Journal of Cognitive Education and Psychology "Cogito, ergo sum." online

Volume 6, Number 1, July 2006

A peer-reviewed publication of
The International Association for Cognitive Education and Psychology

Murphy, R., & Maree, D. J. F. (2006). Meta-analysis of dynamic assessment research in South Africa. *Journal of Cognitive Education and Psychology* [online], *6*(1), 32-60.
www.iacep.coged.org ISSN 1810-7621

Raegan Murphy

Sunnyside, South Africa

and

David J. F. Maree

University of Pretoria

Meta-Analysis of Dynamic Assessment Research in South Africa

Abstract

The need for tertiary education screening in South Africa has highlighted the role dynamic assessment has to play in this regard, and as no cumulative statistical conclusions can be drawn from single case studies, it was considered timely to do so. In order to address this gap in the field, a meta-analysis was conducted on studies that focused on the efficacy of dynamic assessment interventions within various settings. Fewer research studies dealing with dynamic assessment have been conducted in South Africa as compared to overseas research in the same area. The study served a two-fold purpose: first, to assess the significance of the synthesized effect size from a number of individual studies whose original intention was an investigation of the significance of dynamic assessment interventions; second, to compare two meta-analytic software programs that are freely available online. Small to average effect sizes of 0.3354 and 0.3481 were generated respectively by both programs, with the typical effect size ranging from 0.2–0.8. The method and results of this meta-analysis are discussed along with the limitations inherent in both the programs and we conclude with recommendations for further meta-analytic studies in South Africa within the field of dynamic assessment.

Key words: dynamic assessment, South Africa, meta-analysis

Our first aim in this article is to determine the significance of effect sizes across a number of individual South African studies that use dynamic assessment as a manner of intervention to improve on pretest scores. The second aim is to analyze and compare two separate meta-analytic software programs in terms of their robustness and utility value. During 2001-2002 a study was undertaken to detail the then current research situation in dynamic assessment in South Africa (Murphy, 2002). This study was a narrative exposition on the status of this field of enquiry and did not seek to quantify the results but merely to survey the area.

First, the meta-analysis was conducted on the studies surveyed during 2001-2002 in an attempt to emulate and bring to the field of South African dynamic assessment results similar to those offered by Lussier and Swanson (2002). Studies included herein range from those conducted between 1961 and 2001. Only two studies could be found that had since been added to the field of dynamic assessment in South Africa, one of which included the article by Skuy, Gewer, Osrin, Khunou, Fridjon & Rushton (2002), which would have been a welcome addition to this meta-analysis as the empirical study would have added another independent sample. The studies were located by searching South African databases, such as the NRF database, SABINET, other databases, completed masters theses and doctoral dissertations, electronic databases, research conducted at technikons and universities, as well as South African article searches (SAE publications). SABINET is linked to all South African research institutions and only publications that are indexed in individual library collections are included. Thus, this present search did not locate sources that were not indexed; however, any potential null ''file drawer'' results are factored into the meta-analytic results (as originally highlighted by Rosenthal, 1979).

Lussier and Swanson investigated the degree to which effect sizes, as a function of dynamic assessment as opposed to static assessment, were statistically comparable between ability groups. They also investigated the question of whether the effect size was related to dynamic assessment intervention purely as a methodological artifact or whether the effect size was due to the type of research design, intensity of treatment, and nature of instructions given. Lussier and Swanson sourced PSYCINFO for their database of dynamic assessment intervention studies. Their study yielded more effect sizes than did this current meta-analytic study, due entirely to the paucity of primary studies in this field in South Africa in comparison to the field surveyed by Lussier and Swanson. Their study included English research results other than those conducted in the United States of America. Lussier and Swanson initiated their research with a potential database of 303 articles and refined their criteria to a point that allowed for only 30 articles to be included in their final analysis.

Second, a South African meta-analysis was deemed timely due to the sufficient number of studies available for such an analysis to be conducted and more importantly the study was warranted based on the unknown summarized significance of the quantitative effect sizes (based on posttest score results) that dynamic assessment interventions may or may not have evidenced across studies.

Meta-analysis is a quantitative statistical review technique that summarizes the empirical results of any number of studies (Lussier & Swanson, 2002; Wolf, 1987). Isolated studies can never solve any one particular problem, and the foundation of scientific progress can be regarded as the accumulation of knowledge gathered from the results of many studies culminating in a quantitative synthesis of research (Hunter, Schmidt & Jackson, 1982; Wolf, 1987). Meta-analysis is an independent specialty within statistics and is suited specifically to the calculation of the various effect sizes emanating from individual studies and determination of the significance of the combined effect size (i.e., the cumulated effect size for all samples) (Cooper & Hedges, 1994a). In cumulating each study's effect size, the significance of the overall effect size across all studies can be computed. As meta-analysis is a quantitative statistical technique, it cannot comment on the qualitative value that dynamic assessment as an intervention strategy offers individuals.

Third, South Africa is unique in terms of the need to assess many prospective tertiary education students, the majority of whom can be considered as previously disadvantaged. Of

all learners enrolled in tertiary education institutions in South Africa, 60% are previously disadvantaged students (Department of Education, 2003). In other countries these students are almost always considered the minority in terms of number. There is thus an urgent need to fill the assessment gap for potential tertiary education students, and the focus in many of the studies surveyed here have as their sample such individuals.

Murphy (2002) reviewed 29 studies dealing with dynamic assessment in South Africa. Six were purely qualitative studies and one was a validation study. These studies were thus not amenable to a quantitative synthesis. This left 22 studies that were considered for possible inclusion in a meta-analysis, only 7 of which complied with the necessary requirements for a meta-analysis to be conducted using the two software packages. This resulted in the exclusion of 15 studies. The format for the data necessitated by the software packages led to the inclusion of only between-groups research designs. All 15 excluded studies contained data from within-groups research designs and thus could not be included

The small number of studies eventually included in the study may militate against conducting such a study, yet this leaves one of two options open to the prospective researcher: either wait until more studies avail themselves or conduct a study now and conduct another one at a later date. Nevertheless, not only is the original pool of studies small, but the further delimitation of only seven as the final number warrants due caution for any and all conclusions that are reached in this study. Readers are warned at the outset that the results of this study are to be considered tentative. Further details pertaining to all the studies can be sought from the lead author including all primary study characteristics and their statistical results, as well as more specific information. The question to be answered by this meta-analysis is, Does dynamic assessment intervention make a significant difference as opposed to no (static) intervention across separate studies? Regardless of the significance of the original primary findings, a meta-analysis may well reveal that the synthesized effect size is contrary to the individual study significance.

In this article we also investigate the usefulness of two meta-analytic software programs freely available via the Internet. The programs are compared in terms of their ease of use, documentary user-support, and final analysis that is output as results.

The need for a meta-analysis of South African dynamic assessment

As of 2001, a number of empirical studies utilizing dynamic assessment in South Africa had been conducted yielding results mostly in favor of the efficacy of dynamic assessment interventions when compared to static or no interventions (Murphy, 2002). In order to determine whether the cumulative effect of dynamic assessment was in keeping with the individual case study results, a meta-analysis of these studies was deemed necessary. Second, there was no empirical study, which had as yet investigated the efficacy of dynamic assessment across studies. Third, due to the unique nature and challenges facing South African higher education, where 60% of higher education students are considered previously disadvantaged, the case for the utilization of dynamic assessment as a potential entrance assessment tool becomes an even more urgent one (Department of Education, 2003).

The case for dynamic assessment. Dynamic assessment is a manner of assessing individuals in such a way as to allow for and facilitate improvement and subsequent acquisition of cognitive skills within the assessment situation by engaging the individual within the assessment process (Grigorenko & Sternberg, 1998; Lidz, 1997; Minick, 1987). Its basic

philosophy advocates that individuals are continuously changing throughout life and developing expertise (Sternberg & Grigorenko, 2002). As holistic beings, contextual factors as opposed to genetic factors are perhaps more emphasized as playing a greater role in how individuals cope in life and also how they cope within assessment situations even though heritability cannot be ignored (Das, 1987; Guthke, 1993). Research within dynamic assessment typically has as a research design a pretest-mediation-posttest model (Budoff, 1987; Campione, 1996; Elkonin, Foxcroft, Roodt & Astbury, 2001; Hamers & Resing, 1993; Lidz, 1987; Lidz & Pena, 1996). There are variations on this design ranging from purely clinical interventions (Sternberg, 2000) such as those offered by Feuerstein and Jensen (Feuerstein, Feuerstein, Falik & Rand, 2002; Jensen, 2000) to more robust standardized interventions offered by Budoff and Campione for instance (Budoff, 1987; Campione & Brown, 1987).

Dynamic assessment as a method of testing is uniquely placed in South Africa as the majority of learners in this country have suffered moderate to severe educational handicaps due to past segregationist policies, the results of which are still prevalent (Skuy, Gewer, Osrin, Khunou, Fridjon, & Rushton, 2002). As such, dynamic assessment is considered a method less biased towards the socially disenfranchised (Elliott, 2000) and hence more suitable as a viable alternative to current psychometric tests (Hessels & Hamers, 1993; Sewell, 1987; van de Vijver, 1993). Gains in scores between pretest and posttest South African dynamic assessment interventions have evidenced that, in general, dynamic assessment has proved efficacious as a method of helping individuals improve on tasks requiring skills in varying test batteries (Murphy, 2002).

The current educational crisis within South Africa and the assessment of previously disadvantaged learners and their entrance to tertiary educational institutions is of concern. Dynamic assessment may well prove a viable option as choice of assessment instrument if, as evidenced from South African studies (Murphy, 2002), dynamic assessment does in fact have a significant and sizeable effect. In order to determine this, a meta-analysis was conducted on the current research in South Africa, the results of which will help to inform future policy governing the assessment of previously disadvantaged individuals and their entrance to institutions of higher education. Dynamic assessment is not only a method of assessment but serves in the capacity of a mediational tool that can result in effects other than those obtained in pretest-posttest studies. The importance of the qualitative relationship between the assessor and the testee is one such aspect (Lidz & Elliott, 2000) that cannot, for instance, be measured by only studying posttest scores. However, the individual studies did not assess for this relationship and likewise neither did the meta-analysis.

In order to determine the full efficacy of the mediational aspect inherent in dynamic assessment, means other than those used by cumulating effect sizes across studies are necessary. There is thus a limit to which such a meta-analysis can proceed. Effect size results will not necessarily highlight the total effect of dynamic assessment intervention on posttest scores. They will also not necessarily inform the process as to the overall effect that such an intervention will have on individuals undergoing such intervention strategies. For instance, looking at only posttest scores after sessions of dynamic assessment interventions can in no appreciable way be informative regarding any potential long-term effects of cognitive mediation. More qualitative and long-term research investigations are necessary to determine the fuller impact that dynamic assessment may or may not evidence. The effect size results in this study are thus to be interpreted as evidencing either a cumulative effect or lack of such

effect across studies, but the results do not in any manner reflect the value and nature of the full scope that dynamic assessment interventions have to offer.

However, the original intention within each of the individual studies was to determine the significance of dynamic assessment interventions. The meta-analysis merely reviews this endeavor by synthesizing the effect of dynamic assessment interventions. If results of the original studies led to a conclusion that dynamic assessment did or did not have a significant effect, then by extension this too can be applied to the meta-analysis.

Dynamic assessment in South Africa. Since 1961, a number of studies in South Africa have used dynamic assessment interventions as instruments of mediation in order to have as a result increased scores on pretest-posttest research designs. The results of each study when taken in isolation from other studies yield results evidencing the efficacy of these dynamic assessment interventions. However, in order to investigate empirically whether cumulative efficacy is apparent across all studies, a meta-analysis needed to be conducted.

Murphy (2002) reviewed South African dynamic assessment research (1961-2001) and used the primary studies highlighted in the research for purposes of this meta-analysis. Making use of vote-counting, Murphy (2002) concluded that of 29 primary empirical studies 21 revealed that dynamic assessment interventions indeed significantly improved scorers over pretest levels. Bushman (1994) observed that "when effect sizes are medium to small, the conventional vote-counting procedure frequently fails to detect the differences (p. 195). (Two primary empirical studies yielded nonsignificant effects evidencing lack of support for the notion that dynamic assessment interventions significantly improve posttest scores. Six studies yielded confounding results, evidencing both significant and nonsignificant results (i.e., in these studies the same sample was used for more than one experiment). The study concluded, that, based on these findings, the mediational intervention in dynamic assessment in South Africa was efficacious in bringing about significant change from pretest to posttest.

As is at times the case, a meta-analysis will either reveal significant effects across cumulated studies or nonsignificant effects (contrary to those effects evidenced within individual studies). In order to determine whether dynamic assessment was efficacious, an empirical meta-analysis was conducted to either further support the original conclusions in Murphy (2002) or to caution against possible inferences made from the conclusion.

A brief tour of meta-analysis

In essence, meta-analysis seeks to cumulate findings across primary studies, analyze the combined findings, and derive conclusions from the total number of studies. In so doing, it may happen that results counter findings in the primary studies. In other words, what may have seemed to be an effectual experimental intervention in a study may not in fact contribute much proportionally on a larger scale, thus nullifying the original results. This is not necessarily the case in all studies, but such findings are not outside the norm. It may happen that experimental results are indeed robust in terms of results when compared to control groups for individual studies yet when cumulated may result in even greater yields in terms of effectiveness. Meta-analysis typically finds its niche in studies designed to test the differences between experimental and control groups but is not limited to such designs (Chambers, 2004; Kulik & Kulik, 1989). The *d* family of effect sizes is used in this study including Hedge's *g*, Glass' delta and Cohen's *d* (Rosenthal, 1994), which makes use of both

control and experimental groups (Schwarzer, 1989; Strube & Hartmann, 1983). This fact, along with the fact that the two software programs utilized in this study also made use of this family of effects size statistic, led to the preference in this meta-analysis to locate studies with between-groups designs only (i.e. studies with comparisons between experimental and control groups).

The name for this technique was first introduced by Gene Glass in 1976 (Chambers, 2004) and as such is quite recent in terms of statistical methodology; however, Pearson had already, in 1904, taken the average correlation results of medical studies and utilized them in research with similar techniques being used throughout the early half of the twentieth century (Bangert-Drowns, 1992; Cooper & Hedges, 1994a). The need to allocate an effect size for each study in terms of its overall contribution to the final result was an outgrowth of behavioral scientists' need to summarize large databases of literature in as systematic a fashion as possible (Rosenthal, 1979, 1995). The quantitative generalization of such a systematic investigation into the results of many primary studies would seem to offer more value or the "strengthening" of methods (Hall, Tickle-Degnen, Rosenthal, & Mosteller, 1994) in terms of strategic recommendations based on such findings (Arthur, Bennett, & Huffcutt, 1994; Cooper & Lemke, 1991; Hunter, Schmidt, & Jackson, 1982; Strube, 1985; Wolf, 1987). This is in addition to the usefulness of narrative reviews of studies (Strube & Hartmann, 1983). The above-mentioned 2001-2002 narrative study could be crudely construed as a vote-counting method of sorts (Bushman, 1994; Schwarzer, 1989) and hence the need arose to test and quantify similar hypotheses results across independent studies (Kalaian & Raudenbush, 1996). Combining independent studies sharing the same or at the very least similar hypotheses is the necessary requirements when conducting a meta-analysis. Combining studies at will without due consideration of various hypotheses is blatant nonsense. Eysenck (1995) pointed out the folly of unthinking use of meta-analysis in just such a scenario. The scheme used here is a normal advance in any area of research interest, signifying the entry of the particular area into mainstream research territory and alerting the reader to a new body of research (Myers, 1991). The coding scheme for this analysis (Orwin, 1994; Schwarzer, 1989; Stock, 1994) is available on request.

Synthesizing research can never replace the need for reading original sources, and this statement is made apart from the fact that it is merely good practice to do so, but is stated more as a result of the number of discretionary steps taken by meta-analysts when deciding what to include and what to leave out in their final analysis (Arthur et al., 1994). For instance, although fail-safe computation has partially addressed the problem of excluding null outcomes in the meta-analysis (a result partly due to publication bias favoring significant findings for example; Begg, 1994), the onus rests with the researcher to locate any and all information pertaining to the area of study (Glass, McGraw, & Smith 1981; Strube & Hartmann, 1983). It is not the opinion of the authors to obviate the need to study previous research results regardless of the findings of the overall result, as has been alluded to by David Hilbert, albeit in a somewhat different context (Glass, McGraw, & Smith, 1981). The researcher has to decide on the criteria for inclusion of studies, the model assumptions to be used, the use or lack thereof of programs individually tailored to the needs of the particular meta-analytic study, the necessity of inter-rater reliability when coding large numbers of studies, and much else besides (Dickersin, 1994; Hunter et al., 1982; Reed & Baxter, 1994; Rosenthal, 1994; White, 1994; Wortman, 1994).

Meta-analysis approaches are often found to emanate from one of two major groupings; namely, the combination of significance levels and the combination of effect sizes, the latter

being used in this particular study (Strube & Hartmann, 1983). Combining statistical significance levels indicates the degree to which chance plays a role in the findings whereas the combination of effect sizes examines the magnitude of effect across studies (Becker, 1994; Shadish & Haddock, 1994; Wolf, 1987). Two "families'' of effect sizes are available to the meta-analyst and include the *r* family (which also includes *Zr*, Fisher's transformation or *r*); and the *d* family, which includes Hedge's *g*, Glass' delta and Cohen's *d* (Rosenthal, 1994). The latter family of statistics is used within this analysis.

Method

Two-fold purpose of the meta-analysis

The overriding reason for running the meta-analysis was to impart to the field of dynamic assessment pertinent information regarding the scope of dynamic assessment in South African research. Moreover, it was thought prudent to run the analysis on two separate computer software programs in order to highlight advantages and disadvantages of these programs so as to offer the reader a choice of application should further analysis be undertaken. Statistical packages such as SAS and SPSS run standard statistical techniques used in the behavioral sciences and are thus fairly widespread; consequently, details of statistical runs are not often discussed in research reports.

The same cannot be said of meta-analytic packages, which do not run as a program or subroutine within SPSS or SAS, although macros and commercial programs are becoming available. Information pertaining to the freeware as well as SAS and SPSS macros are available at the following comprehensive website as of April 2005: www.um.es/facpsi/metaanalysis/software.php. Some of these programs are reviewed in the professional literature (cf. Arthur, et al., 1994; Normand, 1995). Initially most meta-analytic software was available only for mainframes and not microcomputers but this has now changed (Arthur et al., 1994). Standardized packages are, however, not yet the norm as each package assumes various models as well as theoretical and conceptual underpinnings (Arthur et al., 1994). It is for this reason that these packages are compared and detailed in their functioning. This section thus serves two purposes: to assess the primary study results cumulatively across studies and to evaluate the usefulness of two software programs. Comparison of meta-analytic programs has in the past offered researchers the opportunity to make informed decisions when deciding on the use of one program over another (Arthur et al., 1994; Normand, 1995).

The choice of two programs in particular, namely, "META – Easy to answer'' version III by D. A. Kenny and Meta-analysis program version 5.3 by R. Schwarzer, was made due to their availability and cost (they are both freeware products) and their accompanying recommendations made by peers within the field. Both programs were fairly small to download (91 and 212 kilobytes for both the Kenny and Schwarzer programs respectively), run hassle-free within the Windows environment, and have fairly good to good manuals that accompany the software. The author of the first program (Kenny) is also available for questions about his program. No information was sought from Schwarzer. The two programs are now discussed, after which data analysis of the meta-study follows.

“META – Easy to answer’’ version III by D. A. Kenny

This software program was developed by David A. Kenny at the University of Connecticut, United States, and is a compiled version of a QuickBasic program with a DOS-like appearance, which runs in the windows environment. This is a shareware version offered free of charge and can be downloaded over the Internet from the following address: http://davidakenny.net/meta.htm. Kenny (2003) cautions the user, however, as to its as yet demonstration status and the user is advised to check computational output. This cautionary note propelled further the need for a second program’s analysis, and a double-check of sorts was conducted to compare output. This program computes effect sizes for each study, pools the results and calculates the degree to which the result differs from zero, and tests for homogeneity of effect sizes across studies. It allows for the weighting of studies based on sample size, variance, or user-inputted values.

Program structure. The program encompasses three stages. The first stage seeks overall study information and prompts the user for information such as the chosen statistic for effect size computation, study group characteristics, sample numbers, and degrees of freedom. The second stage seeks the test statistics utilized in the primary data, which are then inputted. The user has to write out by hand the results of each study, should they find this necessary, as this demonstration edition offers neither saving nor printing options. The third stage integrates the input from the first two stages. The results of the third stage are in fact the meta-analytic results of the combined studies. Once again the user has to write out by hand the results of the final analysis.

The data that are entered into META are placed into a data file, which is then used by META to compute the final analysis. This data file consists of numeric characters but occasionally alpha-numeric characters are used. The data file is easy and flexible to manage and changes can be made directly in the data file, although this is not recommended unless the user understands the derivation of all the computations. Studies can be deleted and weightings changed as desired by going through the data within the META program itself as opposed to changing the data within the data file. Both individual study results as well as the final output has to be written out by hand as there is no print and save option available.

Meta-analysis program version 5.3 by R. Schwarzer

This program was developed by R. Schwarzer at the Freie Universität Berlin, Germany and was written in Turbo Pascal 5.0. The program is not public domain but is distributed under the User Supported Software concept (Schwarzer, 1989) and can be downloaded from http://www.RalfSchwarzer.de. Schwarzer’s program allows for the computation of probabilities, effect sizes *d*, and effect sizes *r* (correlations). Depending on the data available to the researcher, any of these three can be selected. Also available is a data editor and a number of utilities that provide transformed data results.

Program structure. The general menu allows the user to access the data file (editor) that has to be entered according to a specific format (depending on which data are entered, i.e., correlations, proportions or d values). The program reads the data file and computes the final result that can be saved and/or printed. The individual results, however, have to be hand written. For this study, the “utilities’’ menu was used to compute the individual study effect sizes based on group mean and standard deviation. The data file can then be assembled.

Depending on which test statistic is available for computation of an effect size and also depending on how the final analysis will be run, Schwarzer's program requires that the data be in certain formats. Effect size computation using *d* facilitates the use of up to 10 groups. For instance, when computing for *d* values, the data file needs to include the study number, sample size for group *n*, sample size for group *s*, effect size, and a reliability coefficient. As reliability coefficients were not available in the primary studies, unity was maintained throughout the studies by inputting 1.0 as suggested by Schwarzer (1989). For probabilities, study number, sample size, and *p*-values are needed; for effect sizes *r*, the study number, sample size, as well as correlations and the variables' reliabilities are needed. In essence, after having computed the effect sizes for each study using the utilities menu and compiling a data file, this program runs the file according to the chosen statistic. Schwarzer (1989) maintains that meta-analysis of effect sizes are superior to those using only combinations of probabilities.

The "effect sizes *d*'' option seeks the number of groups to be compared and also whether standard deviation or variance is available. The "significance of correlation'' option seeks correlation values for the chosen number of groups. Means, variances, and correlations can also be inputted and weighted. Lastly *t* values can be computed for the chosen number of groups. As is evident, this program offers a variety of statistical manipulations for the chosen transformation statistic and is flexible in terms of the data available from the primary research, Schwarzer being cognizant of the broad variety of available primary statistics to the meta-analytic researcher.

Comparison of the two programs

Both programs output similar results in terms of individual studies, although Schwarzer's program is more comprehensive and offers more variation in types of output. Kenny's program offers less variation in output results. During the input stage, Kenny's program requires more information per study, integrates the data file with the output, and allows for a more comprehensive data file. Schwarzer's program is not integrated in similar fashion and does not require as much detail during input. However, Schwarzer's program allows for very comprehensive transformation utilities that can then be used within the data editor for the meta-analysis of choice (either analysis based on *d* values, *r* values as well as cluster analysis output and stem-and-leaf displays). Stem-and-leaf displays were not used for these data, as there were too few effect sizes for a reasonable display; moreover, these displays are better suited to correlation values.

Unlike Kenny's program, in which the data input and calculation take place in seemingly one step, Schwarzer's program requires the user to first compute effect sizes using the "utilities'' option. These results are then recorded by hand and typed into the data file. The final analysis runs the data file. Two separate steps are thus necessitated. For input, Kenny makes use of an effect size that is the equivalent of Schwarzer's ''*g*,'' which is the effect size based on pooled variance. Kenny refers to Schwarzer's ''*g*'' as ''*d*,'' which can lead to some confusion, this being acknowledged by Schwarzer (1989). Kenny makes use of Hedge's unbiased estimator *d* in his meta computation. There is thus a slight difference in the numerical value of the statistic used by both programs.

Individual study results are, however, exactly the same. Kenny's program in essence works on a random model principle (D. Kenny, personal communication, April 11, 2005) and Schwarzer's program presents both fixed and random-effects model results. Fixed-effects

models assume that any differences between samples are due strictly to sampling error with an average effect size simply being an unbiased estimate or simple average of a population effect. Random-effects models assume that differences may also be attributed to aspects other than sampling error, with the assumption that the sample has been drawn from a population; thus there is not only one population effect size but a distribution of population effect sizes, resulting in sample characteristics that are not only dissimilar but also reflect true underlying population differences (Cooper, 1994; Cooper & Hedges, 1994b; Normand, 1995; Schwarzer, 1989; Shadish & Haddock, 1994). Among other indicators, residual variation includes results from chi-square analyses and tests of homogeneity (Chambers, 2004), the significance of which will prompt further investigation into random model usage. Kenny's program, based on a random-effects model, assumes that the study is used as the sampling unit (D. Kenny, personal communication, April 11, 2005; Kenny, 2003), unlike fixed-effects models that use as their sampling unit, individuals within the studies (Rosenthal, 1995).

The manuals that accompany both programs offer the necessary and requisite information in order for correct data input and knowledge of how the program functions within the operating environment. Both allow printing options on only some menus and are up to date and consistent in terms of current research into the statistical area of meta-analysis. Both manuals enable the user to perform the necessary computation in order to obtain output. Schwarzer's manual is, however, more comprehensive, serving as an introduction to and brief overview of the field of meta-analysis. It also elucidates the statistical formulae used within the computations themselves that Kenny's manual does not offer.

Schwarzer's program also offers more variety in terms of output, such as cluster analysis for both *d* and *r* values and visual display of effect sizes that Kenny's does not. Cluster analysis allows the user to search for potential moderating factors that present themselves in terms of how the effect sizes are clustered. Kenny's manual states that the researcher manually looks for moderator variables. The manuals are available for downloading at the same above-mentioned web addresses that are accessed to download the programs. It is advisable that the researcher studies both manuals before attempting to use either program.

Limitations of the programs and violation of assumptions

Neither Kenny's nor Schwarzer's program can handle within-groups studies nor repeated-measures designs and it is for this reason that only between-groups studies were included in the analyses. This criterion results in the further delimitation of the number of studies eventually included in the study and can be considered as a type of selection bias. Some meta-analytic studies are not based on repeated-measures designs (Normand, 1995); however, multiple end-point studies do complicate the methodology involved in synthesizing such data. We suggest that more commercially available software would better serve such studies as macros that run within SAS and SPSS. It is perhaps something to think about in terms of re-conducting this study utilizing more accommodating software so as to include those studies left out of this one. This cannot be said to be a major flaw in the programs because most meta-analyses are utilized for the express purpose of determining effects on experimental groups as opposed to the lack thereof on control groups. It is perhaps telling, then, that local research designs veer more towards within-group analyses and not between-group analyses. The 7 primary studies used for this analysis generated 22 effect sizes. This number was generated due to the input of more than one dependent sample per study. However, 2 primary studies generated three and four independent results, respectively, thus averaging 10 independent effect sizes. An option to average out the effect sizes per study was considered

but rejected as too few effect sizes would have made this endeavor superfluous. This study has thus violated an assumption inherent in both programs, that of independent samples. When interpreting the results it is prudent to keep in mind this violation.

Results

Criteria used for inclusion of studies

Of the original 22 studies, 11 used means and standard deviations, 5 used correlations, 3 used *t* tests, and the remaining 2 used multiple regression as well as discriminant analyses. Of the final 7 chosen for inclusion (i.e., the between-groups studies) 6 used means and standard deviations and 1 used a combination of *t* test results as well as means and standard deviations. For the sake of complete comparison between the two programs, only means and standard deviations were used (including the means and standard deviations for one study which also included *t* test results). Two analyses were conducted as the first analysis yielded effect sizes that were very large. Due to the unusually large effect size results obtained from both programs on an initial data run, the output was studied and three outliers identified. These outliers emanated from two primary studies, namely Lloyd and Pidgeon (1961) (study numbers 20 and 21 samples 1 and 2) and Gewer (1998) (study 19 sample number 4). Effect sizes greater than two to three standard deviations of the mean may be construed as outliers (Chambers, 2004).

Individual effect sizes for these studies ranged from 2.2 to 13.8. An analysis of the primary research yielded the reasons for these effect sizes: the Lloyd and Pidgeon (1961) study presented with very small standard deviations for both the control and experimental groups (0.83, 0.85, 1.08), which was exacerbated by the already large differences between the means of both groups. As the calculation of effect size is very dependent on standard deviation and mean, it stands to reason that this would be the case. Lloyd and Pidgeon (1961) state the following: “it is not thought that the low variance can be attributed to unrepresentativeness but rather to the greater homogeneity of the Natal children when compared to the English children” (p. 150). It is for this reason that these two samples were eliminated from the data during the second analyses (yielding effect sizes of 6.416 and 13.806 respectively). The sample of Gewer (1998) presented with a large difference between both groups (yielding an effect size of 2.211). This particular sample was also eliminated during the second analyses.

Table 1 presents the study numbers of the primary study authors. The results of the initial data run yielded 25 effect sizes. Asterisked studies (19, 20, and 21) are those with outliers and were eliminated during the second analyses, thus lowering the total number of effect sizes down to 22. It is these data (22 effect sizes) with which this study is concerned. Table 2 presents a brief outline of the 7 studies included for the analyses. For studies 12, 13, 16, and 17 the authors divided both experimental and control groups into two groups each, thus yielding four sub-groupings. In order to benefit from an increased sample size for the purposes of the meta-analysis the four sub-groupings were ‘’collapsed’’ into two groups. In order to do so the following formulae were used to calculate means and standard deviations respectively: Average mean: $M= [(M_{Exp} \times n_1) + (M_{Con} \times n_2)]/n_1 + n_2$; Average standard deviation: $SD= [(n_1\text{-}1)SD_1 + (n_2\text{-}1)SD_2)]/(n_1 + n_2 – 2)$.

Table 1. Study Numbers and Sample Classification

Study number	Author and year	Sample number	Sample type
Study 1	Andrews (1996)	sample 1	dependent sample
Study 2	Andrews (1996)	sample 2	dependent sample
Study 3	Andrews (1996)	sample 3	dependent sample
Study 4	Andrews (1996)	sample 4	dependent sample
Study 5	Boeyens (1989)	sample 1	dependent sample
Study 6	Boeyens (1989)	sample 2	dependent sample
Study 7	Murray (1988)	sample 1	dependent sample
Study 8	Murray (1988)	sample 2	dependent sample
Study 9	Murray (1988)	sample 3	dependent sample
Study 10	Murray (1988)	sample 4	dependent sample
Study 11	Murray (1988)	sample 5	dependent sample
Study 12	de Villiers (1999)	sample 1	independent sample
Study 13	de Villiers (1999)	sample 2	independent sample
Study 14	de Villiers (1999)	sample 3	independent sample
Study 15	de Villiers (1999)	sample 4	independent sample
Study 16	Gewer (1988)	sample 1	dependent sample
Study 17	Gewer (1988)	sample 2	dependent sample
Study 18	Gewer (1988)	sample 3	dependent sample
Study 19·	Gewer (1988)	sample 4	dependent sample
Study 20·	Lloyd and Pidgeon (1961)	sample 1	independent sample
Study 21·	Lloyd and Pidgeon (1961)	sample 2	independent sample
Study 22	Lloyd and Pidgeon (1961)	sample 3	independent sample
Study 23	Hoffenberg (1988)	sample 1	dependent sample
Study 24	Hoffenberg (1988)	sample 2	dependent sample
Study 25	Hoffenberg (1988)	sample 3	dependent sample

Table 2. Outline of the 7 Studies Included for the Analysis

Study	Brief description
Andrews (1996)	To determine if cognition is modified in a group administration of the LPAD; and to detect differences in the degree of modifiability
Boeyens (1989)	To evaluate the performance of a learning potential instrument and to investigate the relationship between academic performance and learning potential
De Villiers (1999)	To investigate the practical application of Vygotsky's construct of the zone of proximal development to the selection of disadvantaged students in higher education and to determine alternative predictors of academic performance other than the traditional matriculation examination results used in South Africa
Gewer (1988)	The study investigated the application of dynamic assessment to a sample of black children within a South African township clinic setting
Hoffenberg (1988)	The study aimed to assess the effectiveness of dynamic assessment among a group of academically superior individuals from a disadvantaged black community in South Africa
Lloyd and Pidgeon (1961)	To compare the performance of children from different cultural groups on non-verbal tests, half the children were coached and the other half were not
Murray (1998)	To test the effectiveness of a dynamic assessment approach (LPAD) among groups of socio-politically and educationally disadvantaged Indian and Colored adolescents

Results using META by Kenny

Table 3 presents the results for the meta-analysis using the program by Kenny.

Table 3. Meta-Analysis Results Using Kenny's Program

Study Number	22
Subject *n*	2032
Average effect size	0.3354•
Effect size standard deviation	0.4787
t test of effect size	3.2863* *df* 21
Average *d*	0.3408
Average *r*	0.1540
BESD	0.4230 – 0.5770
Homogeneity of effect sizes Chi Square	103.3336** *df* 21
Average *z*	5.5697***
Fail-safe *N*	156

Note. Figures are accurate to three decimal places.
*$p < .005$, **$p < .001$.

Results using meta-analysis by Schwarzer

Table 4 shows the results for the meta-analysis using the program by Schwarzer.

Table 4. Meta-Analysis Results Using Schwarzer's Program

Statistic	Result	Statistic	Result
Number of effect sizes	22	Total sample size	2032
Unweighted mean of effect sizes *g*	0.4136	SE	0.0967
Observed variance of effect sizes *g*	0.2060	SD	0.4538
Unweighted mean of adjusted effect sizes *d*	0.4055	SE	0.0949
Observed variance of adj. effect sizes *d*	0.1983	SD	0.4454
"Weighted Integration Method"			
Mean effect size *d+*	0.2370	SE	0.0451
Significance *z*	5.1901*		
Variance	0.0020	SD	0.0451
95% Confidence interval	0.1475 to 0.3265	Homogeneity Q	1032.2411** df = 21
"Random Effects Model"			
Mean effect size DELTA	0.3481	SE	0.0910
95% Confidence interval	0.1697 to 0.5266	Significance *z*	3.8237***
Observed variance	0.1983	Error variance	0.0788
Population variance	0.1195	Homogeneity Q	102.6985**** df = 21

table continues

Amount of variance explained by sampling error	39.73 %		
Kraemer (1983) method			
Statistic	Result	Statistic	Result
Mean effect size *d*	0.2587	95% Confidence interval	0.1698 to 0.3481
Population effect size Rho	0.1282	Variance of rho	0.0005
95% Confidence interval	0.0846 to 0.1715	Homogeneity Chi-square	109.9362
Orwin's Fail-safe *n* based on "random effects model" DELTA			
Fail-safe for critical *d* of .20	16.2940	Fail-safe for critical *d* of .50	-6.6824
Fail-safe for critical *d* of .80	-12.4265		

* $p < .001$.

Discussion

Kenny's program

The scrutiny of the normality of data should not be problematic. Meta-analysis cumulates findings resulting in large enough sample sizes to rest on the assumption of normality through reference to the central limit theorem (Normand, 1995). Twenty-two effect sizes with a sample of 2032 yielded an average effect size of 0.3354 which is significant when the *t* test result (two-tailed) is studied, $t = 3.2863$ $p < 0.005$ $df = 21$. The effect size differed significantly from zero, resulting in the conclusion that dynamic assessment did in fact have an effect on posttest scores when studies are cumulated. The *t* test treats study as unit of analysis, but *z* treats person as unit which also happens to be significant in this instance; average $z = 5.5697$ $p < 0.001$. This answers affirmatively the question presented earlier of whether dynamic assessment intervention makes a significant difference as opposed to no (static) intervention across separate studies on posttest scores. This result, as highlighted earlier, cannot comment on the entire value of dynamic assessment as a mediational tool. Potential long-term cognitive effects resulting from mediation programs are not measured in the individual studies and similarly cannot be assessed in the meta-analysis. Caution is attached to the interpretation of this significant value. However, the meta-analytic finding does indicate that the significance of the individual studies is evidenced in the synthesized effect sizes.

The BESD (binomial effect size) measures the estimated difference between the experimental and control groups in terms of proportions. The test of homogeneity relies on the chosen statistic used to compute the effect size, which in this case was d, and thus the program employed Hedge's test of homogeneity. The chi-square statistic is used to compute the test for homogeneity, which in this instance is highly significant, thus indicating that the studies are not homogenous. Effect sizes differ due to factors other than sampling error. The fail-safe number generated by Kenny's program yields 156 null studies which would have to be

generated for this test to be not significant; that is, 156 similar studies will need to be uncovered for this result to be nullified (Strube, 1985; Strube & Hartmann, 1983). Studies were not transformed by any means other than by making use of Hedge's transformation, which, as Kenny correctly points out is not truly a new weighting but merely a sample size correction factor (Kenny, 2003). Hedge's transformation is a correction that improves the sample estimate of the standardized mean difference between the two groups" (Kalaian & Raudenbush, 1996, p.229). No untransformed estimate of average effect size is produced. Determination of effect size used by Cohen ranges from 0.2 (small effect), 0.5 (medium effect) and 0.8 (large effect) (Schwarzer, 1989). Thus the effect size of 0.3354 can be considered as halfway between a small and medium effect size.

Schwarzer's program

Twenty-two effect sizes with a sample of 2032 yielded an average effect size of 0.2370 which, although significant, was calculated using the "weighted integration method.'' The subsequent chi-square statistics yielded a highly significant result, thus prompting the user to investigate the "random effects model'' (Schwarzer's program offers three types of output). The mean effect size delta is 0.3481 and is significant in the random effects model. An aspect, which is worrying, is that 39.73% of the variance explained is due to sampling error. This underlies the original hesitation at running a meta-analysis with data from master's and doctoral studies in which original samples chosen for each study were not randomly chosen. Nevertheless the effect size answers affirmatively the question presented earlier of whether dynamic assessment intervention makes a significant difference as opposed to no (static) intervention across separate studies on posttest scores. Once again, this result is pertinent to the effects of dynamic assessment intervention on the posttest scores and cannot comment on the long-term effects of mediation programs.

The fail-safe number generated by Schwarzer's program is interpreted differently from that of Kenny's. Schwarzer's fail-safe number is the amount of studies needed for critical effect sizes of 0.2, 0.5 and 0.5 respectively. As the fail-safe for both the 0.5 and 0.8 delta levels exceed that of 0.3481 these two fail-safe numbers are meaningless. Although Kenny does not specifically mention how the fail-safe number is computed, it is assumed that Rosenthal's formula has been used. Schwarzer, however, makes use of Orwin's fail-safe number computation, which is an adapted version of the original Rosenthal formula. Kenny's effect size of 0.3354 and Schwarzer's effect size of 0.3481 (delta, random effects model) differs by 0.012 and is slight. Also, values from the various confidence intervals (the 95% confidence intervals described in the weighted integration, random effects and the Kraemer, 1983 models used by Schwarzer, 1989) do not contain zeros, further supporting a significant effect. A zero in the interval could possibly indicate that there is no effect (Chambers, 2004). Any differences between packages usually results in fourth and higher decimal place differences which is acceptable (Arthur et al., 1994). The similarity of output further minimizes any judgment calls used when making the decision to run the analyses on two software programs. Figure 1 illustrates the range of effect sizes in study number order ranging from study 1 to study 22.

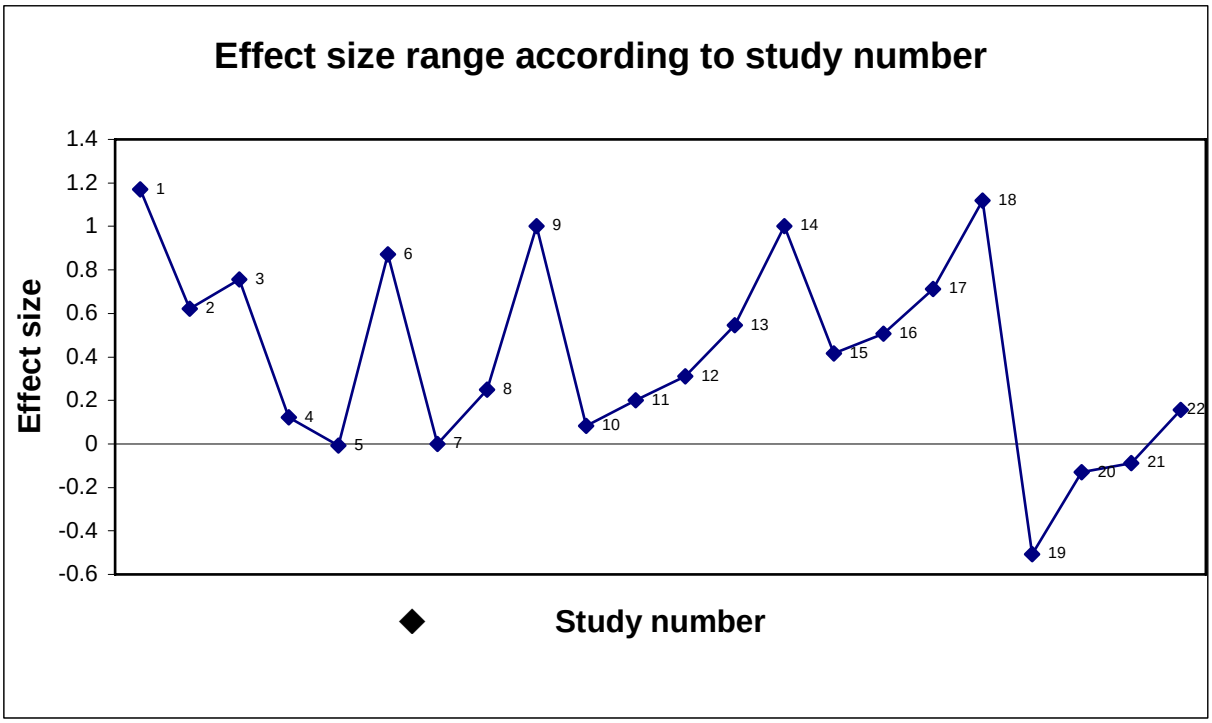

Figure 1. Range of effect size across study number.

A search for moderators

An effect size cluster analysis was conducted in order to determine the heterogeneity of the data set due to the heterogeneity of effect sizes. The chi-square distribution indicated that homogeneity was rejected. The resultant classification of two clusters is evident yet the second cluster comprises only one study and is thus considered an outlier in terms of heterogeneity of effect size when compared to the other 21 studies. A robust search for potential moderators was not carried out due to the small sample size of effect sizes, however this is strongly advised for larger data sets (Chambers, 2004; Eagly & Wood, 1994; Kenny, 2003; Rosenthal, 1995; Schwarzer, 1989). Cluster analysis decomposes the number of effect sizes into smaller sets and effect sizes are rank ordered according to their similarity. This procedure is suited to smaller effect size samples and works less well for unequal sample sizes (Schwarzer, 1989). A cursory glance of the output of effect size cluster analysis (see Table 5) computed on the data file using Schwarzer's program, evidenced two clusters. Only one study is located in the second cluster (this "cluster" is considered an outlier but this terminology is used by Schwarzer); at the 1%, 5% and 10% levels of significance (the study by Lloyd and Pidgeon, 1961; study number 19 in the second data run). According to Schwarzer (1989), "the formulae for the critical values and the computer algorithm for the disjoint cluster analysis have been taken from Mullen and Rosenthal, (1985)'' (p. 33). Additional information is available at the end of Table 5 and yields information on average sample size and standard deviation. A larger database would have allowed more probing investigative analysis in terms of moderator effects (such as the potential moderating effects of gender, age, level of education, and cultural grouping). The quality of the primary studies regarding greater elucidation of sample characteristics and also the small number of studies eventually included did not warrant such an investigation at this stage.

Table 5. Effect Size Cluster Analysis

CLUSTERS AT 1 % 5% and 10% LEVELS OF SIGNIFICANCE		
CLUSTER 1		
StudyID 1	Effect Size=	1.1689
StudyID 18	Effect Size=	1.1180
StudyID 14	Effect Size=	1.0013
StudyID 9	Effect Size=	1.0000
StudyID 6	Effect Size=	0.8704
StudyID 3	Effect Size=	0.7554
StudyID 17	Effect Size=	0.7124
StudyID 2	Effect Size=	0.6205
StudyID 13	Effect Size=	0.5454
StudyID 16	Effect Size=	0.5060
StudyID 15	Effect Size=	0.4148
StudyID 12	Effect Size=	0.3096
StudyID 8	Effect Size=	0.2500
StudyID 11	Effect Size=	0.2000
StudyID 22	Effect Size=	0.1559
StudyID 4	Effect Size=	0.1232
StudyID 10	Effect Size=	0.0830
StudyID 7	Effect Size=	0.0000
StudyID 5	Effect Size=	-0.0082
StudyID 21	Effect Size=	-0.0892
StudyID 20	Effect Size=	-0.1301
CLUSTER 2		
StudyID 19	Effect Size=	-0.5061
Additional Information		
Average Sample Size = 92.3636		
Sample Size Std. Dev= 58.3006		
Correlation between Sample and Effect Sizes = -0.5098		

Limitations of the meta-analysis

First, only 7 studies were included (out of a potential number of 22) for this meta-analysis primarily due to the unavailability of the two software programs to proceed with the accumulation of within-groups studies and repeated-measures designs. Second, of these 7 studies, 22 effect sizes were generated, and of these 22 effect sizes, 12 were dependent samples and 10 were independent samples thus violating the inherent assumption of independence. The results should thus be interpreted with due caution. Although there is

a significant small to medium effect size evident in the South African research literature pertaining to dynamic assessment as an effective intervention strategy in terms of effecting posttest score results, this effect is nevertheless resultant on the quality of studies included for assessment (mostly master's and doctoral studies) the small number of studies finally included (7 from 22 studies) and the fact that 15 of the original studies assessed within-groups results and not between-groups results. This is of itself an important finding which highlights the differences between the South African and overseas research literature in this area. If the original pool of studies had been larger and more varied in terms of quality and had as their research designs between-groups as opposed to within-groups designs, the results may have been even more significant in terms of veracity, applicability and generalizability.

Implications of the findings for dynamic assessment research in South Africa

The effect size of 0.3354 obtained using Kenny's program yielded a significant result with a similar effect size of 0.3481 obtained using Schwarzer's program. In answer to the question of whether dynamic assessment interventions across cumulated South African research indeed had any cumulative significant effect, it can be stated that there is a small to medium effect size across studies. This supports the use of dynamic assessment in South Africa but only as it pertains to the improvement of posttest score results. The study cannot comment on the value of dynamic assessment as a mediational tool in terms of other influences it may evidence, such as the future value of assessor-testee interaction. The significant results from this meta-analysis serve to add credence to an approach that seeks to assess individuals in as unbiased a manner as possible and that has as a core philosophy the understanding of individual change through the learning process. These findings do, however, have to be tempered with the fact that these are results are based on only seven studies.

Recommendations

Meta-analysis is dependent on the accuracy and robustness of primary research data and can never replace the need for primary research. Most of the studies included in this meta-analysis were master's and doctoral studies and the data were not, in most instances, in the correct format for a meta-analysis to be conducted. It is recommended that, in the future, primary empirical dynamic assessment studies be conducted with subsequent meta-analyses in mind. Much worthwhile data was not included in this analysis as a lot of the necessary data information was not included in the primary texts. At times, studies made use of only one group. As more primary research data are added to the field of dynamic assessment in South Africa, it will become increasingly important for cumulative studies to be conducted in order to determine the cumulative efficacy of all the research.

Regarding the utility of the two meta-analytic software programs, both Kenny and Schwarzer's programs yield similar effect sizes and both programs differ in some ways and are alike in others. The programs are freely available over the Internet and are not fully operational (such as the lack of saving and printing facilities). These programs run at no cost and so this cannot be considered a criticism.

Dynamic assessment is more than a tool utilized for increasing scores within pretest-postest research studies. Assessing the synthesized effect sizes of qualitative mediational interventions through meta-analysis may also prove fruitful. However, most quantitative primary studies in South Africa do not study this aspect of dynamic assessment and have as their focus pretest-posttest research designs. It is recommended that primary studies could perhaps quantify such mediational qualitative studies, thus allowing later meta-analysts an opportunity to cumulate effect sizes across studies.

Conclusion

The aim of this study was two-fold: to determine the significance of the efficacy of dynamic assessment as a useful assessment strategy in South African studies in terms of increasing scores on posttest test results and to compare and analyze two meta-analytic software programs. Meta-analysis is a powerful technique, which can aid in the determination of how effectual individual studies are when cumulated. Synthesized study results may evidence results contrary to individual study results. Twenty-two studies were originally coded for inclusion into the meta-analysis but due to the incorrect format of the original data only seven of these studies were included in the final analysis. Effect sizes of 0.3354 and 0.3481 respectively were calculated utilizing two meta-analysis software programs. Due to this small sample size, however, only limited conclusions can be drawn. It was evidenced that the findings were indeed significant. Limitations surrounding the use of the two packages and violations inherent in the analysis of the data were addressed. We suggest that any future South African research designs conducted in the field of dynamic assessment be designed in such a way as to accommodate future meta-analysis, as this technique is dependent on data that can be successfully incorporated into such analyses. The recommendation concerning the utility of the two software programs was based on their ease of use, technical features, similarity of output, and cost-effectiveness.

References

Andrews, S. G. (1996). *A small-scale investigation of the group administration of Feuerstein's learning potential assessment device*. Unpublished master's thesis. University of Natal, Pietermaritzburg, South Africa.

Arthur, W., Bennett, W., & Huffcutt, A. (1994). Choice of software and programs in meta-analysis research: does it make a difference? *Educational and psychological measurement, 54*(3), 776-787.

Bangert-Drowns, R. L. (1992). Review of developments in meta-analytic method. In A.E. Kazdin (Ed), *Methodological issues and strategies in clinical research* (pp. 439-467). Washington: American Psychological Association.

Becker, B. J. (1994). Combining significance levels. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 215-230). New York: Russell Sage Foundation.

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399-409). New York: Russell Sage Foundation.

Boeyens, J. C. A. (1989). *Learning potential: An empirical investigation.* Pretoria, South Africa: Human Sciences Research Council.

Budoff, M. (1987). The validity of learning potential assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 52-81). New York: Guilford Press.

Bushman, B. J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 3-14). New York: Russell Sage Foundation.

Campione, J. C. (1996). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. In H. Daniels (Ed.), *An introduction to Vygotsky* (pp. 219-250). London: Routledge.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82-109). New York: Guilford Press.

Chambers, E. A. (2004). An introduction to meta-analysis with articles from the journal of educational research (1992-2002). *The Journal of Educational Research, 98*(1), 35-45.

Cooper, H., & Hedges, L. V. (1994a). Research synthesis as a scientific enterprise. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.3-14). New York: Russell Sage Foundation.

Cooper, H., & Hedges, L. V. (1994b). Potentials and limitations of research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 521-529). New York: Russell Sage Foundation.

Cooper, H. M., & Lemke, K. M. (1991). On the role of meta-analysis in personality and social psychology. *Personality and Social Psychology Bulletin, 17*(3), 245-251.

Das, J. P. (1987). Forward. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp.vii-xi). New York: The Guilford Press.

Department of Education (2003). *Education statistics in South Africa at a glance in 2001.* Pretoria, South Africa: EMIS Directorate, Department of Education.

De Villiers, A. B. (1999). *Disadvantaged students' academic performance: Analysing the zone of proximal development.* Unpublished doctoral dissertation, University of Cape Town, South Africa.

Dickersin, K. (1994). Research registers. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.71-83). New York: Russell Sage Foundation.

Eagly, A. H, & Wood, W. (1994). Using research synthesis to plan future research. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.485-500). New York: Russell Sage Foundation.

Elkonin, D., Foxcroft, C., Roodt, G., & Astbury, G. (2001). The use of assessment measures in various applied contexts. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context* (pp.265-294). Cape Town, South Africa: Oxford University Press.

Elliott, J. G. (2000). Dynamic assessment in educational contexts: purpose and promise. In C. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp.713-740). New York: Elsevier.

Eysenck, H. J. (1995). Meta-analysis squared – does it make sense? *American psychologist, 50*, 110-111.

Feuerstein, R., Feuerstein, R. S., Falik, L. H., & Rand, Y. (Eds.). (2002). *The dynamic assessment of cognitive modifiability*. Jerusalem, Israel: ICELP Press.

Gewer, A. (1998). *Uncovering potential: Dynamic assessment of non-verbal reasoning ability in educationally disadvantaged children.* Unpublished master's dissertation, University of the Witwatersrand. South Africa.

Glass, G., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*(1), 75-111.

Guthke, J. (1993). Current trends in theories and assessment of intelligence. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues* (pp.13-18). Amsterdam: Swets & Zeitlinger.

Hall, J. A., Tickle-Degnen, L., Rosenthal, R., & Mosteller, F. (1994). Hypotheses and problems in research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.17-28). New York: Russell Sage Foundation.

Hamers, J. H. M. & Resing, W. C. M. (1993). Learning potential assessment: Introduction. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues (*pp.23-41). Amsterdam: Swets & Zeitlinger.

Hessels, M. G. P. & Hamers, J. H .M. (1993). The learning potential test for ethnic minorities. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues* (pp.285-311). Amsterdam: Swets & Zeitlinger.

Hoffenberg, S. R. (1988). *Effectiveness of the Learning Potential Assessment Device with high achieving adolescents from an advantaged community.* Unpublished master's dissertation, University of the Witwatersrand, South Africa.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* London: Sage.

Jensen, M. R. (2000). The mindladder model: Using dynamic assessment to help students learn to assemble and use knowledge. In C. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp.187-227). New York: Elsevier.

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods, 1*(3), 227-235.

Kenny, D. A. (2003). META-analysis: Easy to answer. Version III. Retrieved March 10, 2004, from http://davidakenny.net/meta.htm

Kulik, J. A. & Kulik, C. C. (1989). Meta-analysis in education. *International Journal of Educational Eesearch, 13*, 221-340.

Lidz, C. S. (Ed.). (1987). *Dynamic assessment: An interactional approach to evaluating learning potential.* New York: The Guilford Press.

Lidz, C. S. (1997) Dynamic assessment approaches. In D. P. Flanagan, J. L. Genshaft, & P. L Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp.281-296). New York: Guilford Press.

Lidz, C. S. & Elliott, J. G. (2000). Introduction. In C. S. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp.3-13). New York: Elsevier.

Lidz, C. S., & Pena, E. D. (1996). Dynamic assessment: The model, its relevance as a nonbiased approach, and its application to Latino American preschool children. *Language, Speech and Hearing Services in Schools, 27*(4), 367-372.

Lloyd, F., & Pidgeon, D. A. (1961). An investigation into the effects of coaching on non-verbal test material with European, Indian and African children. *The British Journal of Educational Psychology, 31*(2), 145-151.

Lussier, C. M., & Swanson, L. (2002). Dynamic assessment: A selective synthesis of the experimental literature. In G. M. van der Aalsvoort, W. C. M. Resing & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment and cognitive training: Actual research and perspectives in theory building and methodology* (pp.65-87). Amsterdam: JAI.

The Meta Analysis Unit, Universidad de Murcia. (2005). Retrieved April 13 2005 from www.um.es/facpsi/metaanalysis/software.php

Minick, N. (1987). Implications of Vygotsky's theories for dynamic assessment. In Lidz, C. S. (Ed.), (1987). *Dynamic assessment: An interactional approach to evaluating learning potential* (pp.116-140). New York: The Guilford Press.

Murphy, R. (2002). *A review of South African research in the field of dynamic assessment. Unpublished master's thesis*. University of Pretoria, South Africa. (available online from: http://upetd.up.ac.za/thesis/available/etd-05042002-161239/)

Murray, D. A. (1988). *Effectiveness of Feuerstein's learning potential assessment device in a South African context.* Unpublished master's thesis. University of the Witwatersrand, South Africa..

Myers, D. G. (1991). Union is strength: a consumer's view of meta-analysis. *Personality and social psychology bulletin, 17*(3), 265-266.

Normand, S. T. (1995). Meta-analysis software: a comparative review. *The American statistician, 49*(3), 298-309.

Orwin, R. G. (1994). Evaluating coding decisions. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.139-162). New York: Russell Sage Foundation

Reed, J. G. & Baxter, P. M. (1994). Using reference databases. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.57-70). New York: Russell Sage Foundation.

Rosenthal, M. C. (1994). The fugitive literature. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.85-94). New York: Russell Sage Foundation.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological bulletin, 86*(3), 638-641.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological bulletin, 118*(2), 183-192.

Schwarzer, R. (1989). Meta-analysis programs. Retrieved July 25, 2004, from http://www.RalfSchwarzer.de.

Sewell, T. E. (1987). Dynamic assessment as a non-discriminatory procedure. In C. S. Lidz (Ed.), (1987). *Dynamic assessment: An interactional approach to evaluating learning potentia* (pp.426-443). New York: The Guilford Press.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.261-281). New York: Russell Sage Foundation.

Skuy, M., Gewer, A., Osrin, Y., Khunou, D., Fridjon, P., & Rushton, J. P. (2002). Effects of mediated learning experience on Raven's Matrices scores of African and non-African university students in South Africa. *Intelligence, 30*, 221-232.

Sternberg, R. J. (2000). Prologue. In C. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: prevailing models and applications* (pp.xiii-xv). New York: Elsevier.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge: Cambridge University Press.

Stock, W. A. (1994). Systematic coding for research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.125-138). New York: Russell Sage Foundation.

Strube, M. J. (1985). Combining and comparing significance levels from nonindependent hypothesis tests. *Psychological bulletin, 97*(2), 334-341.

Strube, M.J. & Hartmann, D.P. (1983). Meta-analytic techniques, applications, and functions. *Journal of consulting and clinical psychology, 51*(1), 14-27.

van de Vijver, F.J.R. (1993). Learning potential assessment from a cross-cultural perspective. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issu*es (pp.313-340). Amsterdam: Swets & Zeitlinger.

White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.41-55). New York: Russell Sage Foundation.

Wolf, F. M. (1987). *Meta-analysis: Quantitative methods for research synthesis.* Beverly Hills, CA: Sage.

Wortman, P. M. (1994). Judging research quality. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis (*pp.97-109). New York: Russell Sage Foundation.

Appendix

For purposes of replication two tables are included which detail the data file outputs for Kenny and Schwarzer's programs respectively. Data is presented with an accuracy of up to four decimals.

Table A1. Data Output File Yielded by Kenny's Program

Study Number	Degrees of freedom	Total N per study	Effect size	D	r	Z	t	p2 (confidence interval)	p1 (confidence interval)	n1 (experimental group 1)	n2 (control group 2)	weight
1	19	21	1.1221	1.1689	0.5230	2.4319	2.6752	0.7615	0.2384	10	11	4.5825
2	19	21	0.6204	0.6463	0.3213	1.4201	1.4792	0.6606	0.3393	10	11	4.5825
3	19	21	0.7252	0.7554	0.3687	1.6441	1.7289	0.6843	0.3156	10	11	4.5825
4	19	21	0.1232	0.1283	6.7253	0.2895	0.2938	0.5336	0.4663	10	11	4.5825
5	181	183	-8.1786	-8.2127	-4.1288	-5.5472	-5.5549	0.4979	0.5020	91	92	13.5277
6	181	183	0.8667	0.8704	0.4008	5.6237	5.8872	0.7004	0.2995	91	92	13.5277
7	106	108	0	0	0	0	0	0.5	0.5	54	54	10.3923
8	106	108	0.2482	0.25	0.1251	1.2908	1.2990	0.5625	0.4374	54	54	10.3923
9	106	108	0.9929	1	0.4505	4.8927	5.1961	0.7252	0.2747	54	54	10.3923
10	106	108	8.2456	8.3045	4.1875	0.4303	0.4315	0.5209	0.4790	54	54	10.3923
11	106	108	0.1985	0.2	0.1004	1.0341	1.0392	0.5502	0.4497	54	54	10.3923
12	60	62	0.3057	0.3096	0.1541	1.1965	1.2088	0.5770	0.4229	35	27	7.8740
13	61	63	0.5386	0.5453	0.2670	2.1154	2.1640	0.6335	0.3664	32	31	7.9372
14	100	102	0.9937	1.0012	0.3892	4.0434	4.2260	0.6946	0.3053	79	23	10.0995
15	67	69	0.4101	0.4148	0.2059	1.6977	1.7227	0.6029	0.3970	34	35	8.3066
16	70	72	0.5013	0.5068	0.2354	1.9912	2.0272	0.6177	0.3822	48	24	8.4852
17	70	72	0.7047	0.7123	0.3223	2.7617	2.8495	0.6611	0.3388	48	24	8.4852
18	34	36	1.0931	1.1180	0.4767	2.9388	3.1622	0.7383	0.2616	24	12	6
19	264	266	-0.5047	-0.5061	-0.2462	-4.0592	-4.1274	0.3768	0.6231	133	133	16.3095
20	98	100	-0.1290	-0.1300	-6.5503	-0.6474	-0.6498	0.4672	0.5327	52	48	10
21	98	100	-8.8529	-8.9214	-4.4978	-0.4443	-0.4457	0.4775	0.5224	52	48	10
22	98	100	0.1547	0.1559	7.8443	0.7757	0.7789	0.5392	0.4607	52	48	10

Table A2. Data Output File Yielded by Schwarzer's Program

Study number	Sample size group 1 (experimental group)	Sample size group 2 (control group)	Effect size	Reliability coefficient
1	10	11	1.1689	1
2	10	11	0.6205	1
3	10	11	0.7554	1
4	10	11	0.1232	1
5	91	92	-0.0082	1
6	91	92	0.8704	1
7	54	54	0.0000	1
8	54	54	0.25	1
9	54	54	1.0000	1
10	54	54	0.0830	1
11	54	54	0.2000	1
12	35	27	0.3096	1
13	32	31	0.5454	1
14	79	23	1.0013	1
15	34	35	0.4148	1
16	48	24	0.5060	1
17	48	24	0.7124	1
18	24	12	1.1180	1
19	133	133	-0.5061	1
20	52	48	-0.1301	1
21	52	48	-0.0892	1
22	52	48	0.1559	1

Acknowledgments

The authors are grateful to J. H. Owen and J. Pauw at the Department of Statistics, University of Pretoria, for their assistance with this meta-analysis. This study is based on partial work submitted for Raegan Murphy's doctoral dissertation.

Address for correspondence:

Raegan Murphy
P.O. Box 27846
Sunnyside, 0132
South Africa
raegan@mweb.co.za

Résumé

Une Méta-analyse des Recherches Portant sur l'Évaluation Dynamique en Afrique du Sud

Le besoin d'un état des lieux de l'enseignement supérieur en Afrique du Sud a accentué le rôle que l'évaluation dynamique a à jouer à cet égard et, comme aucune conclusion statistique cumulative n'a été tirée des différentes études de cas, nous avons jugé qu'il était temps de le faire. C'est donc pour combler un manqué dans ce champ qu'une méta-analyse a été menée sur des études qui étaient centrées sur l'efficacité des interventions d'évaluation dynamique dans des contextes variés. Peu d'études de recherches traitant de l'évaluation dynamique ont été entreprises en Afrique du Sud comparé au nombre de recherches réalisées ailleurs sur cette question. L'étude visait un double objectif :

1. évaluer la significativité de la taille de l'effet d'un certain nombre d'études de cas qui avait comme but premier d'évaluer la significativité des interventions d'évaluation dynamique ;
2. comparer deux programmes informatiques de méta-analyses qui sont accessibles gratuitement en ligne.

Une petite partie des effets de taille moyens (compris entre 0.3354 et 0.3481) ont été produits par les deux programmes avec un taille d'effet typique qui s'étend entre 0.2 et 0.8. La méthode et les résultats de cette méta-analyse sont discutés dans cet article ainsi que les limites inhérentes aux deux programmes et à l'étude. Il se conclut par des recommandations pour de futures méta-analyses en Afrique du Sud dans le champ de l'évaluation dynamique.

Resumen

Meta-Análisis de la Investigación sobre Evaluación Dinámica en Sudáfrica

La necesidad de chequear la educación terciaria en Sudáfrica ha destacado el rol que la evaluación dinámica tiene que jugar en ese cometido y debido a que no pueden ser obtenidas conclusiones estadísticas acumuladas procedentes de simples estudios de caso, dicha modalidad de evaluación ha sido considerada la más pertinente. Con el propósito de llenar ese hueco en el campo del meta-análisis, se han llevado a cabo estudios focalizados en la eficacia de las intervenciones de la evaluación dinámica en varios marcos. En Sudáfrica pocos estudios se han ocupado de la evaluación dinámica por comparación a la investigación realizada en otros países en este ámbito.El estudio que se presenta aquí tuvo dos propósitos fundamentales: en primer lugar, evaluar la significación de la amplitud del efecto de un número de estudios individuales, que tenían como intención original investigar la significación de las intervenciones basadas en la evaluación dinámica; en segundo lugar, comparar dos programas informáticos meta-analíticos, que están libremente disponibles on line. Los resultados de ambos programas mostraron un pequeño efecto, equivalente a 0.3354 y 0.3481, en relación con la típica amplitud del efecto, comprendida entre el rango 0.2-0.8. La metodología y los resultados de este meta-análisis se discuten, junto con las limitaciones inherentes a ambos programas y al propio estudio. Finalmente, se presentan conclusiones y recomendaciones para mejorar los estudios meta-analíticos en Sudáfrica en el ámbito de la evaluación dinámica.

Zusammenfassung

Metaanalyse zur Forschung in Dynamischer Diagnostik in Südafrika

Der Bedarf an Screenings im tertiären Schulbereich in Südafrika hat die Bedeutung der dynamischen Diagnostik für diesen Bereich hervorgehoben. Da aus Einzelfallstudien keine kumulativen statistischen Schlussfolgerungen gezogen werden können, wurde versucht, diese Lücke durch eine Metaanalyse über Studien, die die Effizienz von dynamisch-diagnostischen Interventionen innerhalb unterschiedlicher Settings zum Gegenstand haben, zu schließen. In Südafrika sind weniger Studien zur dynamischen Diagnostik durchgeführt worden als in anderen Ländern. Die vorliegende Studie diente dem zweifachen Ziel, die Bedeutung der synthetisierten Effektstärken aus einer Anzahl individueller Studien zu erfassen, deren ursprüngliche Intention die Untersuchung der Bedeutung dynamisch-diagnostischer Interventionen war. Zum zweiten ging es darum, zwei metaanalytische Softwareprogramme zu vergleichen, die über online frei verfügbar sind. Kleine bis durchschnittliche Effektgrößen von 0.3354 und 0.3481 wurden von beiden Programmen gleichermaßen generiert, wobei die typische Effektstärke zwischen 0.2 – 0.8 lag. Die Methode und Ergebnisse dieser Metaanalyse werden zusammen mit den Beschränkungen diskutiert, die in beiden Programmen enthalten sind. Die Studie schließt mit Empfehlungen für weitere metaanalytische Studien in Südafrika innerhalb des Bereichs der dynamischen Diagnostik.