Journal of Psychology in Africa 2010, 20(2), 241–246
Printed in USA - All Rights Reserved

JOURNAL OF PSYCHOLOGY IN AFRICA
ISSN 1433-0237

A Modern Assessment Psychometric Approach to Dynamic Assessment

Marié de Beer
University of South Africa

Address correspondence to Marié de Beer, Department of Industrial and Organizational Psychology, University of South Africa, P.O. Box 392, Pretoria 0003, South Africa. E-mail: dbeerm@unisa.ac.za

Although dynamic assessment (DA) has been hailed as a positive move towards fair assessment, it has generally not been used in educational or industry settings to the same extent that standard (static) tests have been. The present article attempts to elucidate how the use of Item Response Theory (IRT) and Computerised Adaptive Testing (CAT) can address some of the problems typically associated with dynamic assessment. An example of a DA tool that makes use of IRT and CAT, shows acceptable psychometric properties and is comparable to standard tests in terms of ease of administration illustrates the possibility of wider application of DA in both educational and industry settings.

Keywords: dynamic assessment (DA), item-response theory (IRT), computerised adaptive testing (CAT)

In the world history of psychometrics—and in particular in the domain of cognitive assessment—dynamic assessment (DA) has a very special place. DA refers to an assessment approach that includes a learning opportunity during assessment in order to provide information on the current as well as the potential future performance levels of the individual being assessed—typically by means of a test-train-retest process and with the aim of measuring both the level and rate of learning (Caffrey, Fuchs & Fuchs, 2008; Grigorenko, 2009; Lidz, 2009). Vygotsky (1978) is often referred to as the father of DA with his concept of the zone of proximal development (ZPD), but the work of Binet and Simon (1905; 1915) at the beginning of the 20th century can actually be regarded as the first empirical use of dynamic assessment principles for the measurement of learning potential. Binet and Simon were tasked by the French government of the time to develop a measure for conducting assessments among poorly performing school children, to distinguish those who could benefit from further training and educational interventions—to such an extent that they would be able to join normal educational settings—from those who would probably not benefit and would need to remain in special educational settings (Wolf, 1973).

Brief Overview

Despite its original dynamic test characteristics, Binet and Simon‘s test came to be used as a static cognitive ability test and is often indicated as the forerunner of standard static cognitive tests as they are generally used today (Fancher, 1985; Van der Linden, 2008a; Wolf, 1973). A further interesting feature of the Binet-Simon test is that, in principle, it was also an adaptive test, with features such as variable entry level, continuous scoring and a termination rule based on performance (and accuracy of measurement) (Van der Linden, 2008a). In the light of research and work in the field of cognitive assessment that followed in the next 100-plus years, the groundbreaking contribution of these scientists can be appreciated even more.

Although DA has generated considerable interest and research, it has not been absorbed into mainstream assessment commensurate with expectations (Grigorenko & Sternberg, 1998; Lidz, 2009).

Various authors have reviewed the successes, contributions and limitations of dynamic assessment (Beckmann, 2006; Caffrey et al., 2008; Grigorenko, 2009; Grigorenko & Sternberg, 1998; Lidz, 2009; Murphy & Maree, 2006; Sternberg & Grigorenko, 2002). Sternberg and Grigorenko (2002, p. viii) described dynamic assessment as “a wonderful idea whose implementation for the most part has been less than fully satisfactory”. Limiting factors noted by them are the dearth of published research on psychometric information about dynamic assessment tools, as well as practical elements such as measurement problems when comparing pre- and post-test scores and time-consuming administration. According to a number of authors (Grigorenko & Sternberg, 1998; Lidz, 2009), these factors have limited the wider implementation and use of dynamic assessment alongside standard static assessments in mainstream education and other domains. The aim of this article is to elucidate and demonstrate how the use of IRT and CAT can address a number of these limitations.

According to Sternberg and Grigorenko (2002, p. 30), “multiple attempts to quantify learning potential and to transform the construct of dynamic testing into a set of robust psychological diagnostic tools have not produced consistent results. Nevertheless, the idea of dynamic testing is so appealing that, despite its relatively sparse empirical validation, it has been widely discussed and fairly widely used.” They question whether this approach can be supported by results and techniques that show through empirical results that it can provide information “over and above the data collected by conventional tests” (Sternberg & Grigorenko, 2002, p. 181). In a meta analysis of available predictive validity results for DA measures, Caffrey et al. (2008) reported significant unique predictive validity shown by DA measures over and above that shown by traditional cognitive tests in predicting future academic achievement. However, Lidz (2009, p. 16) report that these methods are not generally used due to “insufficient supportive evidence and heavy time requirement.”

Murphy and Maree (2006) provided an extensive overview of dynamic assessment research done in South Africa. Other authors (Caffrey et al., 2008; Grigorenko & Sternberg; Sternberg & Grigorenko; 2002) have also reviewed work done in other regions in the world. It therefore does seem that there are selected areas or specific fields of application where DA is

used, but it has not yet seen the kind of general application that is seen with standard (static) tests (Lidz, 2009). A partial explanation for the limited application of DA could be that within the field itself, different approaches and areas of application exist which share certain commonalities, but which also differ in important ways.

DA for Measurement of Learning Potential

For the purpose of this article, two broad approaches to DA are distinguished, namely a more clinical, diagnostic and remedial approach, on the one hand, and a psychometric, measurement-oriented, comparative approach, on the other. This distinction links to the ideographic versus the nomothetic approaches (Hermans, 1988) and with specific reference to DA was referred to as clinically-oriented DA versus research-oriented DA by Caffrey et al. (2008). With regard to DA, the two approaches share the following similarities: 1) both are based on the Vygotskyan concept of the Zone of Proximal Development (ZPD); 2) both use a test-train-retest strategy—assuming that standard tests do not reflect optimal performance levels; both aim to identify learning potential; and 3) they both provide training relating to the assessment task. However, there are also distinct differences between the two approaches. The first approach is more diagnostic, remedial, individually oriented and typified by the Feuerstein model (Caffrey et al., 2008; Grigorenko, 2009) and often involves assessment of low-performing individuals. In this approach, procedures are not necessarily standardised and the training or learning experience provided is determined by the ability profile, performance and problem areas of the particular individual assessed. Since inter-individual comparison is not the focus here, the lack of standardisation in the learning experience does not represent a limitation. The second approach is more measurement oriented, comparative or norm-based and involves assessment of individuals who perform within the normal range of ability. This latter approach incorporates the principles of dynamic assessment by providing a learning opportunity during assessment, but it is more focused on accurate measurement, the psychometric properties of the instruments used and ease of administration to allow for group assessments, thereby making these tools more comparable with standard assessment measures (Caffrey et al., 2008). Since the learning provided in this approach is standardised, inter-individual comparison is possible. This more standardised, measurement-oriented dynamic assessment approach provides the kind of empirical psychometric evidence that addresses some of the concerns noted by Grigorenko and Sternberg (1998).

Standard or Static Tests

Standard cognitive assessment has a long and extensive history of use in a number of domains – for vocational guidance in education, and in industry for screening and selection, for appointments and promotions as well as training and development. These tests provide useful information for decision making, as evidenced in the predictive validity values typically found. However, in the last few decades there has been an increased need for measures of learning potential to be used alongside standard cognitive assessment (Lidz, 2009; Murphy & Maree, 2006). Standard (static) tests take into consideration only current levels of performance and are based on an implicit assumption that those who are tested had similar opportunities to develop the constructs being measured. Among other things, differences in socioeconomic and educational background have an effect on the development of cognitive-related constructs in particular (Claassen, 1997; Foxcroft, 1997; Paterson & Uys, 2005). Educationally and socioeconomically disadvantaged individuals are often at a further disadvantage when standard cognitive tests are used, because these tests typically include content representing crystallized abilities (ie language proficiency, scholastic content or educational material) influenced by prior learning experiences (Claassen, 1997; Foxcroft, 1997; Van de Vijver, 1997, 2002). In South Africa in particular, differences in socioeconomic and educational levels (StatsSA, 2008) contribute to social and political strife around employment and educational and developmental opportunities.

By contrast, dynamic tests, with their focus on measuring learning potential, make provision for a learning opportunity during assessment, to allow for measurement of the projected future or potential levels of performance (Caffrey et al., 2008; Grigorenko, 2009). They can therefore provide information that can be well used in decisions relating to screening and selection as well as training and development opportunities. This approach takes into consideration the fact that many individuals may have had less than optimal learning opportunities and therefore might not yet have attained their optimal levels of performance. The measurement of learning potential allows for fairer assessment of disadvantaged individuals for two reasons: firstly, the assessment includes a learning experience, and secondly, the test items typically measure fluid reasoning ability, which is less influenced by socioeconomic and educational background and prior learning experiences.

Research has indicated that intelligence quotient (IQ) scores from static tests are influenced by educational opportunity and socioeconomic factors and are subject to change (Claassen, 1997; Grigorenko & Sternberg, 1998; Vincent, 1991). This has partly led to the increased attention that learning potential measures have received. Improved socioeconomic and educational opportunities typically result in increases in the mean IQ scores over and above normal population increases over time (Van de Vijver, 1997; Vincent, 1991). Based on a household survey conducted in 2007, with regard to educational attainment of persons aged 20 years and older, only 33.4% have completed a Grade 12 (senior secondary 23.6%) or higher (9.8%) level of qualification (StatsSA, 2008). The high proportion of this group with attained qualification levels lower than a completed senior secondary level emphasizes the need for the measurement of learning potential to provide training and development opportunities on a fairer basis for individuals at all educational levels.

Learning Potential Assessment

In order to utilize learning potential assessment for participants who perform within the normal range, Vygotsky‘s (1978) work needs to be revisited. Vygotsky (1978) proposed dynamic testing as a means of measuring what he referred to as the ZPD, referring to the zone between an individual‘s independent (actual) and guided or assisted (potential) levels of performance. In Vygotsky‘s view, the task of assessment is to identify not only those cognitive processes that are fully developed, but also those that are in a state of being developed at the time of assessment and which can be identified by incorporating learning as part of the assessment procedure (Dörfler, Golke, & Artelt, 2009). The idea is that a dynamic test would measure a person‘s ability to profit from learning opportunities, guidance and assistance and thereby provide insight into his or her learning potential. The more the individual improves on his or her ini-

tial (pre-test) score, the larger the ZPD will be. However, in some instances, the ZPD (difference score) has been the main focus and has incorrectly been interpreted as a direct measure of learning potential. This could be appropriate for research into learning potential, where participants are either mentally retarded or low-level performers. For these (and only these) individuals, the improvement score or ZPD can be interpreted as a measure of learning potential, since they all start off at a similar (low) level of test performance. Vygotsky clearly indicated that current as well as improved performance levels should be taken into account when interpreting learning potential (Vygotsky, 1978). This allows for the use of learning potential assessment over a broad spectrum of candidates and for comparison of scores across a wide range of ability and potential.

Implications for Practise

From a practical perspective, measurement of learning potential is typically used in the context of learning and training, with the logical result that actual developmental level (pre-test performance) also needs to be considered in learning potential assessment. Measurement of actual and potential levels of performance with consideration of the improvement in performance the individual shows following a learning experience represents a simulation. The assumption is that examinees are likely to show commensurate levels of performance and utilize learning opportunities in a similar fashion in real life.

The reason for assessment is important, since the question at hand is not merely whether individuals show improved performance following a training intervention. For practical purposes, what is important is whether individuals assessed currently (as evidenced in their pre-test level of performance) meet the level of reasoning skills required by the training or development opportunity, or whether they show that they have the potential (as indicated by their post-test level of performance) to reach the required level. The focus is not only on whether the individual will be able to profit from learning, but also on the level of training he or she will be able to cope with – or alternatively, to what degree he or she seems able to cope with a particular level of training offered. Thus in practical terms, current and projected future (potential) levels of performance are compared with the opportunity (i.e. target level of training) to evaluate whether the individual is currently already performing at the required level or shows the potential to perform at the required level.

Predictive validity for dynamic assessment across normal ability ranges uses the pre- and post-test scores as predictors respectively since the difference score is not expected to correlate with criterion performance measures, as explained by Jensen (1963, p. 1):

> *when improvement with practice is thus measured from a different baseline for every subject, the results can be confusing and are often uninterpretable. A subject who is initially good at the task is already near the asymptote of his learning curve and can therefore show but little gain or improvement with practice. The slowest learners can often show the greatest gain. Consequently, correlations between gain scores on various learning tasks and psychometric measures of intelligence usually average close to zero.*

Dynamic assessments are efficiently delivered using IRT approaches and with CAT administration. These applications are considered next.

The Use of IRT and CAT in DA

In IRT correspondence can be established between observed performance and individuals' locations on the latent trait being measured (De Ayala, 2009). This permits accurate measurement of the latent trait being measured – and with regard to DA, also the accurate measurement of difference or improvement scores. Furthermore person and item parameters are theoretically invariant (De Ayala, 2009), meaning item properties are sample independent, while person ability parameter is item independent (Van der Linden, 2008a; Weiss, 1983). This, in combination with the key characteristic of IRT - that item difficulty and person ability are measured on the same scale – makes possible CAT, which involves the ongoing interactive selection of items from a precalibrated item bank to continually match the respondent's estimated ability level (Van der Linden, 2008a; Weiss, 1983). Frey and Seitz (2009) report that in CAT the number of items can be reduced between 50% and 70%, compared to a fixed number of items conventional test - without loss of measurement precision. The IRT-based CAT approach also counters the floor and ceiling effects often encountered in static testing.

IRT and CAT address a number of the concerns regarding dynamic testing (Embretson, 1996; Embretson & Reise, 2000; Sijitsma, 1993a; 1993b), with CAT allowing for shorter testing times and IRT providing improved accuracy of measurement of individual difference scores (Kim-Kang & Weiss, 2008), as well as improved means to compare the scores of the same or different examinees since observed performance is linked to location on the latent trait being measured. "Because of its ability to equate testings and link item pools onto a common metric, IRT has the potential of offering solutions to the problem of measuring gains in achievement levels during the process of instruction" (Weiss, 1980, p. 8). IRT and CAT procedures seem particularly appropriate for learning potential assessment because they improve both measurement accuracy and time efficiency resulting in testing times comparable to those of standard cognitive tests (Dörfler et al., 2009; Van der Linden, 2008b). This modern psychometric approach therefore addresses the further concerns noted by Grigorenko and Sternberg (1998) and Sternberg and Grigorenko (2002) as limiting the broader application of dynamic assessment and provides a modern-day solution to ensure fair, accurate and effective measurement of learning potential.

An Example Illustrating the Use of IRT and CAT in DA in the South African Context

The Learning Potential Computerised Adaptive Test (LPCAT) addresses some of the concerns that have been noted regarding dynamic assessment (De Beer, 2005). Three different types of nonverbal figural item formats are used 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). The LPCAT uses a test-train-retest format with two separate but linked adaptive tests, in an attempt to measure learning potential in the fluid reasoning ability domain so that language proficiency or formal academic qualifications should not impact significantly on performance (De Beer, 2000a; 2000b). Factor analysis results indicate a definite unidimensional structure for the item bank with coefficient alpha values ranging between 0.926 and 0.978 for different groups

(De Beer, 2005). Following IRT-based Differential Item Functioning (DIF) analyses, items that indicated DIF above a set level for sub-groups based on gender, level of education, first language and culture were discarded and not used in the final compilation of the LPCAT (De Beer, 2000b 2005).

Assessment equity issues. In the South African context the Employment Equity Act legislation of 1998 requires empirical evidence of the reliability, validity and fairness of psychological assessment instruments. The LPCAT shows coefficient alpha reliabilities for the item bank ranging between 0.925 and 0.987. Construct and predictive validity values vary depending on the samples and specific comparison measures used but construct validity values ranging between 0.100 and 0.713 and predictive validity values ranging between 0.008 and 0.610 have been reported (De Beer, 2000b, 2006). Furthermore, the following features could be considered positive in terms of fairness in the multicultural and multilingual South African context:

- It is a power test that focuses on measurement of general non-verbal figural (fluid) ability and has no overall time limit.
- It is a dynamic test that allows for learning during assessment, with the focus not only on the current level of performance but also on the potential level of performance after relevant training has been provided.
- It makes use of IRT measurement principles which links performance to the individuals' standing on the latent trait being measured and which allows for more accurate inter individual comparison of scores and more accurate measurement of intra-individual difference scores
- It makes use of CAT, whereby items are administered to match the estimated level of performance throughout and which decreases testing time significantly
- It requires the use of only the space bar and Enter key, for ease of administration and thus allows for assessment of individuals with low literacy (and computer literacy) levels.
- The test instructions are available in all 11 official South African languages and one version of the LPCAT allows for instructions to be read to the individuals being tested thereby not requiring them to do any reading themselves (De Beer, 2005; 2006; 2008). A second version of the test which allows examinees to independently read the instructions and feedback from the screen can be selected during test administration.

Efficient measurement with the LPCAT. The LPCAT incorporates all the advantages of IRT and CAT, 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 (Van der Linden, 2008a; 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, 2000b, 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) show higher predictive validity correlation values compared to the pre-test score (which indicates current level of performance) (De Beer, 2000b; 2008). These results therefore provide support for the concept of measurement of learning potential because they show that allowing credit (whether fully or partially) for learning and improved performance after the learning experience improves the predictive validity, compared to using only current levels of performance (De Beer, 2000b; 2008).

The LPCAT'S measurement of nonverbal figural fluid reasoning ability indicated by levels of performance in the pre- and post-tests have been benchmarked against typical levels of such reasoning ability shown at different educational levels. In this way, the individual's current and projected levels of LPCAT (fluid ability reasoning) performance can be compared with the target or required level of reasoning to ascertain the likelihood of success (or, alternatively interpreted, the amount of effort the individual will have to expend in order to achieve success at the required level). This aligns with Vygotsky's notion (1978, p. 85) that it is "a well known and empirically established fact . . . that learning should be matched in some manner with the child's developmental level".

Benchmarking of qualifications. In South Africa, all qualifications are benchmarked against a National Qualifications Framework (NQF). This framework initially had eight levels, but has recently been expanded to ten levels, as indicated in Table 1 alongside benchmarked LPCAT score ranges and formal qualification levels.

In practice, if individuals are being screened for training and development opportunities or for appointments to positions, the starting point for the decision-making process is to determine the required NQF level (qualification level or the level of training) and the commensurate LPCAT score range benchmarked with empirical assessment results of groups at various educational levels and with predictive validity evidence at all levels (De Beer, 2000b, 2006). Next, the pre- and post-test levels of performance should be compared with the required level (see Table 1), to evaluate whether the individual currently (pre-test level) or potentially (post-test level) is at or comes close to the reasoning level or range required.

An alternative method of interpretation is to interpret the possible gap between the current and/or potential levels of performance and the required levels of performance—if lower than required—as the amount of effort that will be needed from of the individual to attain success at the required level. In particular, in contexts where educational and socioeconomic opportunities are not necessarily equal, current and potential levels of fluid reasoning ability that rely less on scholastic or educational background can assist with decision making and the planning of training and development opportunities. However, it should be noted that since the LPCAT measures fluid reasoning ability only—by means of nonverbal figural patterns—other relevant information such as verbal or numerical skills or specific aptitudes that may be required for the opportunity at hand needs to be obtained from other assessments or sources. It should be noted that dynamic tests are not intended to replace static tests but rather to add information that is not available from static tests, that is, the current as well as potential levels of performance, by allowing for learning during the assessment process (Caffrey et al., 2008).

The LPCAT compares well with standard tests in terms of its psychometric properties, test administration time and ease of interpretation, and is considered fair for use with disadvantaged examinees (De Beer, 2000b; 2006). Construct and predictive validity results at various educational levels have been reported (De Beer, 2000b; 2006; Van der Merwe & De Beer, 2006).

Table 1

National Qualifications Framework (NQF) 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

Reflective Summary

To return to the initial aim, this article attempted to elucidate how IRT and CAT can be used to address a number of problems associated with DA and explained how these modern assessment principles and methods were used in the construction of the LPCAT.

The LPCAT is has shown satisfactory psychometric properties and is comparable with standard tests in terms of ease of administration and testing time, thereby also addressing those concerns mentioned by Grigorenko and Sternberg (1998). It can be administered individually or in groups, has standard instructions, makes use of 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 fact that the pre- and post-test scores can be directly linked to benchmarked educational levels improves the practical utility of the results.

It serves as a practical example of the implementation of IRT and CAT to enhance the utility of DA alongside standard tests in educational and industry settings.

References

Beckmann, J.F. (2006). Superiority: Always and everywhere? On some misconceptions in the validation of dynamic testing. *Educational and Child Pychology, 23*(3), 35–49.

Binet, A., & Simon, T. (1905/1916). *The intelligence of the feeble-minded*. Baltimore, MD: Williams & Wilkins.

Binet, A., & Simon, T. (1915). *A method of measuring the development of the intelligence of young children*. Chicago, IL: Chicago Medical Book Co.

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment. *Journal of Special Education*, *41*(4), 254–269.

Claassen, N. C. W. (1997). Cultural differences, politics and test bias in South Africa. *European Review of Applied Psychology, 47*(4), 297–307.

De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: The Guilford Press.

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. (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. (November, 2008). 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.

Dörfler, T., Golke, S., & Artelt, C. (2009). Dynamic assessment and its potential for the assessment of reading competence. *Studies in Educational Evaluation, 35*, 77–82.

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, *8*(4), 341–349.

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, South Africa, 19 October 1998

Fancher, R. E. (1985). *The intelligence men: Makers of the IQ controversy*. New York, NY: WW Norton.

Foxcroft, C. D. (1997). Psychological testing in South Africa: Perspectives regarding ethical and fair practices. *European Journal of Psychological Assessment*, *13*(3), 229–235.

Foxcroft, C. D. (2004). Planning a psychological test in the multicultural South African context. *South African Journal of Industrial Psychology*, *30*(4*),* 8–15.

Frey, A., & Seitz, N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation, 35*, 89–94.

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.

Hermans, H. J. M. (1988). On the integration of nomothetic and ideographic research methods in the study of personal meaning. *Journal of Personality, 56,* 785–812.

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.

Jensen, A. R. (1963). Learning ability in retarded, average, and gifted children. *Merrill-Palmer Quarterly, 9*(2), 123–140.

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. (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. Ruijssenaars (Eds.), *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. Ruijssenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues* (pp. 175–194). Amsterdam, Netherlands: Swets & Zeitlinger.

StatsSA, *vide* Statistics South Africa.Statistics South Africa. (2008). General household survey 2007. Report published by Statistics South Africa.

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

Van de Vijver, F. (1997). Meta-analysis of cross-cultural comparisons of cognitive test performance. *Journal of Cross-cultural Psychology*, *28*(6), 678–709.

Van de Vijver, F. (2002). Cross-cultural assessment: Value for money? *Applied Psychology: An International Review*, *51*(4)*,* 545–566.

Van der Linden, W. J. (2008a). Some new developments in adaptive testing technology. *Zeitschrift fur Psychology [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.

Vincent, K. R. (1991). Black/white IQ differences: Does age make the difference? *Journal of Clinical Psychology*, *27*(2), 266–270.

Vygotsky, L. S. (1978) *Mind in society: The development of higher-order psychological processes*. Cambridge, MA: Harvard University Press.

Weiss, D. J. (1980). *Final report: Computerised adaptive performance evaluation*. Minneapolis MN: University of Minnesota, Department of Psychology.

Weiss, D. J. (Ed.). (1983). *New horizons in testing: Latent trait test theory and computerised adaptive testing*. New York, NY: Academic Press.

Wolf, T. (1973). *Alfred Binet*. Chicago, IL: University of Chicago 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.