Graduate
Management
Admission
Council®

Use of CAT in Dynamic Testing

Marié de Beer
University of South Africa

Presented at the CAT-Based Tests Poster Session, June 8, 2007

Abstract

The assessment of learning potential—referred to as *dynamic testing* and generally using a test-train-retest approach—has gained some ground in the field of cognitive assessment. Despite general support for this concept and approach, both measurement and practical concerns have hampered the acceptance of these measures in general assessment contexts. Item response theory (IRT) and computerized adaptive testing (CAT) provide answers to a number of the main concerns in this field, such as long administration time, lack of standardization of administration procedures, and problems concerning measurement accuracy. The Learning Potential Computerised Adaptive Test (LPCAT), is a dynamic learning potential test that has addressed a number of the general concerns regarding dynamic testing through use of IRT and CAT. The development of the LPCAT is discussed. Empirical research results for construct and predictive validity are provided in support of its psychometric properties. Examples are provided of the qualitative interpretations of performance levels from the pre-test and the post-test that are available from the LPCAT.

Acknowledgment

Presentation of this paper at the 2007 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC®.

Copyright © 2007 by the author.

All rights reserved. Permission is granted for non-commercial use.

Citation

De Beer, M. (2007). Use of CAT in dynamic testing. In D. J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing.* Retrieved [date] from www.psych.umn.edu/psylabs/CATCentral/

Author Contact

Marié de Beer, Department of Industrial and Organisational Psychology, University of South Africa, P.O. Box 392, Pretoria, 0003. Email: dbeerm@unisa.ac.za

Use of CAT in Dynamic Testing

Cognitive assessment has always fascinated researchers and has a long and at times controversial history (Fancher, 1985; Jensen, 1981). The first cognitive test of the kind that is still used today was the test developed by Alfred Binet and Theodore Simon in the early 1900s (Binet & Simon, 1905/1916, 1915; Wolf, 1973). Their test—generally known as the Binet-Simon test and later as the Stanford-Binet after it had been standardized for use in the United States—has significance for cognitive assessment in the 21st century for two reasons. First, the Binet-Simon test represents the first test that could be described as a learning potential test. The instruction from the French Ministry of Education to Binet and Simon was to develop a test that could distinguish—among low performing learners in special schools—between those learners who could benefit from "learning exercises" and show that they could benefit from learning to improve on their present levels of performance (therefore enabling placement in normal schools), and those learners who seemed not to benefit from "learning exercises" and therefore probably could be classified as mentally retarded (Wolf, 1973). Allowing for improvement in cognitive performance after a learning experience reflects the current ideas around the measurement of learning potential.

Second, this first example of a cognitive test can also be seen as the first adaptive test, although it took another 70 years or so and the development of computer technology to bring about modern-day computerized adaptive testing (Hambleton & Swaminathan, 1985; Hambleton & Zaal, 1991; Lord, 1980; Reckase, 1989; Van der Linden & Hambleton, 1997; Warm, 1978; Weiss, 1983). Most of the elements of the computerized adaptive tests (CATs) of today were present in the Binet-Simon test, namely: variable entry level (depending on the estimated level of the individual to be assessed); scoring of each question as it was answered; a decision on the level of question to administer next based on the response pattern up to that point in time; and variable termination – again depending on the answer pattern shown by the individual. Similar to modern-day CATs, individuals did not receive the same questions or even the same number of questions, but the scores obtained could be compared.

Dynamic assessment generally refers to testing procedures that include a learning experience as part of the assessment with the aim of measuring learning potential (Campione, 1989; Hamers & Resing, 1993; Laughon, 1990; Lidz, 1987a, 1987b, 1991; Murphy, 2002) so as 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 current levels of performance when relevant learning opportunities could be provided. Vygotsky's (1978) theory of the Zone of Proximal Development (ZPD) is generally acknowledged as the theoretical foundation upon which dynamic assessment and the measurement of learning potential have been built (Brown & Ferrara, 1985; Brown & French, 1979). Vygotsky's theoretical concept of the ZPD refers to the distance between the actual level of performance (i.e., present performance level without help) and the potential level of performance (i.e., performance level after some form of help or learning has been provided).

The typical test-train-retest approach of dynamic assessment is based on Vygotsky's theory of the ZPD and provides information not only about the present level of an ability, but also about the potential level that could be attained if relevant learning opportunities could be provided. This approach takes into consideration that not all individuals have had similar or optimal prior learning opportunities and therefore might not yet have attained their optimal levels of performance. The pre-test reflects the present (actual) level of performance—similar to that which is typically assessed in standard tests. The training that is included is aimed at providing further examples, hints, strategies and guidelines that will highlight important aspects of information required to help solve similar questions to those administered in the test. After the training, the post-test then provides an indication of the potential future level of performance that could be attained if relevant training or learning opportunities could be provided. This entire process (introduction, pre-test, training and post-test) forms part of a single test administration session of approximately one hour. The assumption is that examinees are likely to utilize real-life learning opportunities in a similar way.

It is important to note that a small improvement score does not necessarily imply limited learning potential. The current and projected future levels of performance (as well as the resulting difference or improvement score) are all important in determining overall learning potential. For example, someone who is already currently performing at a postgraduate level in the pre-test might not show any improvement but nevertheless maintains a postgraduate level of performance in the post-test. This definitely does not mean that this individual has no learning potential. The interpretation of such a result would be that this individual shows that he/she should be able to cope with and benefit from training provided up to a postgraduate level—therefore indicating a high level of learning potential.

Individuals from poor educational and socioeconomic backgrounds are often at a disadvantage when standard cognitive tests are used, because these tests often include content that reflects language proficiency or educational background—therefore representing crystallized abilities which are influenced by prior learning experiences (Claassen, 1997; Foxcroft, 1997; Van de Vijver, 1997, 2002). The measurement of learning potential allows for fairer assessment of disadvantaged individuals. The reason for this is twofold, namely that assessment includes a learning experience and that the content used reflects fluid ability which is less influenced by prior learning experiences.

Dynamic testing and the measurement of learning potential have received considerable attention (Carlson, 1989; Grigorenko & Sternberg, 1998; Murphy, 2002). However, despite the initial promise of this field, it is not yet fully utilized in general testing contexts. Some of the reasons why these tests have not really been incorporated have to do with the time that it takes to administer (compared to standard tests), problems with standardization impacting on measurement accuracy, and the lack of research and psychometric information available. Researchers have lamented the fact that limited empirical research is hampering its progress (Grigorenko & Sternberg, 1998; Gupta & Coxhead, 1988; Guthke, 1992, 1993a, 1993b; Guthke & Stein, 1996; Murphy, 2002).

Implied in the use of the term learning potential is the assumption that intelligence as measured by standardized tests is changeable, as indicated by improvement in scores obtained with standard tests when a relevant learning opportunity or some form of help is provided. One of the reasons why learning potential measures have received increased attention is that research results indicate that intelligence quotient (IQ) scores are linked to educational opportunity and socioeconomic level and subject to change (Claassen, 1997; Grigorenko & Sternberg, 1998; Vincent, 1991). Improvement in the socioeconomic and educational opportunities of the disadvantaged group typically result in increases in the mean group score which exceed the normal population increases over time (Van de Vijver, 1997; Vincent, 1991).

South African researchers have contributed both in the development of instruments for the measurement of learning potential (De Beer, 2000a, 2000b; Taylor, 1994) and also in research contributing to the available information on the validity of dynamic assessment measures (Boeyens, 1989; De Beer, 2006; Lopes, Roodt & Mauer, 2001; Shochet, 1992, 1994; Taylor, 1994; Van Eeden, De Beer, & Coetzee, 2001). Murphy (2002, 2007) provided an extensive overview of South African research on dynamic assessment.

The Learning Potential Computerised Adaptive Test (LPCAT) was developed in South Africa and addresses some of the concerns that have been noted regarding dynamic assessment. It uses non-verbal figural reasoning content in 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).

The LPCAT

Purpose

Item response theory (IRT) and CAT address a number of the concerns regarding dynamic testing (Sijtsma, 1993a, 1993b). Having made significant strides in the last 30 to 40 years, IRT and CAT have 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. “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). It therefore allows a modern-day solution to ensure fair, accurate and effective measurement of learning potential. It also allows for the evaluation of differential item functioning (DIF) to investigate bias (Osterlind, 1983; Wainer, 1993), which is of particular importance in the multicultural South African context (Employment Equity Act, 1998). The shortened testing time through use of CAT represents a further advantage.

IRT and CAT procedures seem particularly appropriate for learning potential assessment, because they improve both measurement accuracy and time-efficiency. In addition to the aforementioned advantages, CAT also allows for additional qualitative information to be made available by plotting performance levels of examinees throughout the assessment process.

The purpose of the LPCAT is to provide a dynamic (test-train-retest) measuring instrument for the measurement of learning potential by including a learning opportunity during assessment. Furthermore, use of IRT allows for improved measurement accuracy and comparison of the pre-test and post-test scores (something that has been considered a problem in dynamic assessment using standard tests). Lastly, the separate yet linked CAT pre-test and post-test incorporate all the advantages related to CAT, namely 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-test and the post-test.

Development

Items. The LPCAT was developed as a dynamic test for the measurement of learning potential by means of non-verbal figural reasoning. The test comprises two linked CATs with a standardized training or learning session between the two adaptive tests and takes about an hour to administer. Although the two adaptive tests—a pre-test and a post-test respectively—use separate item banks, they are nevertheless linked in that the exit level of performance (θ estimate) in the first test or pre-test is used as the entry level in the second test or post-test. This further improves measurement accuracy of overall performance in the non-verbal figural reasoning domain.

In the development of the LPCAT three different types of non-verbal figural item formats were used—figure series, figure analogies and pattern completion. These item formats were chosen since they measure fluid general reasoning ability and do not rely on language proficiency or scholastic background. These items have been shown to be more culture-fair (Claassen, 1997; Hugo & Claassen, 1991; Owen, 1998) than items with verbal content.

Initially 270 new questions were developed—90 of each type—and these were administered to a representative sample of 2,450 examinees representing most of the South African culture and language groups (De Beer, 2000b). After using IRT and classical test theory item analysis to evaluate the psychometric properties of the items and to investigate possible bias in terms of gender, educational level, language and culture, 77 items were discarded due to either not meeting the criteria set in terms of psychometric properties or exceeding the cutoff set in terms of bias (De Beer, 2000b, 2004).

The remaining items of each item type (65 figure series, 58 figure analogy and 65 pattern completion items) were arranged in ascending order of item difficulty. The items then were allocated to the pre-test and the post-test sequentially in a 1:2 ratio (one item to the pre-test and the next two to the post-test). This was done separately for each of the three item types to ensure an even spread of item types and item difficulties in the pre-test and post-test. Approximately one-third of the selected items were thus allocated to the pre-test (N = 63) and the remainder to the post-test (N = 125).

McBride (1997) suggested that the number of items in the bank should exceed by a ratio of 5 or 10 to 1, the number of questions an individual examinee will encounter. For the LPCAT, the number of items in the respective item banks exceed (by a ratio of between 5 and 8 for the pre-test and by a ratio of between 7 and 10 for the post-test), the number of questions an individual examinee will encounter. Fewer items are administered in the pre-test (between 8 and 12) than in the post-test (between 10 and 18). The pre-test provides an initial general level of performance. In the post-test, the pre-test level of performance is used as entry level, and therefore more efficient measurement of performance is possible. This requires more items at each difficulty level in the post-test.

Adaptive test termination depends on stopping rules, which include both the minimum and maximum number of items as well as accuracy of measurement. For the LPCAT pre-test, the variance used for termination – that is, the accuracy - of the θ estimate was set at 0.10, while for the post-test it was set at 0.05 (Assessment Systems Corporation, 1995; De Beer, 2000b). Although there is no overall time limit on the test, for practical purposes a three-minute time limit per test questions in the adaptive test parts was set. If a question remains unanswered when the time limit is reached, the question will be replaced by an easier next question as the previous question will have been indicated as unanswered (Assessment Systems Corporation, 1995).

The sample that was used to determine the item parameters, represented a mid-secondary educational level (a mixed group of grade 9 and grade 11 learners was used) (De Beer, 2000b).

The θ estimates for the pre-test and the post-test are converted to a T-score with a mean of 50 and standard deviation of 10. A score of 50 was therefore taken to be equivalent to a mid-secondary level of general reasoning ability. In order to determine the typical performance levels of individuals at other educational levels, sample groups representing a variety of educational levels were used to determine further benchmark scores at different educational levels as shown in Table 1.

Table 1. LPCAT Scores and Educational Levels

LPCAT T-Score	Stanine Score	ABET* / NQF** Level	Educational Level
27-33	1	ABET level 1	Grade 0 – 3
34-37	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
53-57	6	NQF level 4-5	Grade 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.

* Adult Basic Education and Training (ABET)
** National Qualifications Framework (NQF)

Scores. The scores that are provided by the LPCAT are a pre-test score, a post-test score, a difference score (numerical difference between the post-test score and the pre-test score) and a composite score. The composite score is calculated by adding a proportion (actual improvement divided by maximum possible improvement from that pre-test level) of the difference score to the pre-test score. This allows for the fact that it is more difficult to show improvement when the initial level is high compared to a lower initial level of performance. (De Beer, 2000b). It has been found that the post-test score (which allows for maximal credit for the learning that has been achieved as reflected in the post-test score) and the composite score (which allows for partial credit for the learning that has been achieved) generally show higher correlations with criterion measures compared to the pre-test score (which indicates current level of performance). This provides support for the concept of measurement of learning potential, as it shows 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).

The benchmark performance levels and typical score ranges of groups at different educational levels (see Table 1) are used for the interpretation of the LPCAT results. It is always recommended that the LPCAT results should be used together with other test results, since it only measures non-verbal figural reasoning, and learning potential using this type of content, and therefore does not provide information on other aptitudes and abilities, interests, or individual personality preferences. Furthermore, results in terms of the level of performance should be interpreted by comparing it with a given educational level (for training, for instance) and then interpreted in terms of whether the LPCAT levels are commensurate with the intended level of training to be provided. “Differences between levels of current or potential reasoning ability and the level at which training is considered could be interpreted as the amount of effort that will be required from the individual to attain success at the particular training level” (De Beer, 2006, p. 9).

Because language proficiency and/or other specific skills and aptitudes might be particularly important in specific screening and selection decisions, these need to be determined by means of other measuring instruments. However, the LPCAT is particularly useful in providing information on the general reasoning level at which the individual is currently performing as well as the level to which the individual could be developed—provided that relevant training and learning opportunities could be provided. Since language proficiency is important in any tertiary learning environment, it would need to be assessed and possibly improved before someone from a disadvantaged (educational) background could fully benefit from training at the levels indicated in the LPCAT—since the latter is based only on non-verbal figural reasoning.

Administration. The administration time of the LPCAT is approximately one hour. This includes the introduction and orientation, pre-test, training, and post-test, and results are available immediately after completion of the test. This time is quite comparable to the typical time required for administration of standard cognitive tests. Group administration of the LPCAT is possible, although the size of the group will be determined by practical considerations such as the number of computers available for testing and other logistical considerations. Individuals may either read the instructions, feedback and explanations from the screen (in English), or the instructions can be read to the individual from the user’s manual (De Beer, 2000a). For the latter version of the LPCAT, no instructions appear on the screen and the instructions that are read to the individuals are available in the User’s Manual in all 11 official languages of South Africa (De Beer, 2000a). It would be quite simple to translate these instructions to any other language. For ease of use, the individual needs to use only the space bar and the enter key in the testing process and therefore participants are not required to be computer literate—a definite advantage when testing low literacy or disadvantaged groups.

Psychometric Properties

Research on the LPCAT has indicated satisfactory psychometric properties. It also complies with the Employment Equity requirements in South Africa, which requires all psychological tests to be scientifically developed, shown to be valid and reliable, and biased and fair to all groups (Claassen, 1997; Employment Equity Act, 1998; Foxcroft, 1997).

Reliability/precision. Both classical test theory methods and IRT-based methods were used to evaluate measurement precision. In terms of classical test theory, coefficient alpha was high—ranging between 0.925 and 0.981 (De Beer, 2000b), indicating high internal consistency and homogeneity of the items used. In IRT and adaptive testing, (im)precision is measured by the inverse of the item information available at a given ability level. The expected information for the LPCAT pre-test is 17.811 and for the post-test 32.699 (De Beer, 2000b). Since the standard error is equal to the reciprocal of the square root of test information, these values for the pre-test and post-test translate into expected (average) standard errors of 0.24 and 0.17 theta units respectively. This means that generally speaking, roughly 68 percent of the estimates will fall between -0.24 and +0.24 (for the pre-test) and between -0.17 and +0.17 (for the post-test) from the estimated ability level in standard θ units. Translated to T-scores, this means that roughly 68 percent of the T-score estimates will fall between -2.4 and +2.4 T-scores from the estimated ability level (for the pre-test) and for the post-test roughly 68 percent of the T-score estimates will fall between -1.7 and +1.7 T-scores from the estimated post-test ability level. The estimated standard error values at various θ values are provided in the LPCAT Technical Manual (De Beer, 2000b).

Validity. For validity evaluation the following aspects and measures of validity were included: content validity, construct validity (convergent and discriminant) and criterion-related validity (concurrent and predictive).

Content validity was evaluated by a panel of experts of the Human Sciences Research Council. The non-verbal figural item content was deemed appropriate for a measuring instrument of learning potential and the item types represented the typical formats found in culture-fair measures (Claassen, 1997; Owen, 1998). Construct validity was evaluated by comparing the LPCAT results with a variety of other standard cognitive tests for groups at different educational levels. Correlations typically ranged between 0.4 and 0.7 (De Beer, 2000b, 2006). Concurrent and predictive validity was evaluated primarily by the use of academic performance as a criterion. Correlations typically ranged between 0.1 and 0.5 (averaging around 0.3 to 0.4) and were generally statistically significant. A pattern found is that post-test and composite scores typically showed higher correlations with the academic criterion results than the pre-test scores (De Beer, 2000b; Van der Merwe & De Beer, 2006)

Qualitative Interpretation of LPCAT Results

Over and above the scientific, statistical and psychometric ways of evaluating LPCAT results and interpreting its utility, there is also a qualitative aspect to the interpretation of LPCAT results. In terms of the scores that are provided, the pre-test score represents the performance reached at the end of the pre-test, while the post-test score represents the performance reached at the end of the post-test (see Figure 1). When examining the graph of the pre-test and post-test, it is possible that higher levels of performance were reached during either the pre-test or the post-test, and that these higher levels of performance are not reflected in a total score.

An individual’s LPCAT graph shows the estimated ability or performance level of the individual throughout the adaptive pre-test and post-test. These graphs (see Figure 1 for examples) can be interpreted as follows:

1. When the pre-test and post-test graphs are reasonably flat—that is, performance generally remains at the same level—the future performance of the individual could be expected to remain at a level similar to the current level of performance.
2. When there is a definite positive gradient in the graphs, the interpretation is that future levels of performance could be expected to be higher than the current level of performance—provided that relevant training and learning opportunities could be provided.
3. A graph that shows a negative gradient—which is contrary to the expected flat graph or graph with a positive gradient—could indicate possible loss of concentration or motivation, or in extreme cases a negative attitude toward assessment. In some or other way, the individual is not maintaining levels of performance attained earlier.

4. In cases where the individual attains a higher level of performance during the test but this level of performance is not maintained, credit could nevertheless be given for the peak levels reached, although this might not be reflected in the end of the pre-test or end of the post-test scores provided.

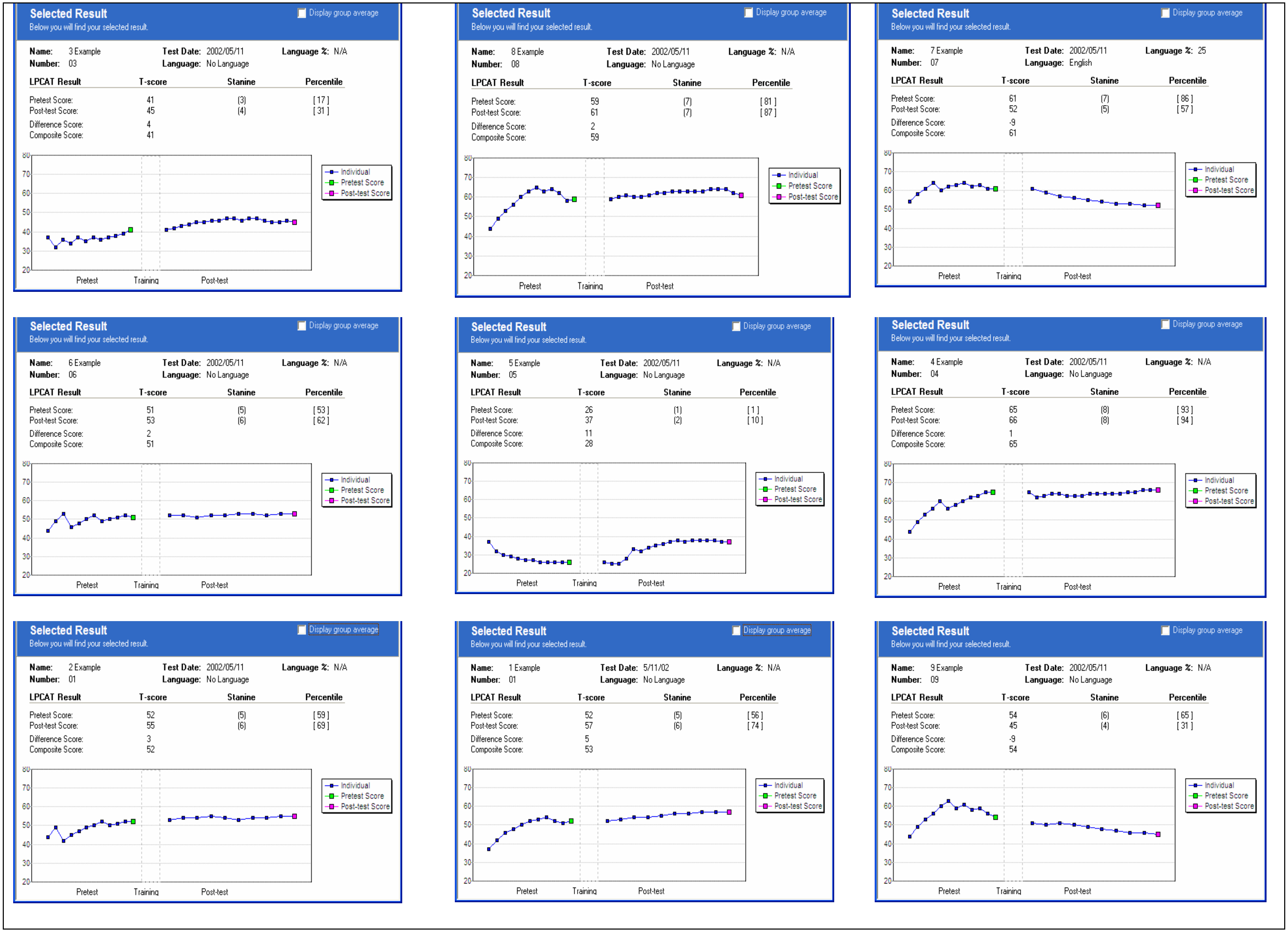
Conclusions

IRT and CAT have made possible improved psychometric quality and practical utility of a dynamic test-teach-retest instrument for the measurement of learning potential based on non-verbal figural content. A learning potential test like the LPCAT makes provision not only for differences between individuals and groups, but also for ongoing changes within different individuals and groups. It provides useful information in the domain of general reasoning ability and future developmental potential for individuals and, since it uses only non-verbal figural content, it could provide useful information for various culture, language, and educational groups. The LPCAT provides useful information for training and development, as training could be matched with present and potential future levels of reasoning ability.

The visual way in which adaptive test results are be presented allows for qualitative interpretation in terms of the level of performance throughout the test administration. In addition to providing useful predictive validity results in terms of academic training, it also provides qualitative information in terms of the individuals level of performance throughout the pre-test and the post-test.

In particular in the South African multicultural context (Claassen, 1997; Foxcroft, 1997; Meiring, 2007; Shuttleworth-Jordan, 1996) where there are still large differences between groups in terms of educational and socioeconomic factors, measures of learning potential provide a more equitable approach to assist in decision making regarding training and development opportunities.

Figure 1. Examples of LPCAT Results



Selected Result □ Display group average
Below you will find your selected result.

Name: 3 Example **Test Date:** 2002/05/11 **Language %:** N/A
Number: 03 **Language:** No Language

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	41	(3)	[17]
Post-test Score:	45	(4)	[31]
Difference Score:	4		
Composite Score:	41		

Selected Result □ Display group average
Below you will find your selected result.

Name: 8 Example **Test Date:** 2002/05/11 **Language %:** N/A
Number: 08 **Language:** No Language

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	59	(7)	[81]
Post-test Score:	61	(7)	[87]
Difference Score:	2		
Composite Score:	59		

Selected Result □ Display group average
Below you will find your selected result.

Name: 7 Example **Test Date:** 2002/05/11 **Language %:** 25
Number: 07 **Language:** English

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	61	(7)	[86]
Post-test Score:	52	(5)	[57]
Difference Score:	-9		
Composite Score:	61		

Selected Result □ Display group average
Below you will find your selected result.

Name: 6 Example **Test Date:** 2002/05/11 **Language %:** N/A
Number: 06 **Language:** No Language

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	51	(5)	[53]
Post-test Score:	53	(6)	[62]
Difference Score:	2		
Composite Score:	51		

Selected Result □ Display group average
Below you will find your selected result.

Name: 5 Example **Test Date:** 2002/05/11 **Language %:** N/A
Number: 05 **Language:** No Language

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	26	(1)	[1]
Post-test Score:	37	(2)	[10]
Difference Score:	11		
Composite Score:	28		

Selected Result □ Display group average
Below you will find your selected result.

Name: 4 Example **Test Date:** 2002/05/11 **Language %:** N/A
Number: 04 **Language:** No Language

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	65	(8)	[93]
Post-test Score:	66	(8)	[94]
Difference Score:	1		
Composite Score:	65		

Selected Result □ Display group average
Below you will find your selected result.

Name: 2 Example **Test Date:** 2002/05/11 **Language %:** N/A
Number: 01 **Language:** No Language

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	52	(5)	[59]
Post-test Score:	55	(6)	[69]
Difference Score:	3		
Composite Score:	52		

Selected Result □ Display group average
Below you will find your selected result.

Name: 1 Example **Test Date:** 5/11/02 **Language %:** N/A
Number: 01 **Language:** No Language

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	52	(5)	[56]
Post-test Score:	57	(6)	[74]
Difference Score:	5		
Composite Score:	53		

Selected Result □ Display group average
Below you will find your selected result.

Name: 9 Example **Test Date:** 2002/05/11 **Language %:** N/A
Number: 09 **Language:** No Language

LPCAT Result	T-score	Stanine	Percentile
Pretest Score:	54	(6)	[65]
Post-test Score:	45	(4)	[31]
Difference Score:	-9		
Composite Score:	54		

References

Assessment Systems Corporation. (1995). *User's manual for the MicroCAT testing system (Version 3.5).* St. Paul: Assessment Systems Corporation.

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

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

Boeyens, J. C. A. (1989). *Learning potential and academic performance*. Unpublished Master's dissertation, University of South Africa.

Brown, A. L., & Ferrara, R. A. (1985). Diagnosing zones of proximal development. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskyan perspectives* (pp. 273-305). Cambridge: Cambridge University Press.

Brown, A .L., & French, L. A. (1979). The zone of potential development: Implications for intelligence testing in the year 2000. *Intelligence, 3*, 255-273.

Campione, J. C. (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities, 22*, 151-165.

Carlson, J. S. (1989). Advances in research on intelligence: The dynamic assessment approach. *The Mental Retardation and Learning Disability Bulletin, 17(1)*, 1-20.

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: Production printers (UNISA)

De Beer, M. (2000b). *Learning Potential Computerised Adaptive Test (LPCAT): Technical Manual*. Pretoria: Production printers (UNISA)

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. (2006). Dynamic assessment: Practical solutions to some concerns. *South African Journal of Industrial Psychology,* 32(4),1-7.

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, No 55 (1998). *Government Gazette, 400 (19370). Cape Town, 19 October 1998.*

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

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

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 Press.

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.

Guthke, J., & Stein, H. (1996). Are learning tests the better version of intelligence tests? *European Journal of Psychological Assessment, 12(1),* 1-13.

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.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhof Publishing.

Hambleton, R. K., & Zaal, J. N. (Eds.). (1991). *Advances in educational and psychological testing: Theory and applications*. Boston: Kluwer Academic Publishers.

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: Human Sciences Research Council.

Jensen, A. R. (1981). *Straight talk about mental tests*. London: Methuen.

Laughon, P. (1990). The dynamic assessment of intelligence: A review of three approaches. *School Psychology Review, 19(4),* 459-470.

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.

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

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.

Meiring, D. (2007). *Bias and equivalence of psychological measures in South Africa.* Ridderkerk: Labyrint Publication.

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

Murphy, R. (2007.). *Exploring a meta-theoretical framework for dynamic assessment and intelligence.* Unpublished Doctoral Thesis. University of Pretoria.

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

Owen, K. (1998). *The role of psychological tests in education in South Africa: Issues, controversies and benefits*. Pretoria: Human Sciences Research Council.

Reckase, M. D. (1989). Adaptive testing: The evolution of a good idea. *Educational Measurement: Issues and Practice, 8(3),* 11-15.

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.

Shuttleworth-Jordan, A. B. (1996). On not reinventing the wheel: a clinical perspective on culturally relevant test usage in South Africa. *South African Journal of Psychology, 26(2),* 96-102.

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, *Learning Potential Assessment: Theoretical, methodological and practical issues* (pp. 117-134). Amsterdan: Swets & Zeitlinger.

Sijtsma, K. (1993b). 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. (1994). 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.

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., & Hambleton, R. K. (Eds). (1997). *Handbook of modern item response theory*. New York: Springer.

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.

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.

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.

Warm, T. A. (1978). *A primer of item response theory (Technical Report 941078)*. Oklahoma City, OK: U. S. Coast Guard Institute.

Weiss, D. J. (1980). Final Report: Computerized 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 computerized adaptive testing*. New York: Academic Press.

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