

## Longitudinal Predictive Validity of a Learning Potential Test

Marié de Beer

University of South Africa

Address correspondence to Marié de Beer, Department of Industrial and Organizational Psychology, Unisa, P.O. Box 392, Unisa, Pretoria 3000, South Africa. E-mail: dbeerm@unisa.ac.za

**The aim of this article is to provide empirical psychometric evidence of the (longitudinal) predictive validity of a learning potential measure—the Learning Potential Computerised Adaptive Test (LPCAT)—in comparison with standard static tests with school aggregate results as the criterion measure. Participants were 79 boys (mean age 12.44, SD = 0.44) and 72 girls (mean age 11.18, SD = 0.42) attending two private schools. Correlation and regression analyses were used to evaluate the predictive validity of the learning potential and standard test scores for school aggregate academic results as criterion measure. Results indicate that learning potential scores were statistically significant predictors of aggregate academic results and provided results that were comparable to those of the standard test results—providing empirical support for the use of learning potential tests in mainstream educational settings.**

*Keywords: dynamic assessment (DA); longitudinal predictive validity; item response theory (IRT); computerised adaptive testing (CAT); Learning Potential Computerised Adaptive Test (LPCAT)*

The measurement of learning potential has received considerable attention over the years (Budoff, 1987; Caffrey, Fuchs, & Fuchs, 2008; Embretson, 2004; Feuerstein, 1979; Grigorenko, 2009; Grigorenko & Sternberg, 1998; Haywood, 2008; Lidz, 1987a, 1987b, 1991, 2009; Murphy & Maree, 2006). This approach to cognitive assessment—which typically involves a learning opportunity during assessment in a test-train-retest process - is also referred to as dynamic assessment (DA). Based on Vygotsky's concept of the zone of proximal development (ZPD) (Vygotsky, 1978), it allows for measurement of current as well as projected future levels of performance (Lidz, 1991). It is generally considered fairer for the assessment of disadvantaged individuals because it allows for learning during assessment. However, despite sustained interest in DA over the years, it has not yet been incorporated into mainstream assessment in education or in industry alongside standard assessment methods (Grigorenko & Sternberg, 1998; Lidz, 2009; Sternberg & Grigorenko, 2002). Reasons given are that the test-train-retest strategy often results in longer testing times, measurement problems with comparison of resulting pre-test and post-test scores and limited available psychometric information on DA measures (Beckmann, 2006; Lidz, 2009; Sternberg & Grigorenko, 2002).

Continuing research to investigate the psychometric properties of DA tools is important if they are to be considered and used more widely in educational and industry settings (Caffrey et al., 2008; Lidz, 2009). Provision of information that allows for comparison of the predictive validity of DA versus standard measures—such as the results presented in this article—could be useful in this regard.

### The Learning Potential Computerised Adaptive Test (LPCAT)

The Learning Potential Computerised Adaptive Test (LPCAT: De Beer, 2000a, 2000b, 2005) uses IRT and CAT to address issues about DA noted by other authors (Embretson, 1996; Embretson & Reise, 2000; Sijtsma, 1993a, 1993b).

Three different types of nonverbal figural item formats are used in the LPCAT namely figure series, figure analogies and pattern completion. Since they measure fluid general reasoning ability, they do not rely on language proficiency or scholastic background and are deemed to be more equitable than items with verbal or scholastic content (Claassen, 1997; Foxcroft, 2004; Hugo & Claassen, 1991; Owen, 1998). In the typical test-train-retest approach of most DA tests, the LPCAT uses two separate but linked adaptive tests with a focus on measurement of fluid ability by means of nonverbal figural content (De Beer, 2005). A unidimensional structure was indicated by factor analysis of items in the item banks and coefficient alpha values ranging between 0.926 and 0.978 for different groups have been reported (De Beer, 2005).

### The Legal-professional Context

The South African Employment Equity Act (EEA, 1998) requires psychological tests used to show evidence of the acceptability of their psychometric properties and fairness towards different groups (Foxcroft, 2004; Paterson & Uys, 2005). Research results on the LPCAT have shown evidence on its reliability and validity for different groups (De Beer, 2000b, 2005, 2006; Van der Merwe & De Beer, 2006; Van Eeden, De Beer & Coetzee, 2001). The fact that differential item functioning (DIF) analysis was used during test construction to eliminate biased items (De Beer, 2004), that the LPCAT incorporates DA and CAT in a power test with non-verbal figural content, uses only the space bar and Enter keys during administration and that the test instructions are available in all 11 South African indigenous languages goes some way with regard to the fairness requirement in the South African context.

The use of IRT and CAT provide several advantages to DA, namely accurate measurement of difference scores, shortened test times, equivalent measurement accuracy at all levels of ability and use of items that match the estimated ability level of individuals being tested throughout both the pre- and post-tests (Kim-Kang & Weiss, 2008; Van der Linden, 2008a, 2008b;

Weiss, 1983). In addition to the pre-test, post-test and difference scores, the LPCAT also provides a composite score which represents a global potential score. The composite score represents a reasoned combination of the pre- and post-test scores which allows for the fact that it is more difficult to show improvement when the initial pre-test performance level is high compared to a lower initial level of performance (De Beer, 2005). Research results for the LPCAT have generally shown that the post-test score (which allows for maximum credit for the learning that has been achieved) and the composite score (which allows for partial credit for the learning that has been achieved) improve predictive validity compared to the pre-test score (which indicates current level of performance) (De Beer, 2005; 2008).

The South African National Qualifications Framework (NQF) provides level descriptors for all formal qualifications. In terms of LPCAT performance levels, the performance levels of groups at specific educational levels have been benchmarked against the NQF framework (see table 1).

These levels of performance can be used as a guideline in interpreting LPCAT results—to determine how close the individual is to the indicated LPCAT level associated with a particular qualification level—or in other words—how much effort will be required of the individual to reach the target level. Construct and predictive validity results for the LPCAT for samples at various educational levels have been reported (De Beer, 2000b, 2005, 2006; Van der Merwe & De Beer, 2006; Van Eeden et al., 2001).

### Goals of the Study

The present study provides longitudinal predictive validity results for the LPCAT and standard proficiency tests using aggregate academic performance as criterion measure. Results are presented for two separate studies which both involved gathering of criterion data two years after the initial assessment. In light of the stated need for validity evidence of DA results (Caffrey et al., 2008; Lidz, 2009; Sternberg & Grigorenko, 2002) these results were deemed appropriate to present as evidence for the longitudinal predictive validity of a learning potential test in comparison with standard (static) tests. The specific research questions were:

1. Are LPCAT learning potential results statistically significant predictors of aggregate academic performance at a secondary school level over a two-year period of time?
2. Are LPCAT learning potential results comparable to standard test results in terms of predicting aggregate academic performance at a secondary school level over a two-year period of time?

## Method

### Participants and Setting

Two independent sample groups from a boys' only school (N=79; mean age 12.44, SD=0.44) and a girls' only school (N=72 mean age 11.18, SD=0.42) respectively were both assessed at a senior primary level. The male sample group were at grade 7 level at the time of the assessment and was evaluated prior to entry to the junior secondary (Grade 8) level. The female sample group were at grade 6 level at the time of the assessment and was also evaluated prior to entry to the specific school. Assessment in both cases was aimed at obtaining information on the performance levels of individuals and of the groups on the selected measures to inform the schools of the profile of prospective pupils.

Both sample groups included pupils from different language and culture groups but the specific demographic information of individual pupils in terms of language and culture groups was not made available to the researcher.

### Instruments and Measures

Participants took the English proficiency test (Proficiency Test English second language intermediate level) (Chamberlain & Van der Schyff, 1991), two subtests of a standardised aptitude test (for English language and for calculations respectively) (Claassen, Van Heerden, Vosloo & Wheeler, 2000), and the LPCAT (De Beer, 2005). In addition, their aggregate academic performance scores were collected two years after the initial assessments were done.

**English proficiency test.** The English proficiency test (Proficiency Test English second language intermediate level) (Chamberlain & Van der Schyff, 1991) was administered to the

Table 1

*National Qualifications Framework Levels with Commensurate Qualifications and LPCAT Score Ranges*

LPCAT T-score range	LPCAT Stanine score	ABET* / NQF level	Educational
20 – 32	1	Abet level 1	Grade 0 – 3
33 – 37	2	Abet level 2	Grade 4 – 5
38 – 42	3	Abet level 3	Grade 6 – 7
43 – 47	4	Abet 4 / NQF 1	Grade 8 – 9
48 – 52	5	NQF level 1 – 3	Grade 10 – 12
53 – 54	6	NQF level 4 – 5	Grade 12+ (Higher Certificate)
55 – 57	6	NQF level 6	Diploma/Advanced Certificate
58 – 62	7	NQF level 7	3-year Degree/Adv. Diploma
63 – 68	8	NQF level 8	Honours / 4-year Degree/ Postgraduate Diploma
69 – 80 (65+)	9	NQF level 9	Advanced Degree (Master's)
69 – 80 (65+)	9	NQF level 10	Advanced Degree (Doctorate)

*Note:* Abet=Adult Basic Education and Training

male sample group. The English proficiency test contains written sections and tests knowledge of English grammar, vocabulary and reading comprehension by requiring the participants to read a passage and choose short answers using 40 multiple-choice format questions (Chamberlain & Van der Schyff, 1991). The reliability for this instrument is reported in the test manual as 0.89 (Chamberlain & Van der Schyff, 1991). Although the norms of this test are somewhat outdated, no other English proficiency test based on national samples was available at the time, and measurement of English proficiency was not deemed to have changed dramatically in the time concerned.

**Aptitude tests.** Subtests of a standardised aptitude test (for English language and for calculations respectively) (Claassen, et al., 2000) were administered to the female sample group. The English subtest (30 questions) gives an indication of proficiency in English, while the calculations subtest (40 questions) measures the ability to solve simple calculations quickly and accurately (Claassen et al., 2000). The reliability for these two subtests is reported as 0.86 for the English subtest and 0.78 for the calculations subtest (Claassen et al., 2000). These subtests were used as the school required comparative information with regard to English and calculations levels of the group and of individuals.

**The LPCAT.** The LPCAT— administered to both groups— is a dynamic computerised adaptive test for the measurement of learning potential (De Beer, 2005; 2006). Coefficient alpha reliability values for the LPCAT are reported as ranging between 0.926 and 0.978 (De Beer, 2005) and validity values with academic marks as criterion scores for a junior secondary group at a similar educational level as the current samples ranged between 0.44 and 0.54 for Grade 8 pupils and between 0.55 and 0.66 for Grade 9 pupils (De Beer, 2000b, 2006).

**Academic performance.** The aggregate academic performance of learners two years after the initial assessment was used to measure the academic performance results. For the male sample, academic marks for 15 school subjects were provided as percentage scores, while for the female sample academic marks for 13 school subjects were provided on a seven-point scale. For the female sample, in all but one subject, more than one mark was available per subject—hence a mean score was first calculated for the specific subject before aggregating all subject (mean) scores. For both samples, subject marks were added together in an aggregate score that was used as the criterion measure.

## Procedure

The two studies were conducted independently but for both groups, academic criterion results were collected two years after the initial assessment. In both cases, assessments were conducted at the school's request and the samples of convenience used cannot be considered representative of the general South African school population. Initial assessment of the male sample was done in 2004 when the learners were in Grade 7 and their academic results obtained in 2006 at the end of their Grade 9 school year. Assessment of the female sample took place in 2006 when the learners were in Grade 6 and their academic results were obtained in 2008 when they were in Grade 8. Assessments took place during normal school hours and tests were administered by registered psychologists. Standard instructions for administration were followed. The learners were tested in groups for both the paper-and-pencil-based and the computer-based assessments. The academic results of both groups were obtained after two years and the aggregate score used as the criterion measure, thereby providing longitudinal predictive validity information for all the predictor measures.

## Data Analysis

Descriptive, correlation and regression analyses were performed to answer the research questions. Descriptive statistics and correlation results were used to inspect the magnitude of scores and the relationships between variables. Regression analysis was used to evaluate the predictive validity of the LPCAT and to compare the predictive validity of the LPCAT results and standard test results for predicting aggregate academic performance after a two-year period of time.

## Results

Descriptive results provide an overview of the mean scores and dispersion of scores (see table 2 and table 3).

In terms of the LPCAT scores, the mean pre- and post-test scores for the male sample were at a NQF 6 to NQF 7 level—substantially higher than the learners' actual academic level at the time of their assessment. For the female sample, the mean pre- and post-test scores were at a NQF 5 level—slightly above the learners' academic level at the time of their assessment.

The correlation results show that for the male sample (see table 4), the LPCAT composite score shows the highest correlation with the dependent variable, while for the female sample (see table 5) the LPCAT post-test shows the highest correlation with the dependent variable. These correlations are statistically highly significant and also practically significant (Henson &

Table 2  
*Descriptive Statistics for the Male Sample Group*

	Minimum	Maximum	<i>M</i>	<i>SD</i>
LPCAT pre-test	44	62	54.52	4.38
LPCAT post-test	44	65	56.10	4.73
LPCAT difference score	-4	8	1.59	2.89
LPCAT composite score	44	62	54.78	4.39
English proficiency raw score	16	40	31.42	4.06
Aggregate academic score (% scale)	713	1736	1039.33	161.75

Note. N=79

Table 3

*Descriptive Statistics for the female Sample Group (N=72)*

	Minimum	Maximum	M	SD
LPCAT pre-test	37	65	50.01	5.56
LPCAT post-test	37	62	50.87	5.31
LPCAT difference score	-7	8	0.86	2.89
LPCAT composite score	37	65	50.13	5.53
English aptitude score	9	30	23.22	5.14
Calculations aptitude score	3	20	8.53	3.78
Aggregate academic score (7-point scale)	60.92	81.83	71.46	4.99

*Note:* N=72

Table 4

*Intercorrelations for the Predictor (Test) and Criterion (Aggregate Academic) Results for the Male Sample*

Variable	1	2	3	4	5	6
1-LPCAT pre-test	-					
2-LPCAT post-test	.804** (.000)	-				
3-LPCAT difference	-.206 (.068)	.416** (.000)	-			
4-LPCAT composite	.993** (.000)	.849** (.000)	-.121 (.286)	-		
5-English raw score	.298** (.008)	.405** (.000)	.213 (.060)	.328** (.003)	-	
6-School aggregate (% scale)	.471** (.000)	.390** (.000)	-.077 (.498)	.482** (.000)	.457** (.000)	-

*Note.* N=79; \*\* p<.01; \* p<.05

Table 5

*Intercorrelations for the Predictor (Test) and Criterion (Aggregate Academic) Results for the Female Sample*

Variable	1	2	3	4	5	6	7
1-LPCAT pre-test	-						
2-LPCAT post-test	.859** (.000)	-					
3-LPCAT difference	-.347 (.003)	.182 (.125)	-				
4-LPCAT composite	.998** (.000)	.876** (.000)	-.312** (.008)	-			
5-English raw	.211 (.076)	.263* (.025)	.078 (.516)	.213 (.072)	-		
6-Calculations raw	.273* (.020)	.278* (.018)	-.014 (.908)	.280* (.017)	.343** (.003)	-	
7-School aggregate (7-point scale)	.440** (.000)	.499** (.000)	.068 (.571)	.445** (.000)	.234* (.048)	.403** (.000)	-

*Note.* N=73; \*\* p<.01; \* p<.05

Smith, 2000). For the male sample, the LPCAT pre-test and composite score correlations with the dependent variable (school aggregate academic results) are numerically higher than the correlation of the English proficiency raw score with the dependent variable. However, statistical comparison of the magnitude of the correlations indicates no statistically significant difference (University of Amsterdam online statistical analysis).

Contrary to the results of a number of other studies (De Beer, 2000b), for the male sample, the LPCAT pre-test scores indicated a higher correlation with the aggregate academic score than the post-test score, albeit only marginally. The correlation of the composite score with the dependent variable is however higher than that of the pre-test score with the dependent variable, in line with the trend found in other studies (De Beer, 2000b, 2008).

In the case of the female sample, the correlations of the LPCAT pre-test, post-test and composite scores with the dependent variable are all numerically higher than those of the English aptitude and calculations aptitude raw scores respectively with the dependent variable. Once again, statistical comparison of the magnitude of the correlations indicates no statistically significant difference. The results for the female sample are similar to the majority of other LPCAT predictive validity results in that both the post-test score and the composite score show a higher correlation with the dependent variable than the pre-test score (De Beer, 2000b; 2008).

Multiple regression analyses using the Enter method were performed to evaluate the predictive validity of the LPCAT results for predicting aggregate academic results (research question 1) and also to compare the predictive validity of LPCAT learning potential results and standard test results for predicting aggregate academic results (research question 2). The standard test results and the LPCAT score showing the highest correlation with aggregate academic performance were entered as predictor variables. The latter was deemed appropriate due to

multicollinearity between the LPCAT pre-test, post-test and composite scores respectively. Academic aggregate score was used as the dependent variable for both sample groups.

Table 6 shows multiple regression analysis for the male sample, using as independent variables the LPCAT Composite score and English Raw score with the Enter method was statistically significant ( $p < .01$ ) in predicting academic aggregate results.  $R^2 = .33$ . Adjusted  $R^2 = .32$ . Altogether 33% (32% adjusted) of the variability of Academic Aggregate score was predicted by the two variables that were entered into the regression. Both variables entered contributed significantly ( $p < .01$ ).

Table 7 shows multiple regression analysis for the female sample, using as independent variables the LPCAT Post-test score, Calculations raw score and English Raw score with the Enter method was statistically significant ( $p < .01$ ) in predicting academic aggregate results.  $R^2 = .32$ . Adjusted  $R^2 = .30$ . Altogether 32% (30% adjusted) of the variability of Academic Aggregate score was predicted by the three variables that were entered into the regression. Two of the three variables (LPCAT Post-test and Calculations raw score) contributed significantly ( $p < .01$  and  $p < .05$  respectively).

From the correlation and regression results it appears that the LPCAT scores were significant predictors of academic performance two years after the initial assessment, therefore showing statistically highly significant longitudinal predictive validity. English proficiency and calculations aptitude respectively also made further statistically significant contributions to the accuracy of predicting aggregate academic results.

**Discussion**

For both samples, an LPCAT score (composite score in the case of the male sample and post-test score in the case of the female sample) showed the highest predictive validity with the academic aggregate score. In the case of the male sample, the English score also contributed significantly to the prediction of

Table 6

*Summary of Standard Multiple Regression Analysis using the Enter Method for the Male Sample Pupils' Aggregate Academic Results*

Variable	B	SEB	Beta	T	P
LPCAT composite score	13.717	3.658	.372	3.749	.000 **
English raw score	13.331	3.951	.335	3.374	.001 **

Note:  $R^2 = .33$ ; Adjusted  $R^2 = .32$ ; \*\*  $p < .01$ ; N=79

Table 7

*Summary of Standard Multiple Regression Analysis using the Enter Method for the Female Sample Pupils' Aggregate Academic Results*

Variable	B	SEB	Beta	t	P
LPCAT Post-test score	.389	.099	.413	3.908	.000 **
Calculations raw score	.367	.143	.278	2.560	.013 *
English raw score	.029	.105	.030	.279	.781

Note:  $R^2 = .32$ . Adjusted  $R^2 = .30$ ; \*\*  $p < .01$ ; \*  $p < .05$ ; N=72

the dependent variable (school aggregate results). In the case of the female sample, the calculations score also showed a further significant contribution to the accuracy of the prediction of the dependent variable (school aggregate results) over and above that already covered by the LPCAT post-test score.

Since similar results were reported for both groups, which were totally independent and from different schools, this provides empirical evidence in support of the utility of learning potential scores in the normal educational environment. This can be seen as a (partial) replication study, since the assessments at the two schools took place in different years in independent studies. The fact that the criterion results were obtained two years later—which implies the longitudinal predictive validity value of the learning potential results—further contributes to the available empirical evidence in support of the use of learning potential measures in mainstream educational environments. The results were not only statistically highly significant but also practically significant (Henson & Smith, 2000) with correlation effect sizes between medium and large.

To return to the initial aim, the results of the present study provide empirical evidence for the predictive validity of a learning potential or DA measure (see also Caffrey et al., 2008 as well as Grigorenko and Sternberg, 1998). Both sets of results also indicate that a learning potential measure contributes similarly to the accurate prediction of academic results in comparison with and alongside static tests. Furthermore, the LPCAT is comparable with standard tests in terms of ease of administration and testing time, thereby also addressing that concern of Grigorenko and Sternberg (1998). It can be administered in groups, has standard instructions, uses IRT-based CAT which enhances measurement accuracy and testing time, and because it is computerised, the results are available immediately on completion of the assessment. The direct link of the pre- and post-test scores to benchmarked educational levels enhances the practical utility of the results.

Based on these results, the LPCAT was shown to be suitable as a measuring instrument to predict academic performance, in combination with other results and information - in this case the English proficiency and calculations assessments. Since only nonverbal figural items are used in the LPCAT, verbal ability and other specific information required for the particular decision would still need to be obtained from other instruments.

### Limitations and Recommendations

The main limitation of the present study was that independent convenience samples were used comprising only 79 boys and 72 girls respectively, and that these samples therefore cannot be considered to be representative of all the students in the two schools or the South African student population in general. These factors limit the statistical power of the study and the generalisation possibilities of the results. The academic results for the two sample groups were also different in that percentage scores were obtained for the male group but the academic results for the female group were provided on a seven-point scale—thereby limiting direct comparison and precluding combining of the two groups. Differences in the test batteries administered also precluded combining of the two groups and is a further limitation.

Future research could expand on this study by examining larger and more representative samples to further investigate the relationship between different other static measures (such as English proficiency or numerical ability) and learning poten-

tial, on the one hand, and students' aggregate academic performance, on the other. Comparison with other more conventional static measures such as the Ravens could also provide useful additional information to further evaluate the utility of the LPCAT and other learning potential measures. It should be kept in mind that other specific aptitudes or skills would still need to be assessed by means of standard tests as measures of learning potential cannot provide all relevant information and are not intended to replace such standard measures.

According to Sternberg and Grigorenko (2002, p. 180), "it is difficult to argue that this approach has proven its usefulness and has shown distinct advantages over traditional static testing, relative to the resources that need to be expended." Research evidence in support of the utility of DA measures are accumulating (Bekmann, 2006; Caffrey et al., 2008; Haywood, 2008). The approach described and the results provided in the present study help to show that practitioners can use dynamic testing alongside mainstream cognitive assessments to contribute measurement information and improve decisions in contexts that require the kind of information made available through dynamic tests.

### References

- Beckmann, J. F. (2006). Superiority: Always and everywhere? On some misconceptions in the validation of dynamic testing. *Educational and Child Psychology, 23*(3), 35–49.
- 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, NY: Guilford Press.
- Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment. *The Journal of Special Education, 41*(4), 254–269.
- Chamberlain, J. C., & Van der Schyff, E. (1991). *Manual for proficiency test English second language advanced level*. Pretoria, South Africa: Human Sciences Research Council.
- Claassen, N. C. W. (1997). Cultural differences, politics and test bias in South Africa. *European Review of Applied Psychology, 47*(4), 297–307.
- Claassen, N. C. W., Van Heerden, J. S., Vosloo, H. N., & Wheeler, J.J. (2000). *Manual for the Differential Aptitude Test Form R (DAT-R)*. Pretoria, South Africa: Human Sciences Research Council.
- De Beer, M. (2000a). *Learning Potential Computerised Adaptive Test (LPCAT): User's Manual*. Pretoria, South Africa: Production Printers.
- De Beer, M. (2000b). *Learning Potential Computerised Adaptive Test (LPCAT): Technical Manual*. Pretoria, South Africa: Production Printers.
- De Beer, M. (2004). Use of differential item functioning (DIF) analysis for bias analysis in test construction. *South African Journal of Industrial Psychology, 30*(4), 52–58.
- De Beer, M. (2005). Development of the Learning Potential Computerised Adaptive Test (LPCAT). *South African Journal of Psychology, 35*(4), 717–747.
- De Beer, M. (2006). Dynamic testing: Practical solutions to some concerns. *South African Journal of Industrial Psychology, 32*(4), 8–14.
- De Beer, M. (2008, November). *Addressing some of the problems of dynamic assessment by using Item Response Theory and Computerised Adaptive Testing*. Paper presented

- at the 20th Anniversary Conference of the International Association for Cognitive Education and Psychology, Lake Louise, Alberta, Canada.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, 8(4), 341–349.
- Embretson, S. E. (2004). The second century of ability testing: Some predictions and speculations. *Measurement and Interdisciplinary Research and Perspectives*, 2(1), 1–32.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Employment Equity Act 55 of 1998. *Government Gazette*, 400(19370). Cape Town, 19 October 1998
- Feuerstein, R. (1979). *The dynamic assessment of retarded performance*. Baltimore, MD: University Park Press.
- Foxcroft, C. D. (2004). Planning a psychological test in the multicultural South African context. *South African Journal of Industrial Psychology*, 30(4), 8–15.
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, 42(2), 111–132.
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, 124(1), 75–111.
- Haywood, H.C. (2008). Twenty years of IACEP, and a focus on dynamic assessment: Progress, problems, and prospects. *Journal of Cognitive Education and Psychology*, 7(3), 419–442.
- Henson, R. K., & Smith, A. D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA task force report and current trends. *Journal of Research and Development in Education*, 33(4), 285–297.
- Hugo, H. L. E., & Claassen, N. C. W. (1991). *The functioning of the GSAT Senior for students of the Department of Education and Training*. Pretoria, South Africa: Human Sciences Research Council.
- Kim-Kang, G. & Weiss, D. J. (2008). Adaptive measurement of individual change. *Zeitschrift fur Psychologie [Journal of Psychology]*, 216(1), 49–58.
- Lidz, C. S. (1987a). *Dynamic assessment: An interactional approach to evaluating learning potential*. New York, NY: Guilford Press.
- Lidz, C. S. (1987b). Historical perspectives. In C.S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 3–32). New York, NY: Guilford Press.
- Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment*. New York, NY: Guilford Press.
- Lidz, C. S. (2009). *Dynamic assessment in school psychology*. National Association of School Psychologists Communique, 16–18.
- Murphy, R., & Maree, D. J. F. (2006). A review of South African research in the field of dynamic assessment. *South African Journal of Psychology*, 36(1), 168–191.
- Owen, K. (1998). *The role of psychological tests in education in South Africa: Issues, controversies and benefits*. Pretoria, South Africa: Human Sciences Research Council.
- Paterson, H., & Uys, K. (2005). Critical issues in psychological test use in the South African workplace. *South African Journal of Industrial Psychology*, 31(3), 12–22.
- Sijtsma, K. (1993a). Classical and modern test theory with an eye toward learning potential testing. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijsenaars, *Learning potential assessment: Theoretical, methodological and practical issues* (pp. 117–134). Amsterdam, Netherlands: Swets & Zeitlinger.
- Sijtsma, K. (1993b). Psychometric issues in learning potential assessment. In J. H. M. Hamers, K. Sijtsma & A. J. J. M. Ruijsenaars, *Learning potential assessment: Theoretical, methodological and practical issues* (pp. 175–194). Amsterdam, Netherlands: Swets & Zeitlinger.
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge: Cambridge University Press.
- University of Amsterdam online statistical analysis. Retrieved from [http://www.fon.hum.uva.nl/Service/Statistics/Two\\_Correlations.html](http://www.fon.hum.uva.nl/Service/Statistics/Two_Correlations.html)
- Van der Linden, W. J. (2008a). Some new developments in adaptive testing technology. *Zeitschrift fur Psychologie [Journal of Psychology]*, 216(1), 3–11.
- Van der Linden, W. J. (2008b). Adaptive models of psychological testing. *Zeitschrift fur Psychologie [Journal of Psychology]*, 216(1), 1–2.
- Van der Merwe, D., & De Beer, M. (2006). Challenges of student selection: Predicting academic performance. *South African Journal of Higher Education*, 20(4), 547–562.
- Van Eeden, R., De Beer, M., & Coetzee, C. H. (2001). Cognitive ability, learning potential, and personality traits as predictors of academic achievement by engineering and other science and technology students. *South African Journal of Higher Education*, 15(1), 171–179.
- Vygotsky, L. S. (1978) *Mind in society: The development of higher-order psychological processes*. Cambridge, MA: Harvard University Press.
- Weiss, D. J. (Ed.). (1983). *New horizons in testing: Latent trait test theory and computerised adaptive testing*. New York, NY: Academic Press.

### Author Note

This article is based in part on a presentation made at the 20th Anniversary Conference of the International Association for Cognitive Education and Psychology (IACEP), held in Lake Louise, Alberta, Canada from 2–5 November 2008. Attendance of the conference was made possible by a grant from the National Research Foundation (NRF) of South Africa (Grant number IFR2008051700003). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and therefore the NRF does not accept any liability in regard thereto.

