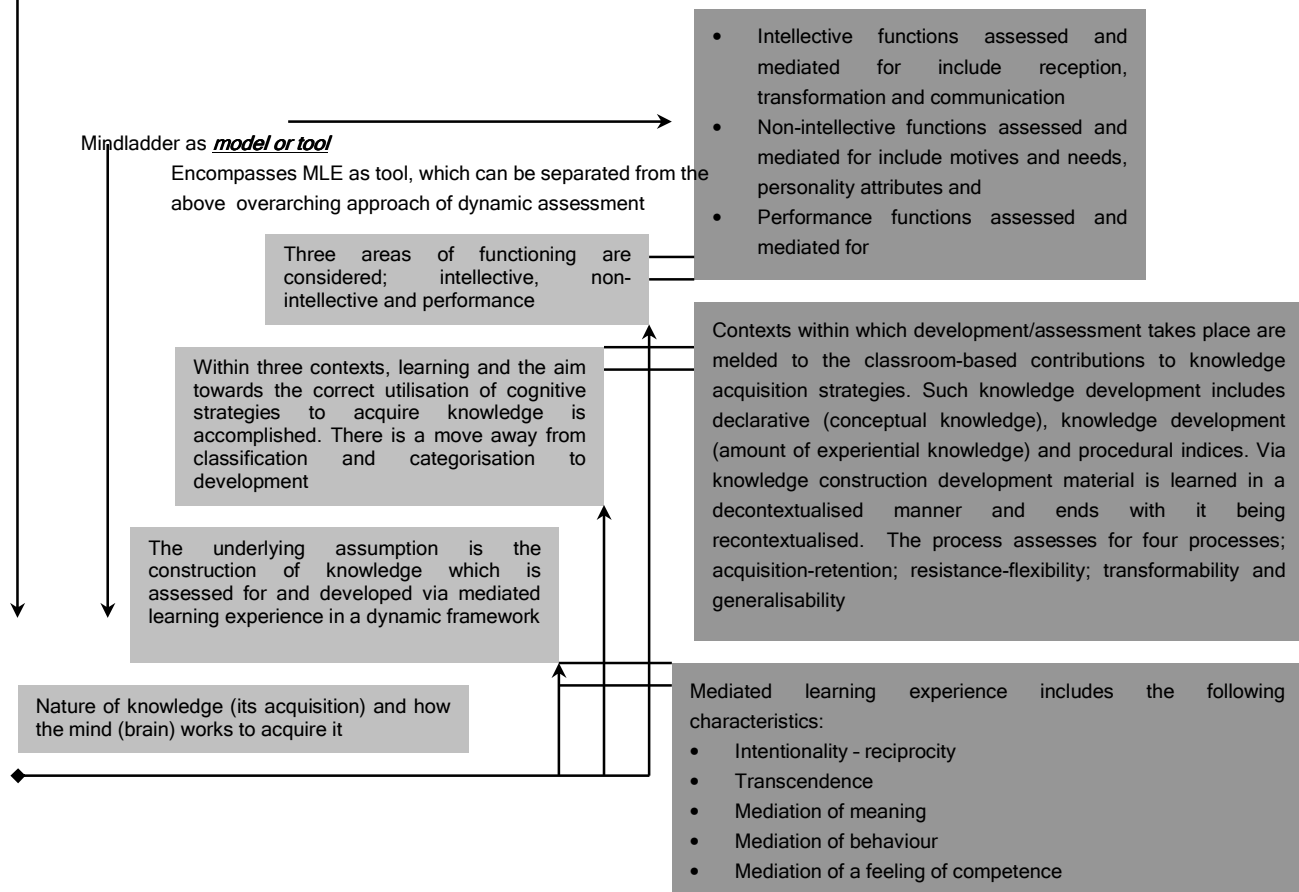
### 5.2.12 The Mindladder model: using dynamic assessment to help students learn to assemble and use knowledge

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 though 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)<sup>13</sup> 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 imbibed 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 intellective 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.

<sup>13</sup> 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 practises. 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.

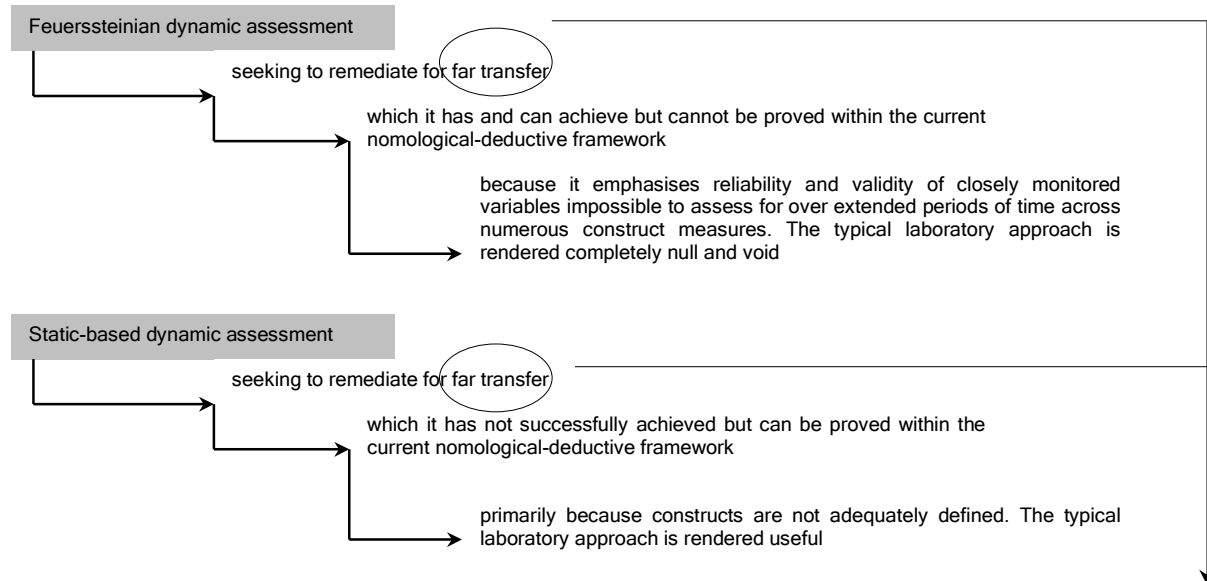
### 5.2.13.8 C(iii) Prime considerations

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)<sup>14</sup></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

<sup>14</sup> 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?