# DYNAMIC TESTING:
# PRACTICAL SOLUTIONS TO SOME CONCERNS

MARIÉ DE BEER
*Department of Industrial and Organisational Psychology*
*UNISA*

## ABSTRACT

Although the concept of dynamic testing is generally well supported, some concerns have been raised about the general and more widespread practical application and use of these procedures. The main problem areas that have been noted are the amount of time needed to administer dynamic assessment measures (compared with standard tests), the lack of standardisation of procedures (impacting on the accuracy of measurement and interpretation) and the limited validity information available for dynamic measures. This article provides background on and research results of the Learning Potential Computerised Adaptive Test (LPCAT), a dynamic learning potential test. The methodology used in this instrument addresses some of the concerns that have been noted about dynamic assessment, namely the use of standardised procedures; the improvement of measurement accuracy with the use of Item Response Theory (IRT); and the use of Computerised Adaptive Testing (CAT) to improve the time efficiency of dynamic testing and obtaining empirical research results for construct and predictive validity.

**Key words**
Dynamic testing, learning potential

Dynamic testing as a research area is attracting increased attention. It is lauded as an approach that lends itself to fairer multicultural cognitive assessment, on the one hand, whereas some concerns have been voiced about its broader implementation, on the other. These and other issues will be discussed in this article, with specific focus on new statistical and testing methods that can help to address some of the concerns. Empirical results providing support for the psychometric properties of the LPCAT are provided, as well as a general evaluation of the way in which it addresses some of the concerns voiced by Grigorenko and Sternberg (1998).

### Historical beginnings of cognitive assessment

Recurring patterns are often evident in history. From about the turn of the previous century, scientists and researchers contributed fundamentally to the subject field of psychological assessment. With specific reference to cognitive assessment, Binet and Simon's (1905/1916) contribution can be appreciated to a greater extent in the light of some of the more recent trends in cognitive assessment in terms of the content used, the general approach adopted and the specific test techniques applied.

Binet and Simon (1905/1916) were tasked by the French government of the time to develop an instrument to identify, amongst low-performing school children, those who could benefit from further training from those who would probably not (Wolf, 1973). This focus, of paying attention to current performance, but also allowing for a possible improvement in performance if relevant opportunities, exercises and training can be provided, is at the heart of dynamic testing. Binet and Simon did ground-breaking work in terms of the measurement of cognitive ability and were the first researchers to use cognitive tasks and not physiological or reaction-time measures in the measurement of cognitive ability (Wolf, 1973). Hence, most cognitive tests developed in the last one hundred plus years, reflect tasks similar to those that Binet and Simon introduced.

In addition, the test developed by Binet and Simon (1905/1916) had all the characteristics of a dynamic (learning potential) test, with a basic assumption that what is measured is changeable. The latter assumption refers specifically to the concept of learning potential – that is, maintaining or improving on current levels of performance when relevant learning opportunities are provided – which relates back to the brief Binet and Simon received from the French educational authorities.

Furthermore, adaptive testing also has its roots in the work of Binet and Simon, because, according to Reckase (1988), the procedure they followed is analogous to what happens in adaptive testing: Firstly, the Binet-Simon test had a variable entry level – with the examiner starting to administer items at the individual examinee's level of estimated ability. Secondly, items were scored during administration and the results used for further branching and selection of additional items. Lastly, the test featured a variable termination criterion which resulted in different individuals receiving varying numbers of items. The test was terminated when a ceiling level was reached (Weiss, 1983)

It is interesting to note that all of the above are features of CAT procedures, although the latter is somewhat more sophisticated with the use of computer technology. With hindsight, these early researchers made a huge contribution to the field of cognitive assessment – which can be better appreciated when considering how the field has evolved over time.

### Modern trends in cognitive assessment

Testing in multicultural contexts has become one of the key concerns in cognitive assessment. In an attempt to provide more equitable cognitive assessment in the last few decades, dynamic testing and the measurement of learning potential have received increasing attention both locally and internationally (Lidz, 1987a, 1987b; Murphy, 2002, 2006). Societal needs drive research, and dynamic testing was first intensively researched in the 1960s and 1970s as measures that: (1) could provide more culture-fair assessment; (2) would be useful for comparing results obtained in culturally diverse populations; (3) would be appropriate for testing individuals with deprived educational experiences; and (4) could measure learning potential distinct from what has been learned – regardless of the culture, population, or social group of a tested individual (Grigorenko & Sternberg, 1998).

Measurement of learning potential typically involves a test-train-retest strategy with some form of help or training provided as part of the assessment process. Hence, it specifically provides useful information for training and development purposes. The provision of a learning opportunity in the test administration provides fairer assessment of the disadvantaged groups in particular. By providing a learning opportunity in the assessment, the focus is not only on the present level of performance (possibly reflecting limitations of the examinee's background), but also on the potential future levels of performance that can be achieved if relevant training can be provided.

While learning potential assessment provides alternative and supplementary information in the cognitive reasoning domain, researchers have lamented the fact that limited empirical research is hampering its progress (Grigorenko & Sternberg, 1998; Gupta & Coxhead, 1988; Guthke, 1992, 1993a, 1993b). Although the concept of dynamic assessment is generally well supported, some doubts have been raised about the general and more widespread practical application and use of these procedures.

The article written by Grigorenko and Sternberg on dynamic assessment in 1998 provides a thorough review of work done in this field up to that point in time. Although they were generally appreciative of and encouraging about the work done in this field, they still indicated quite clearly that specific elements would need to be addressed before this approach could optimally contribute to and be considered fully part of assessment practices in general.

Dynamic assessment has been proposed as a way of uncovering information about the extent to which developed abilities reflect latent capacity (ie the difference between latent capacity and developed abilities). In dynamic tests, what is tested is not merely previously acquired knowledge, but also the capacity to master, apply and reapply knowledge taught in the dynamic testing situation. The goal of dynamic testing is to see whether and how the subject will change if an opportunity is provided. However, according to Grigorenko and Sternberg (1998), multiple attempts to quantify learning potential and transform such testing into robust psychological diagnostic tools have not produced consistent results. There is a paucity of published empirical data on the reliability and validity of dynamic testing.

The principal application of dynamic testing is with disadvantaged individuals (ie people with unequal learning opportunities because of deficient previous education) who often perform poorly on conventional static tests. Dynamic testing should reduce the effect of educational inequalities by providing what are seen as more compassionate, fair and equitable means for assessing learning capacity.

Vygotsky's theory of the Zone of Proximal Development (ZPD) appears to have been the first nearly complete theory of dynamic testing (1978). ZPD reflects development itself – it is not only about what one is, but what one can become; it is not only what has developed but what is developing and can be viewed as a means to improve the testing of individual mental functioning. Experimental validation of the ZPD construct is extremely rare – with little empirical validation. Research conducted to date has not produced convincing quantitative empirical data to support the broad claim that ZPD-based teaching results in better educational and cognitive outcomes.

Some of the limitations of dynamic testing mentioned are that many dynamic test applications use standardised psychometric instruments in dynamic mediational modes, thus not using instruments specifically developed for dynamic testing. Furthermore, target groups for dynamic testing are often low performers, which limits the applicability of dynamic testing for groups at varying ability levels. In dynamic assessment, the role of the examiner varies from very important (clinical diagnostic orientation) to more limited (measurement or psychometric orientation). In general, the more clinical and unstandardised the approach – possibly with diagnosis or enrichment as the focus – the less comparable the results of different individuals become.

South African researchers have contributed both in the development of instruments for the measurement of learning potential and also in research contributing to the available information on the validity of dynamic testing measures (Boeyens, 1989; De Beer, 2000a, 2000b, 2003, 2005; Lopes,

Roodt & Mauer, 2001; Shochet, 1992, 1994; Taylor, 1992, 1994a, 1994b; Van Eeden, De Beer & Coetzee, 2001). Murphy (2002, 2006) provides an extensive overview of South African research in dynamic assessment, providing a mixture of positive and negative results with a variety of approaches and methodologies used. For the current article, some background on and research results for the LPCAT - which was developed in South Africa, and which addresses some of the concerns that have been noted about dynamic testing - will be provided.

**Dynamic testing**
Dynamic testing refers to testing procedures that include a learning experience as part of the assessment to obtain information not only about the outcome of learning up to that point in time, but also about the potential to learn and possibly improve on levels of performance when relevant learning opportunities can be provided. The aim of the dynamic test-teach-retest approach, with the focus on measurement of learning potential, is to provide learning opportunities in the assessment situation to enable examinees to optimise their test performance (Campione & Brown, 1987; Hamers & Resing, 1993; Lidz, 1991). This approach acknowledges the differences with which examinees come to the testing situation. The pretest provides an indication of the present (actual) level of performance attained – similar to that which is typically assessed in standard tests. The training is aimed at providing further examples, hints and guidelines that will highlight important aspects of information required to help solve similar questions. The post-test then provides an indication of the potential future level of performance – that which the examinee is likely to attain if further training can be provided. The assumption is that examinees are likely to utilise real-life learning opportunities in a similar way. *It is important to note that a small improvement score does not imply limited learning potential, because the current and projected future levels of performance (as well as the resulting difference or improvement score) are all relevant in determining overall learning potential.*

In practice, results should preferably be interpreted as the current and potential levels of cognitive reasoning ability, which could be attained if relevant training were to be provided. Thus, if an individual is currently performing at tertiary level (in the pre-test) and maintains that level of performance in the post-test – even with a zero improvement score – then indications are that he or she will be able to cope with and benefit from training up to tertiary level. Differences between levels of current or potential reasoning ability and the level at which training is considered can be interpreted as the amount of effort that will be required from the individual to attain success at the particular training level. The use of dynamic testing, which incorporates a test-train-retest strategy, allows for the measurement of learning potential – a field that is gaining ground in cognitive assessment – because to some extent it allows for more equitable assessment of people coming from different educational and socioeconomic backgrounds and resulting disparities in prior learning opportunities.

**Clarification of the concept of learning potential**
One of the reasons for the more recent development of learning potential as a theoretical concept and continued efforts to measure it is that research results indicate that intelligence quotient (IQ) scores are subject to change (Grigorenko & Sternberg, 1998). Changes in IQ test scores are usually linked to educational opportunity, language proficiency and general socioeconomic level, with differential changes in test scores between cultural groups (Claassen, 1997; Vincent, 1991). Where certain culture groups or other subgroups are disadvantaged, an improvement in the socioeconomic and educational opportunities of the disadvantaged group results in increases in the mean group

score which are beyond the normal population increases over time (Van de Vijver, 1997; Vincent, 1991).

Standard tests of cognitive ability generally measure the products of prior learning and hence rely heavily on the assumption that all examinees have had comparable opportunities to acquire the skills and abilities being measured. This assumption is false when individuals from different socioeconomic and cultural backgrounds are compared. In the measurement of general reasoning ability, persons from poor educational and/or socioeconomic backgrounds are often at a disadvantage when standard cognitive tests are used, because these tests often rely quite heavily on crystallised abilities which are influenced by prior learning experiences (Claassen, 1997).

Whereas ability refers to that which is available on demand, potential is concerned with what could be, and is based upon the possibility of change (Taylor, 1992, 1994a, 1994b; Von Hirschfeld, 1992; Zaaiman, Van der Flier & Thijs, 2001). Learning potential has to do with an overall cognitive capacity and includes both present and projected future performance. Implied in the use of the term is the assumption that intelligence – that which is measured with psychometric tests – is changeable, as indicated by improvement in scores obtained with standard tests when a relevant learning opportunity or some form of help can be provided.

In recent years, Vygotsky's (1978) theory of the ZPD has generally been acknowledged as the theoretical foundation upon which dynamic assessment and the measurement of learning potential has been built. Vygotsky (1978, p. 87 [own italics]) clearly indicated that the ZPD should be used as a tool by means of which

> we can take account not only of the cycles and maturation processes that have already been completed but also those processes that are currently in a state of formation, that are just beginning to mature and develop ... allowing not only for what already has been achieved developmentally but also for what is in the course of maturing.

While Vygotsky included both the initial level of functioning and the ZPD in explaining his theory, the difference score or ZPD has often (incorrectly) been referred to as that which indicates "potential". Because much of the early research in dynamic assessment involved low-ability examinees with similar (low) initial levels of performance, the initial focus was only on the ZPD or difference score obtained. This can only be done in special cases where the initial levels of performance are equal. However, in all other cases, when one needs to interpret the results of individuals where there are differences in the initial level of the performance and quite likely also differences in the ZPD, pre-test and post-test results must be included in the interpretation because, for all such cases, the use of the ZPD (difference) scores without reference to the level at which they occur provides incomplete information. Vygotsky's (1978) proposed use of both the actual developmental level (level of initial performance) and the ZPD is thus essential to achieve logical and practically useful interpretations (De Beer, 2005).

De Beer, (2002c, pp. 98-99) states the following:

> If someone were to say that a university mathematics professor has no learning potential, quite a few eyebrows would be raised. A person who functions at such a high level should by all accounts be able to cope better than most people with virtually any new learning situation. If the focus is on the ability to learn, then credit also needs to be given to learning that has already been accomplished and which forms part of the learner's repertoire. The professor will probably obtain a very high score on the initial (actual)

level of performance and consequently can show only limited improvement. Within the restrictive framework of considering only the difference score as the score that indicates learning potential, it is therefore possible to say that she has very little learning potential. To take the example to the extreme, when selecting someone for further training, this professor could find herself being dropped in favour of a primary school pupil who showed more improvement, since the latter's difference score (ZPD) is larger – and this, in spite of the fact that the overall level of performance of the primary school pupil is substantially below that of the professor. It is clear, especially when one acknowledges that measurement of mental development is used in the framework of learning and training environments, that actual developmental level (pretest performance) cannot be overlooked in dynamic assessment. If it is assumed that by learning potential, we mean the potential to benefit from and cope with new learning situations, it is clear that Vygotsky's interpretation of using both the actual level of development and the ZPD should be adhered to.

In general, individuals with larger ZPD scores are likely to improve their performance, whereas those with smaller ZPD scores are likely to maintain their present level of performance. However, it is clear that the present level of performance as well as improvement shown should be considered to determine appropriate decisions in terms of learning opportunities.

Learning potential for the LPCAT is defined as a combination of the pretest performance and the magnitude of the difference between the pretest and the post-test scores. Since the LPCAT measures learning potential over a broad range of ability levels, it is essential that the improvement score should not be used alone, but that present level of performance as well as the projected future level of performance and the resulting difference score should also be taken into account.

### Time- and cost-effective dynamic assessment administration procedures

CAT is one of the most exciting developments to flow from IRT. It is based on the premise that "an examinee is measured most effectively when the test items are neither too difficult nor too easy for him" (Lord, 1980, p. 150). CAT involves the interactive selection of items during test administration which means that item difficulty is matched to the examinee's (estimated) ability level throughout the test session. The item selected each time is the one that provides the best information at the examinee's current estimated level of ability. The interactive selection of appropriate items from an item bank throughout the test is possible because the difficulty level of items and the examinee's estimated ability level are on the same scale. Furthermore, CAT brings about a significant saving in testing time.

CAT makes possible measurement that is equivalent in precision at different ability levels, since the termination criterion can be linked to the level of accuracy of measurement that has been achieved. Another important factor is that adaptive tests are power tests and not timed tests. In adaptive testing procedures, it is possible to administer varying numbers and different sets of items to individuals while scores remain comparable – since they reflect the level of the underlying trait. This same principle allows direct comparison of pretest and post-test scores of the same examinee as well as comparison of scores of different examinees, making CAT uniquely suitable and appropriate for the measurement of learning potential (Sijtsma, 1993).

Owing to the use of computerised adaptive testing, the LPCAT takes approximately one hour to administer – including the introduction and orientation, pretest, training and post-test –

with the results available immediately after completion of the test. This is quite comparable to the time required for administration of standard cognitive tests. The LPCAT can be administered to groups and the size of the group is determined by the number of computers available for testing. Instructions are either read on the screen – with all explanations, instructions and feedback appearing on the screen – or read out to the examinee. For the latter version of the LPCAT, no instructions appear on the screen and the instructions to be read are available in the User's Manual in all 11 official languages of South Africa (De Beer, 2000a). These instructions can also easily be translated into any other language, which would allow for administration of the LPCAT in any language of choice.

The fact that the testing time is comparable to that of standard tests, that it can be administered to groups of examinees and that the results are available immediately after completion of the test, provides definite time and cost advantages and improves ease of administration.

### Measurement accuracy

Measurement problems in dynamic testing have included subjective scoring of some procedures, problems with measurement accuracy of, in particular, the difference or improvement scores, the lack of standardisation, which limits generalisation and comparison and the practice effect when the same instrument is used in both the pretest and post-test. Many of these factors can be addressed by use of IRT.

The development of IRT in the last 30 to 40 years has introduced significant changes in psychometric theory and test development (Embretson, 1996; Embretson & Reise, 2000). The main advantage of IRT for learning potential measurement lies in the improved accuracy of measurement of difference scores, as well as improved means to compare scores of the same or different examinees. It allows a modern-day solution to ensure both fair and accurate measurement of learning potential. IRT and CAT procedures seem particularly appropriate for learning potential assessment because they improve both measurement accuracy and time efficiency.

A further extremely useful application of IRT is differential item functioning (DIF) analysis to investigate bias (Osterlind, 1983; Wainer, 1993). Separate item characteristic curves can be drawn for different subgroups, thereby allowing for visual representation of item characteristics per subgroup, thus providing for comparison of subgroups and investigation of item bias (De Beer, 2004).

### Empirical research results (psychometric properties)

Although some South African researchers have shown positive results with the use of dynamic assessment methods (Boeyens, 1989; De Beer, 2002, 2003; Shochet, 1992, 1994; Lopes, Roodt & Mauer, 2001), ongoing research is required to provide solid evidence in support of the use of these measures. Grigorenko and Sternberg (1998) stated in particular that, despite the obvious advantages and usefulness offered by dynamic assessment techniques, convincing empirical data are needed to ensure their further and ongoing general use.

The LPCAT is intended to serve as a screening instrument that can be used mainly to counter inadvertent discrimination against disadvantaged groups. By using nonverbal figural patterns to measure reasoning ability, it is not dependent upon either language proficiency or prior school learning. Results of the test-train-retest procedure indicate present level of reasoning performance as well as the projected or potential future level of general reasoning performance after relevant training. The LPCAT was developed as a dynamic computerised adaptive test specifically for South Africa's multicultural context, aimed at addressing the need for a fair, psychometrically sound and time-efficient measure of learning potential in the domain of general nonverbal figural reasoning. It addresses the typical concerns of cross-cultural assessment in terms of the construct measured, methods used and the investigation of item bias (De Beer, 2000a, 2000b, 2004; Van de Vijver, 2002).

The LPCAT uses nonverbal, figural items that can be administered to all culture groups. It focuses on learning potential and assesses not only present level of performance, but also the level to which examinees are able to improve their performance after relevant training. The training provided as part of the administration in the test-train-retest approach, is standard – similar to typical group training situations – thus allowing for comparison between individuals who were given the same standard training.

The LPCAT makes use of CAT to save administration time without forfeiting quality or accuracy of measurement. In the CAT process, items are sampled without replacement from the specified item bank and administered to the examinee until one of the termination criteria is met. The IRT-based measurement allows for more accurate measurement of difference scores.

The results of the LPCAT are in graph form (see Figure 1) indicating performance throughout the pre-test and post-test. In this way it provides continuous information on the level of performance during the test.

Multicultural samples were utilised for item analysis, standardisation and validation of the test to provide information about the psychometric properties and use of the test for multicultural assessment. Coefficient alpha Internal consistency reliability scores ranged between 0.925 and 0.987 for different groups (De Beer, 2000b). Furthermore, the typical performance levels of groups at various educational levels on the final computerised adaptive form of the test were determined for interpretation of results in terms of educational level. These levels are provided in Table 1. Therefore, although individuals may not have the formal educational qualification, their results may show them to be performing at a level of reasoning ability typical of a particular educational level. The results of the LPCAT reflect the current and projected future levels of mental reasoning – comparable to typical levels of education – irrespective of age or attained level of education.

**TABLE 1**
**LPCAT SCORES AND EDUCATIONAL LEVELS**

| LPCAT T-score | Stanine score | ABET/NQF level | Educational level |
|---|---|---|---|
| 20-33 | 1 | ABET level 1 | Grades 0 – 3 |
| 34-37 | 2 | ABET level 2 | Grades 4 – 5 |
| 38-42 | 3 | ABET level 3 | Grades 6 – 7 |
| 43-47 | 4 | ABET 4/NQF 1 | Grades 8 – 9 |
| 48-52 | 5 | NQF level 1-3 | Grades 10 – 12 |
| 53-57 | 6 | NQF level 4-5 | Grades 12+ tertiary |
| 58-62 | 7 | NQF level 6 | First degree |
| 63-68 | 8 | NQF level 7 | Higher degree |
| 69-80 | 9 | NQF level 8 | Advanced postgrad |

The results for samples at different educational levels provide empirical support for the construct and predictive validity of the LPCAT (De Beer, 2003, 2005). A summary of these results are provided in Table 2.
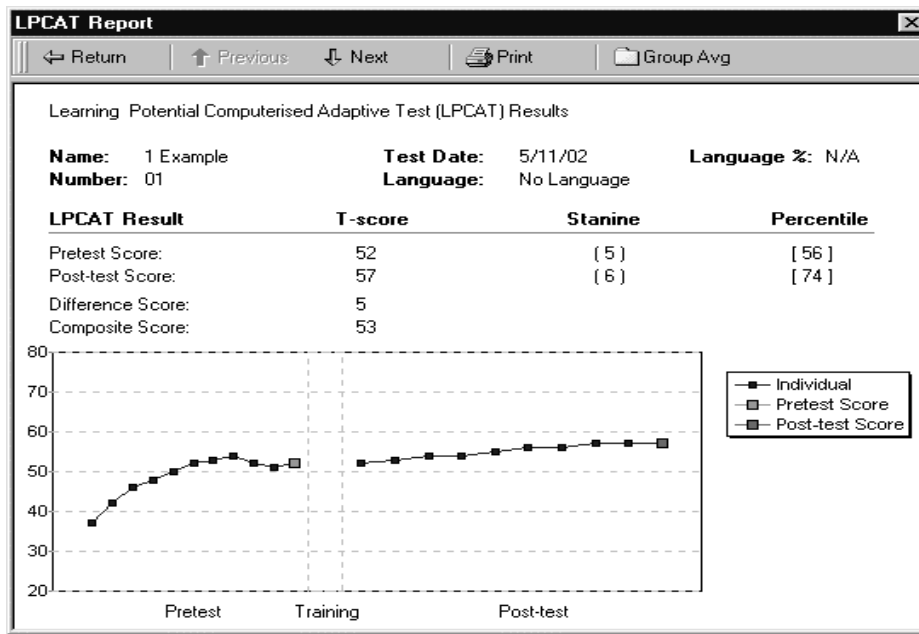
**Figure 1: Example of LPCAT results grap**

| Group/level | Mean post-test scores | Construct validity | Predictive validity |
|---|---|---|---|
| Adult learners (low literacy) | 37,76 (N=194) | 0,408 – 0,645 (Paper-and-pencil games) | 0,437 – 0,610 (ABET training) |
| Primary school (High ability group) | 51,11 (N=56) | 0,353 – 0,474 (Junior aptitude test) | 0,334 – 0,480 (Academic results) |
| Secondary school (Grade 8) | 47,83 (N=128) | 0,613 – 0,691 (General scholastic aptitude test) | 0,530 – 0,600 (Academic results) |
| Bridging group | 57,61 (N=158) | 0,661 (General ability test) | 0,313 – 0,525 (Academic results) |
| First-year students | 59,15 (N=92) | 0,653 – 0,713 (General scholastic aptitude test) | 0,138 – 0,460 (Academic results) |
| Postgrad students | 63,36 (N=75) | 0,100 – 0,376 (Cognitive process profile – various subscales) | 0,008 – 0,166* (Academic results) |

\* Because the academic results were not from same institution, this could have affected the results.
\# (Detailed results in De Beer, 2003)

The results indicate acceptable construct validity with the LPCAT shown to measure a similar general reasoning construct as measured by standard cognitive tests – although it focuses on fluid ability and does not use content that relies on language proficiency or formal previous education. In terms of the prediction of training or academic results, correlations at school and junior tertiary levels are generally acceptable. At postgraduate level, the particular university level group whose results were used, consisted of participants from seven different universities, and the incomparability of academic marks across universities may have affected the results. Furthermore, at that level, restriction of range could also have affected the correlation results. Further research at university level with larger samples from a single institution could provide useful further information.

Although the predictive validity of standard cognitive tests in academic environments is often better than that of nonverbal dynamic tests, the dynamic results are nevertheless useful and more fair for disadvantaged and multicultural groups, because they do not rely on prior learning. As previously indicated, standard cognitive tests often include material that is more closely related to typical academic content, and disadvantaged individuals often perform poorly on these tests. Thus, although standard tests have better predictive validity for academic outcome, they are based on the false premise that all examinees have had similar educational opportunities, and as a result, maintain an unfair disadvantage for those individuals who may not have had optimal educational opportunities.

## DISCUSSION

The LPCAT uses nonverbal figural reasoning content in a test-train-retest format 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). In the multicultural and socioeconomically and educationally diverse South African context, this addresses some of the concerns about the fairness of assessment prescribed by legislation (The Employment Equity Act 55 of 1998).

The results indicate that the LPCAT provides useful information in terms of indicating the level of general reasoning ability and learning potential shown by individuals. It can indicate at what academic level an individual is likely to be able to perform or the amount of effort required from an individual to achieve success at a certain level. There is furthermore adequate variance within the different levels to show that it can indicate different levels of performance for persons at approximately the same academic level.

In terms of construct validity, the results indicate that the LPCAT does measure the general reasoning ability measured by other cognitive tests. With regard to predictive validity for the prediction of (mostly academic) criterion results, acceptable and useful results are shown at most levels, providing support for using the LPCAT for screening and selection.

Using Grigorenko and Sternberg's (1998) four-point system for the evaluation of empirical data available for the LPCAT, the results may be summarised as follows:

1. In terms of the *comparative informativeness (psychometric characteristics, quality and informativeness of the obtained data)*, the results for the LPCAT indicate acceptable psychometric properties in terms of construct and predictive validity. Furthermore, it provides four scores (pretest, post-test, difference and composite scores) as well as a graphic representation of performance in the pre-test and post-test, and allows for a precise evaluation of changes in levels of performance throughout (see Figure 1), providing additional information not available in standard tests. The pre-test reflects current level of performance, while the post-test reflects projected or potential future level of performance. The difference score reflects undeveloped capacity, while the composite score is a reasoned combination or global potential score which takes into account at what level what magnitude of improvement was shown. It should again be emphasised that learning potential is defined as a combination of the present and projected future levels of performance and should not be incorrectly linked only to the difference or improvement score. Investigations on gain scores are considered important (Te Nijenhuis, Van Vianen & Van der Flier, 2006), but has an altogether different focus.

2. An acceptable power of prediction is shown for academic results at various levels (the relationship between the information collected and the criteria used to assess validity).

3. With regard to the degree of efficiency (time and effort invested in consideration of the uniqueness of information obtained – compared with conventional testing), with a typical testing time of approximately one hour and the option to administer it to groups, the LPCAT is fairly comparable to conventional tests in terms of testing time and efficiency.

4. Lastly, the results for groups at different educational levels indicate a *robustness of results (results shown to be replicable across studies and research groups)*.

The overall results provide support for the use of the LPCAT as a screening instrument to assist in decision making for training and development, when combined with other information such as language proficiency, specific aptitude, interests or personality. In particular, it provides useful information for the appropriate level of training for individuals. The advantage of the LPCAT is that performance is not reliant on language proficiency or formal academic qualification, making it a culture-fair measure to include in assessment batteries in the South African context.

The use of dynamic assessment in South Africa has been limited by misperceptions regarding its nature (Murphy, 2006) as well as perceptions regarding cost and practical considerations of its implementation. A measuring instrument like the LPCAT, which makes provision not only for differences between culture groups, but also for ongoing changes within different groups, can provide useful information in the domain of general reasoning ability and future developmental potential for people of different cultures and at different developmental levels. The LPCAT can provide useful information for training and development - hence training can be matched with present and potential future levels of reasoning ability. It thus helps to provide optimal developmental opportunities for individuals over a wide spectrum of ability, while taking into account that prior learning opportunities may have been extremely different. Although this is quite a new field in South Africa, initial results indicate support for the use of dynamic assessment (Murphy, 2006), however, ongoing research is imperative.

## REFERENCES

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

Boeyens, J. (1989). *Learning potential: An empirical investigation. Report PERS-435*, Human Sciences Research Council. Pretoria: Human Sciences Research Council.

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. 82-115). New York: Guilford Press.

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

De Beer, M. (2000a). *Learning Potential Computerised Adaptive Test (LPCAT): User's Manual*. Pretoria: University of South Africa.

De Beer, M. (2000b). *Learning Potential Computerised Adaptive Test (LPCAT)*: Technical Manual. Pretoria: University of South Africa.

De Beer, M. (2000c). *The construction and evaluation of a dynamic computerised adaptive test for the measurement of learning potential*. Unpublished doctoral thesis, University of South Africa.

De Beer, M. (2002). *Utility of Learning Potential Computerised Adaptive Test (LPCAT) scores in predicting academic performance of bridging students: A comparison with other predictors*. Poster presented at the 5th Annual Society for Industrial and Organisational Psychology Congress, Pretoria, 14-17 June 2002.

De Beer, M. (2003). *A comparison of learning potential results at various educational levels*. Paper presented at the 6th annual conference of the Society for Industrial and Organisational Psychology, Sandton Conference Centre, 25-27 June 2003.

De Beer, M. (2004). Use of differential item functioning (DIF) analysis for bias analysis in test construction. *South African Journal of Industrial Psychology*, ??(??), pp. ??-??.

De Beer, M. (2005). *CVDynamic assessment: Practical solutions to some concerns* (?????). Paper presented at the 10th International Conference of the International Association for Cognitive Education and Psychology, Durham, UK: 10-14 June 2005.

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 (1998). *Government Gazette*, 400 (19370). Cape Town, 19 October 1998.

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

Gupta, T.M. & Coxhead, P. (1988). Why assess learning potential? In R.M. Gupta & P. Coxhead (Eds), *Cultural diversity and learning efficiency: Recent developments in assessment* (pp. 1 - 21). Hong Kong: Macmillan.

Guthke, J. (1992). Learning tests: the concept, main research findings, problems and trends. *Learning and Individual Differences*, *4* (2), 137-151.

Guthke, J. (1993a). Current trends in theories and assessment of intelligence. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijssenaars (Eds), *Learning potential assessment: Theoretical, methodological and practical issues* (pp. 13-20). Amsterdam: Swets & Zeitlinger.

Guthke, J. (1993b). Developments 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. 43-68). Amsterdam: Swets & Zeitlinger.

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. 23-42). Amsterdam: Swets & Zeitlinger.

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

Lidz, C.S. (1991). *Practitioner's guide to dynamic assessment.* New York: Guilford Press.

Lopes, A., Roodt, G. & Mauer, R. (2001). The predictive validity of the Apil-B in a financial institution. *Journal of Industrial Psychology, 27* (1), 61-57.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Murphy, R. (2002). *A review of South African research in the field of dynamic assessment.* Unpublished MA thesis. Pretoria: University of Pretoria.

Murphy, R. (2006). A review of South African research in the field of dynamic assessment. *South African Journal of Psychology, 36* (1), 168-191.

Murphy, R. & Maree, D.J.F. (2006). Meta-Analysis of Dynamic Assessment Research in South Africa. *Journal of Cognitive Education and Psychology, 6* (1), 32-60.

Osterlind, S.J. (1983). *Test item bias.* Beverly Hills: Sage.

Reckase, M.D. (1988). *Computerized adaptive testing: a good idea waiting for the right technology.* Paper presented at the meeting of the American Educational Research Association, New Orleans, April 1988.

Shochet, I.M. (1992). A dynamic testing for undergraduate admission: The inverse relationship between modifiability and predictability. In H.C. Haywood & D. Tzuriel (Eds), *Interactive testing* (pp. 332-355). New York: Springer-Verlag.

Shochet, I.M. (1994). The moderator effect of cognitive modifiability on a traditional undergraduate admissions test for disadvantaged black students in South Africa. *South African Journal of Psychology, 24* (4), 208-215.

Sijtsma, K. (1993). Psychometric issues in learning potential assessment. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijssenaars, *Learning Potential Assessment: Theoretical, methodological and practical issues* (pp. 175-194). Amsterdan: Swets & Zeitlinger.

Taylor, T.R. (1992). *Beyond competence: Measuring potential in a cross-cultural situation fairly: Potential in Psychometrics* (part two). Paper presented at the 1992 South African Psychometrics Congress, June 1992.

Taylor, T.R. (1994a). A review of three approaches to cognitive assessment, and a proposed integrated approach based on a unifying theoretical framework. *South African Journal of Psychology, 24* (4), 184-193.

Taylor, T.R. (1994b). *Learning potential: Theory and practical assessment.* Paper presented at the Annual Department of Industrial Psychology Conference on Fairness in Testing and Assessment (30 August 1994). Pretoria: Holiday Inn.

Te Nijenhuis, J., Van Vianen, A.E.M. & Van der Flier, H. (2006). *Score gains on g-loaded tests: No g.* (In print, Intelligence – accepted 11 July 2006).

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 assessement: Value for money? *Applied Psychology: An international Review, 51* (4), 545-566.

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.

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

Von Hirschfeld, S. (1992). *The psychological concept of potential.* Paper presented at the South African Psychometrics Congress, June.

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

Wainer, H. (1993). Model-based standardised measurement of an item's differential impact. In P.W. Holland & H. Wainer (Eds), *Differential item functioning* (pp. 123-135). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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

Zaaiman, H., Van der Flier, H.R. & Thijs, G.D. (2001). Dynamic testing in selection for an educational programme: Assessing South African performance on the Raven Progressive Matrices. *International Journal of Selection and Assessment, 9* (3), 258-269.

## AUTHOR'S NOTE