TABLE OF CONTENTS

List of Tables

CHAPTER 1 Introduction and background to the study

"The guiding principle of dynamic assessment is that if you wish to understand how a child learns, it is best to engage the child in the learning process."

- Carol Lidz

"Broadly defined, dynamic assessment is naturally linked with intervention. In essence the goal of dynamic assessment is to evaluate, to intervene, and to change. The goal of dynamic testing, however, is much more modest: it is to see whether and how the subject will change if an opportunity is provided."

- Grigorenko and Sternberg

1.1 Background to the research

The impetus for this study has been the need to investigate learning potential assessment techniques currently available, looking at learners in the education environment. Psychometric ability tests assess the ability, whether innate or acquired, to accomplish certain tasks necessitated by various fields and disciplines. The question is whether psychometric ability tests assess the ability to learn certain skills. Learners who perform poorly in conventional IQ tests may prove to be good performers on dynamic assessments. That is, they have the ability to benefit from tuition based on their ability to learn.

One important question that arises is whether static tests assess learning potential. To date, educational institutions and the private sector have relied mostly on unstandardised tests, which indicate proficiencies or lack thereof in certain skills or abilities. They do not encompass the added notion of being able to test for the ability or the potential to learn the specific task at hand. This is by and large lacking in industry and schools where many employees and students are sent out to enhance their capabilities by completing courses pertinent to their job requirements or intended direction of study. As to whether these courses are indeed advantageous in the sense of people having acquired any new information and/or skills and being able to retain the information is questionable. If educators continue to assess learners based on static assessments such as conventional IQ tests, and from this infer who will receive special treatment (and in so doing stigmatise learners) a potential learner is put at a distinct disadvantage, being unable to fulfil his/her abilities. Learning potential theory states that potential may reside in some learners, a potential which may remain untapped if the learner is assessed in a conventional manner.

Another question that needs to be addressed is which candidates are suitable to enter certain fields. This leads to the necessity to determine whether or not effective placement assessment can be done (particularly at tertiary institutions) without allowing for potential development. Candidates may perform poorly in ability tests but this in no way suggests an inability to learn that particular skill, seeing that most of these conventional test batteries rely almost exclusively on prior performance. Psychometric testing has developed from the conventional (structural) approach in the early twentieth century, where static measurements were commonplace through to the information processing approach developed in the 1960s with the advent of the personal computer (Taylor, 1994). However, neither approach takes cognizance of the potential which resides in individuals to accomplish tasks which may not have been attempted had potential for these tasks been identified. Another aspect of this developmental or dynamic approach is the suitability of tests to overcome cultural, educational and environmental bias, evidenced in many static conventional tests.

Due to previous disadvantages which many students experienced and still do in many cases today, learning potential tests appear to offer the test-taker an opportunity not afforded by conventional test models. Learning potential tests may be a better alternative or at least serve as a complementary assessment method to test such individuals in South Africa. This study examines the dynamic tests that are currently available in South Africa, or the dynamic procedures that are currently being used and attempts to assess how they compare with conventional tests. It is important to investigate the

ways in which testing agencies and placement bureaux at tertiary institutions have adapted conventional tests. This study attempts to answer these pertinent questions so that heightened awareness may bring about a change in the current thinking of ability testing as well as to overcome inhibitory factors encountered thus far in static tests.

As the inequalities of the past have been and are being assessed, the emergent democracy has institutionalised equal opportunities for all learners. Yet, due to a multitude of sometimes uncontrollable factors, equal educational opportunities have not always been realised. There exists a niche for the use of dynamic assessment on a larger scale in South Africa. This fact has been realised by many local researchers (Andrews, 1996; Boeyens, 1989a, 1989b; De Beer, 2000; De Villiers, 1999; Henley, 1989; Lipson, 1992; Shochet 1986; Van Aswegen, 1997; Zolezzi, 1992) but to more effectively understand and deal with the situation and in the attempt to develop such tests, a greater understanding of what research has gone before and the conclusions reached have first to be attended to.

The idea and motivation behind this dissertation research was initially instigated by an awareness of the lack of meaningful ways of analysing students' academic performances based on prior test results. Even though aptitude tests may reveal that students should follow certain career paths, the main question is whether or not these so-called static tests can really predict (in any way) cognitive changes that might occur over the crucial period of the start of academic studies at a tertiary institution. The notion of predictability does not seem, at first, to be compatible with the ideas propounded by dynamic assessment, for predictability assumes an already known course of action which may or may not be fulfilled. However, dynamic assessment reveals that, as a basic premise such "known paths" are not knowable and that no test can adequately predict what is to become of any individual. Though this may be the case, an attempt can still be made to allow for at least an "unknown" factor encompassed within static tests.

1.2 Problem statement

Dynamic assessment is a field which, although researched for a number of years overseas, particularly in Europe,[1] has yet to find its place in broad South African research circles. Apart from some professional and academic institutions which have tried to implement dynamic assessment as part of selection and placement initiatives, and fewer organisations which have stressed the remedial aspect of dynamic assessment, it would seem that most research emanates from graduate studies at a select few universities, valuable information that when pieced together can provide enough ammunition to further bolster the necessity to make use of dynamic assessment (in its varied contexts). Researchers who are currently working in this field (or related fields who share an interest in what this field has to offer) are not always aware of the number of studies that have been conducted, an area which this study seeks to remedy. The problem addressed is what is currently being done and has been done in the last two decades in South African research in the field of dynamic assessment.

1.3 The aim of the study

The main aim in this study is to provide information to those researchers, practitioners and other interested parties concerned with research that has already been conducted in South Africa in the field of dynamic assessment, on future implications and directions that the field has taken and is likely to take. It strives to serve as an information tool for the above-mentioned, combining as much of the research available as possible on the topic in South Africa, and in so doing, also serve as a repository of summations of various research projects. It offers summaries of results, main findings and issues in each project thus allowing the reader to gain some sense of the "bigger" picture of the research currently available locally. It is in essence a large literature study; which has to be summarised in

1 Engelbrecht (1999) states that more research into dynamic assessment seems to have emanated from Europe than the United States.

order to deliver a coherent and comprehensive account of the goings on in this field today. To date, as far as can be ascertained, no such literature summary is available.

Dynamic assessment is the next "movement" in the history of assessment and yet, with its varied inhibitory issues, will at least cement the foundation of an even more progressive theory of assessment. It is at least a start, a start which seems to be finding tentative holds in some areas but which needs to become established holds eventually. In essence, the aim of the study is to provide practitioners, researchers and any other interested stakeholders, with information that can aid in designing, developing and implementing of dynamic assessment in as varied a context as possible, beginning within the realm of primary, secondary as well as tertiary education.

People change throughout life, cognitively, physically and psychologically. This is not to say that the overall design of cognition, physicality and psychological make-up are not to a certain extent wired into human beings. The plasticity of humans in all these areas does not negate the pre-wired nature of our being. However, it can be argued that static tests do not allow for this notion of change to be assessed. In order to address this issue in South Africa and before a proper research design can be devised, the nature of dynamic assessment in this country needs to be addressed.

The results and figures obtained in local studies, the conclusions reached and the recommendations made for future research can only benefit those who may in the future want to enter into the field. Instead of re-inventing the dynamic assessment wheel, researchers can turn towards a source which will offer succinct information, objective comment and pertinent findings of previous results from such studies. The study aims to go one step further than the conventional literature survey which usually divulges a brief and limited source of information pertinent to the particular study under investigation. This survey addresses broad findings within the field.

1.4 Format of the presentation of the study

The impetus behind the research has been dealt with in this first Chapter and the reasons as to why it is deemed necessary have also received attention.

Chapter Two focuses on dynamic assessment as representing the next phase in test theory and practice. A review of dynamic assessment's various theoretical underpinnings, origins as well as practices receives attention, broadly defining the main groupings and orientations within this field. Much of the dynamic assessment theory has been developed overseas.

Chapter Three centres on the methodology used in the study, the means by which information was obtained as well as the various databases that were accessed and additional information was come by in order to bring together the various results.

Chapter Four forms the crux of the dissertation, as well as the bulk, detailing the findings, methods and designs used in each study, making comparisons throughout with other studies, and continues by integrating theoretical constructs, conclusions as well as recommendations suggested for further studies in the field. No study is favoured as this survey merely aims to elucidate the findings of each study, and any comment is strictly of an academic nature.

Chapter Five discusses the results of the survey and compares the findings across the board with Chapter Six serving as the concluding chapter which also imparts final recommendations for future studies in this area.

Chapter Two now follows with a detailed overview of the theory behind dynamic assessment

CHAPTER 2 Dynamic assessment - the next phase in test theory and practice - theoretical underpinnings

"If the misery of the poor be caused not by the laws of nature, but by our institutions, great is our sin."

- Charles Darwin

2.1 Introduction

The following section investigates the development of dynamic assessment, the nature of the course it has taken since the early part of the twentieth century and its often estranged and tenuous link with classical test theory, which is not, as it is commonly understood, its predecessor. The early notions and ideas of dynamic assessment have been with the intelligence research community since the inception of modern test theory and have, in fact, progressed in parallel. The main contributors and their theories and main assumptions in the field dynamic assessment receive attention as does contentious issues such as important classical test theory concepts - reliability, validity and scoring.[1]

2.2 Intelligence - the elusive multifaceted diamond

Definitions of intelligence have been the bane of many psychologists' research endeavours and are not without a contingency of researchers who maintain that such an elusive concept will never be adequately defined. Numerous intelligence researchers have sought to define intelligence, and bar the more popular understandings of what it supposedly refers to, there has never been, there is not and will probably never be consensus as to what intelligence refers to. There are, of course, numerous definitions of intelligence (Sternberg, 1997). This does not imply a fatalistic stance on the topic where future attempts to define it will be relegated to the heap of other definitions. Perhaps the progress of science will have it such that the definitions become more closely aligned. Without exception, treatize, theses, articles, papers, books and other accredited literature both locally and internationally explicate the meanings and definitions of intelligence as understood throughout the ages. It is not the aim of this study to try and add further defining criteria to this already plagued field.[2] However, the **manner** in which different types of intelligence are measured is the focus of this study. It is important to determine whether dynamic assessment seeks to measure intelligence or the propensity towards intelligent behaviour or whether it is merely another attempt to introduce more definitions and jargon to the intelligence field.

2.2.1 Intelligence classifications and definitions

"Concepts of 'intelligence' are attempts to clarify and organize [a] complex set of phenomena. Although considerable clarity has been achieved in some areas, no such conceptualization has yet answered all the important questions and **none commands universal assent**" [own emphasis][3] (Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg & Urbina, 1996, p.77). Neisser et al. (1996) add that current theories of intelligence (of which there are many) contend that there are many kinds of intelligence and that psychometric tests capture only a select few of these types.

Clearly there is a divide between those who advocate the use of psychometric testing and the opposing camp who vehemently deny its overall usefulness. After all, if there is no consensus as to a

1 Please note that some references cited are relatively dated and are featured only because it was deemed necessary to evidence the progression within the field of dynamic assessment.

2 Such a term as "plagued" is not without merit, especially if one studies the amount of literature available on the topic of intelligence research.

3 Unless stated otherwise, as in this quote, all emphases cited in quotations are those of the original author and not the author of this MA study.

definition of intelligence, why are psychometric tests used at all? Apart from the fact that different types of psychometric tests exist for the sole purpose of investigating specific intelligence issues within narrowly confined contexts (Daniel, 1997), the psychometric school has received grave criticism with due legitimacy. This argument begs the question of the suitability of traditional or conventional psychometric tests. Neisser et al. (1996) discuss the current efforts in the intelligence field and systematise the conceptualisation of intelligence, referring to a number of broad classifications within the field, namely:

- The psychometric approach - involving intelligence tests, intercorrelations between such tests, the notion of the as yet ill-defined concept of "G" [4] or commonly referred to as "g"
- Multiple forms of intelligence - theories developed by Gardner and Sternberg[5]
- Cultural variation - aspects of intelligence considered intelligent in one culture and not in another
- Developmental progressions - Piaget and Vygotsky,[6] typically understood as developmental theories of intelligence emphasising opposing views as to what enables intelligence to flower, namely, biological preparedness (Piaget) or proximal development via social mediation (Vygotsky)[7]
- The biological approach - in which the aspects of brain anatomy and the workings of the brain receive more attention in the intelligence research.

Daniel (1997) compartmentalises the field into three "status of instruments", namely:

- Tests based on psychometric abilities with its concurrent emphasis on "g" and abilities commonly "…identified through factor analysis of sets of diverse cognitive tasks" (p. 1039)
- Tests based on neuropsychological-processing models and theoretically driven tests which differ in outlook and outcome from the psychometric approach, models which emphasise the workings of the brain and its constituent regions responsible for different aspects of functioning. Even though these models are said to correlate with psychometric tests, this "…does not preclude them from measuring a different system of processes, …[they] are overlapping but independent" (p. 1040)
- Dynamic assessment, which Daniel states as referring to diverse approaches to intelligence which share some common assumptions. These shared assumptions include the adverse role played by what is measured, i.e. that familiarity with certain types of thinking are necessary in order to complete certain psychometric tasks, secondly, that what is measured should be a good predictor of learning and that information gathered should be utilised for educational planning and, perhaps most importantly, the fact that dynamic assessment seeks to create enduring changes through modifiability. Daniel states further that dynamic assessment is akin to neuropsychological based assessment owing to their emphasis on cognitive processes and the "teachability of those processes" (p. 1041).

In most case-studies and formulations of intelligence, many variables are considered, most of all the environment, which itself ranges from the built environment to the social environment. Environmental effects on intelligence are implicated in many studies dealing with intelligence (Neisser et al. 1996; Sternberg, 1997).

Taylor (1994) offers an integrated approach to cognitive assessment, based on three traditional approaches to assessment. Through the fusion of these three approaches, Taylor seeks to accommodate their theoretical underpinnings and thus construct a test battery reflective of all three approaches, which can be used within the local multicultural context. The approaches stipulated are:

---

4 Advocated by Spearman and Thurstone and subsequently used and commented on by numerous intelligence researchers thereafter (Carroll in Devlin, Fienberg, Resnick & Roeder, 1997; Du Bois, 1970; Sahakian, 1981; Thomson, 1974).

5 Gardner's "multiple intelligences" theory as well as Sternberg's hierarchical theory of intelligence (Neisser et al. [1996]).

6 Although Piaget and Vygotsky are cited most frequently as representing the developmental approach, one has to take cognizance of theorists such as Bruner and Freire (Bradbury & Zingel, 1998; Moll, 1989).

7 It is Vygotsky's developmental approach that is most closely aligned with the present study in the field of dynamic assessment.

- The conventional or structural approach, which is characterised by a factor-analytic approach to the study of cognition, personality, interests and so forth, relying heavily on the scientific method, is data driven and "…rather theory-weak" (p. 185)
- The information processing approach, in which four main paradigms are discussed at length by Taylor. As opposed to the structural approach to cognition, the information processing approach is characterised by what Taylor refers to as "…[its] fine-grained [nature], they tap one or a few specified cognitive activities or processes" (p. 186). A limitation imposed by this approach, states Taylor is the need for computer administration which limits the number of persons that can be tested at any one time, yet the constructs measured by this approach are more clearly defined than those of the structural approach
- The learning or dynamic approach, characterised for Taylor mainly by the nature of the assessment procedure, and its use of "novel tasks" and measurement of adaptation to these tasks "…as a result of repeated exposure, instruction, examples or hints" (p. 88). Taylor cites the work of numerous researchers in the field of dynamic assessment, concluding that although it lends itself to cross-cultural research, it nevertheless compares unfavourably with structural tests, for instance, when considering variance in results. In other words, some testees score lower in conventional tests than in dynamic tests.

2.3 The origins of dynamic assessment - early twentieth century ponderings

Dynamic assessment is often greeted with initial enthusiasm and a positive frame of mind and is thought to be quite a unique approach to assessment. Dynamic assessment as an idea or philosophy in fact finds its roots in the early twentieth century, with the work of Binet. Binet's original intention with his 1905 test was "…that it should be used as an empirical guide to identify children who need special help and that emphasis should be on **improvement through special training**" [own emphasis] (De Beer, 2000, p. 42). Binet advocated process assessment but did not devise a test for such an assessment (Grigorenko & Sternberg, 1998). Although the aim of this section is not to elaborate on the history of intelligence testing, it cannot be ignored, as it has had such an overwhelming influence on the development of dynamic testing. A brief summary outlines the course and development of intelligence testing as a precursor to dynamic assessment.

Of interest is that mental testing has been around for over four thousand years, with the Chinese having developed standardised "civil service testing programmes" (Thorndike in Flanagan et al., 1997, p. 3), which was implemented in an attempt to select people for government service and was done because China had no ruling class at the time (Du Bois, 1970). In fact, in 1219 testing took place at the University of Bologna in which examinations were first conducted in private and then conducted in public. In England, oral examinations took place where they were introduced in 1636 at Oxford and written tests were implemented as early as 1803 (Du Bois, 1970). The need for universal education and the need to determine who would benefit from such education lead to the development of testing instruments and, spurned by the quantitative sciences, psychology found its niche in quantifiable testing procedures. Testing for the American civil service also took place as a means of evaluating candidates' suitability for government (*ibid*.).

Francis Galton measured sensory acuity and is considered the founder of the scientific study of human differences, James Catell coined the term "mental test" and initiated testing at two universities in America and Alfred Binet,[8] who believed that intelligence could only be measured by investigating mental tasks, set about devising a means to differentiate between mental retardation and other cognitive malfunctions which lead to poor scholastic ability (Ittenbach, Esters & Wainer in Flanagan et al., 1997; Thorndike in Flanagan et al., 1997). Together he and his colleague, Simon, introduced, the Binet-Simon (1905) scale which was intended to be an unbiased measure of intelligence (Ittenbach, Esters & Wainer in Flanagan et al., 1997). It was Henry Goddard who took the Binet-Simon scale to the United States (1908) with Lewis Terman adapting the Binet-Simon scales in his own work; and in 1916 and again in 1937 and 1960 revisions were made to the scales and the system was renamed the

8 Binet had in fact measured skulls, and "flirted with craniometry" following in the footsteps of Paul Broca (Gould, 1996, p. 176).

Stanford-Binet scales [9] (Du Bois, 1970; Gould, 1996; Huysamen, 1980; Thorndike in Flanagan et al., 1997).

At the time of these many developments the intelligence quotient (proposed by the German psychologist William Stern in 1911) was at once hailed and rejected. Either way, there was much controversy surrounding the idea of mental age divided by chronological age, as the rate of change during childhood and adolescence differed, leading to disparities. During the army testing programme which tested army recruits en masse in the first world war, tasks on the Army Alpha[10] were biased due to the number of testees who were not English speaking, English being the language of the tests. These test results "revealed" the very low intelligence level of average American citizens. The 1920s was a period in which intelligence testing flourished much to the dismay of those for whom it was not normed nor standardised.

Before the first quarter of the twentieth century had reached its end, the intelligence community was polarised between Charles Spearman's two factor theory of intelligence which maintained that there were "…methods for determining a general factor ["g"] underlying a group of tests" (Du Bois, 1970, p. 42; Thorndike in Flanagan et al., 1997). Spearman is also credited as having developed the idea of test reliability. Rejecting the notion of a general factor of intelligence, advocates of multiple intelligence such as Edward Thorndike and Leon Thurstone[11] came to the fore. Development of clinical batteries applicable to every phase in human development as well as the emerging technicalities with which to deal with these dimensions of intelligence led to the development of factor analysis.[12] During this period the Wechsler scales for children and adults began to compete with the Binet-Simon test and this proved quite successful both with civilian and military populations, as the test also made use of non-verbal tasks (as the Binet-Simon test was quite heavily loaded on verbal tasks). The period from 1925 - 1975 unfortunately was not a period in which great strides (in terms of theory) were made in intelligence testing, and in fact many tests were only revised in the 1960s (an exception being the Binet-Simon scales which were revised in 1916 and 1937 respectively). The work of Raymond Cattell and John Horn did however contribute ideas about the nature of intelligence such as fluid and crystallised intelligence.

The 1960s was a period characterised by lawsuits and unfair discrimination based on seemingly biased test results. Defenders of these tests such as Arthur Jensen and Richard Hernstein have caused debates and battles that have since not been settled.[13] Each side has merit, a discourse which cannot now be delved into. Since the late seventies, progress has been made in terms of addressing the vast disparities of the past in testing, such as culture, gender and class bias. Latest developments have also seen the advent of item response theory and computer adaptive testing, all welcome additions and changes to the field of intelligence testing (Devlin, Fienberg, Resnick & Roeder, 1997; Thorndike in Flanagan et al., 1997).

The development of ideas within the field of dynamic assessment can be clearly traced throughout the last century. Notions of "learning to learn" were identified, the ability to learn, progress of learning, and propensities to improve when given sufficient training, were already being pondered in the 1920s and

9 An adapted version of the Binet-Terman scales was devised by Dr Eybers in Bloemfontein in 1925 for use on white South African children (Huysamen, 1980).

10 Developed by Arthur Otis for the 1917 testing.

11 In fact there were many theorists engaged in this debate with Spearman's hierarchical one-factor theory of intelligence being supported in part by Arthur Jensen, P.E. Vernon and the multifactored theory being supported in part by Howard Gardner, J.P. Guilford, John Horn, and Robert Sternberg (Ittenbach, Esters & Wainer in Flanagan et al., 1997).

12 Cohen (1959) in fact conducted on the first factor analysis of the Wechsler Intelligence Scale for Children (Kamphaus, Petoskey & Morgan in Flanagan et al., 1997).

13 An example of this is Hernstein and Murray's now infamous "Bell Curve: Intelligence and class structure in American life" (1994) which has been held to wide acclaim and also been attacked for its racist bent. Volumes of books, papers and articles have since been written expressing these dual sentiments.

1930s (Lidz, 1987). It was during the 1940s that the notion of equating intelligence with learning was questioned and the evidence suggesting that intelligence and learning did not co-vary as measures of intelligence arose to spark more controversy. Hamers and Resing (in Hamers, Sijtsma & Rujssenaars, 1993) state that the theoretical and practical beginnings of dynamic assessment did in fact emerge at much the same time that intelligence testing came to the fore (but did not of course remain at the forefront in intelligence testing early on). Guthke (1982) notes three main reasons why the development of dynamic assessment was temporarily halted; the premature death of Vygotsky,[14] the temporary rejection of psychometrics by Marxist psychologists and Vygotsky's overemphasis on imitation. When similar ideas did surface the world-over, some independently, not all referred back to the original ideas of Vygotsky. Vygotsky was rediscovered "…at a very special historical-psychological moment: on the wave of the growing popularity of Piaget's theory and the revival of cognitive psychology in the United States" (Kozulin, 1990, p.278).

The 1950s was a period noted for its emphasis on coaching and an attempt to equalise the opportunities of disadvantaged learners when taking tests. Studies were conducted in which the effects of coaching were investigated and the differences between practice and coaching highlighted. Non-coaching research also made an appearance in this decade and the search for other alternatives which did not rely so much on coaching, but turned to hints and testing-the-limits. By providing hints to the learner one was able to investigate the limit a learner was able to attain, based on the number of hints needed to compete certain exercises.

The 1960s saw a surge in the test-teach-test model, espoused so often today as a defining criterion for dynamic assessment. Testing for educability became a focus of research and transfer (retention) scores were seen to be indicative of IQ levels. Arthur Jensen's (1961) (as cited in Lidz, 1987) research into IQ and children from both low and high socio-economic standing, indicated that although biological determinism accounted for 80% of intelligence variance, the remaining 20% was pliable and that the environment played a role in intelligence and could not be ignored. Jensen's emphasis on mediation allowed for it to be used as a differentiating factor between high and low SES children. "During the late 1960s, challenges to traditional approaches to assessment of intelligence became more profound and discussions of basic issues and underlying assumptions began to appear" (Lidz, 1987, p. 15).

The 1970s was perhaps the most notable decade in terms of expertise in the field, during which a number of prominent researchers in the field came to the fore. Of note is the introduction of the work of Feuerstein and Vygotsky to mainstream American academia. Dynamic assessment, it must be recalled, finds its origins in Europe, although a change had started to occur in America concurrently. Lidz (1987) states that the 1970s research output can be divided into three types of issues discussed in published literature, namely:

- Growing dissatisfaction with conventional types of assessment (particularly within the education system) (Lidz, 1981)
- Attempts at devising dynamic assessment devices
- Isolated research projects supporting dynamic assessment either directly or indirectly.

It was during this decade that aspects such as the narrowing of gaps between learners from different cultures became evident when tested dynamically as well as the fact that mere repetition of tests did not suffice for long term change or modifiability. Growing concern over the issues and problems associated with static testing came to the fore in Israel and America, where the work of Feuerstein and Budoff, Carlson and Lidz started to surface in the dynamic assessment field (Grigorenko & Sternberg, 1998). This decade also saw the controversy surrounding the predictive validity of dynamic assessment partially solved by certain researchers in the field. Growing uneasiness with traditional psychometrics led to uncomfortable associations with IQ tests and the like. The 1980s saw a spread of dynamic assessment research into deaf and reading-disabled populations (Lidz, 1987) and research as well as testing was no longer confined to the educable mentally retarded populations. It

14 Vygotsky died when quite young (1896 - 1934) from tuberculosis (Luria in Vygotsky, 1978).

was during the 1980s that research into dynamic assessment started to take off in South Africa, albeit not quite to the same extent to that of overseas research.

Hamers and Resing (in Hamers, Sijtsma & Ruijssenaars, 1993) discuss the theoretical and practical roots of learning potential assessment with reference to Vygotsky' ZPD (noting Vygotsky's important contribution to the learning potential approach), cognitive psychology (emphasising the search for underlying constructs in psychology), cognitive correlates approach (in which the basics of cognition are explored such as speed of processing), cognitive components approach (in which components are viewed as elementary processing units), cognitive training approach (in which large scale efforts at problem solving are attempted) and learning ability (in which both notions of intelligence and learning ability are expressed through ideas such as Vygotsky's ZPD). Together with these roots of learning potential, Guthke adds that research into thinking and problem solving, educational psychology as well as personality psychology influenced the progression of learning potential (in Hamers et al., 1993b). Within this paradigm of assessment there has been a move away from prediction-orientated assessment (although still a very important aspect) towards instruction-orientated assessment (Ruijssenaars, Castelijns & Hamers in Hamers et al., 1993). Further changes and developments within dynamic assessment (as this field does not remain stationery, as with any other field) include learning process analysis to more fully understand what occurs when learning takes place and initiatives such as video analysis, learning curve analysis, and detailed investigations of what exactly results in incorrect answers to tasks have been developed (*ibid*.).

2.3.1 Current dynamic assessment - a brief overview

Perhaps the most concise and comprehensive overview to be published in the past four years in the field of dynamic assessment, is the article by Grigorenko and Sternberg (1998) in which the authors delineate the major trends occurring within the field, convey main findings, explicate the negative and positive results of studies as well as documenting the latest theoretical and practical aims of the field.

Reference to American and western European studies form the bulk of the review although South African researchers Mervin Skuy and Ian Shochet are referred to in this article. These authors' main concern with this field of research is with the "…paucity of evidentiary support for the utility of the operationalizations of the constructs that been proposed to date" (*ibid.,* p. 76). The authors investigate the voluminous literature with specific reference to the method of testing used, target population chosen, format of testing procedure, nature of materials used, outcomes and predictive power of the approach used. Classifications of dynamic assessment, according to the authors can be listed as follows

- Metacognitive intervention in which the main aim is to teach (and modify) and is characterised by the research of Feuerstein (see section 2.9.1)
- Learning within the test - characterised by Campione and Brown's graduated prompting as well as Guthke's learning test concept (see sections 2.9.4 and 2.9.6)
- Restructuring the test situation - characterised by Budoff's training tests as well as Carlson and Wiedl's optimisation of test administration (testing-the-limits) approach (see sections 2.9.2 and 2.9.3)
- Training of a single cognitive function - characterised by Swanson's working memory battery.[15]

Four criteria are used to assess the studies grouped under this classification system and are scrutinised according to their informativeness, power of prediction, degree of efficiency and the robustness of the results. As these approaches are discussed, the authors' conclusions regarding each approach are highlighted with emphasis on the more contentious issues raised within each approach.[16]

15 See Grigorenko and Sternberg, 1998, p. 82, Table 1 in which the various dynamic testing approaches are systematically listed.

16 This is done so as to give a more balanced and accurate view of the different approaches available.

Metacognitive intervention

It is recognised that Feuerstein's LPAD is useful within the intended target population of handicapped adolescents and children but is not as useful when administered to mainstream children as "…mediation may in effect be doing nothing more than teaching to the tests" (*ibid*., p. 88). The LPAD is also seen to make use of ill-defined and overlapping concepts (see Andrews, 1996, section 4.3.11.4 who states similar findings in South Africa; Buchel & Scharnhorst in Hamers et al., 1993), while reliability issues have not been adequately investigated (Buchel & Scharnhorst in Hamers et al*.,* 1993) and administration is not standardised and, most important of all, is the length of time needed to administer the assessment battery.

Learning within the test

The problem area with this approach, as the authors perceive it, lies with the different meanings that may be associated with different hints aimed at learners with different cognitive abilities and that any improvement in task performance might not necessarily be due to the nature of hints given and transferability, but that the learner may well have increased concentration. Moreover, this approach also does not detail many findings on reliability and those which have dealt with the issue seek to address the reliability of change as opposed to the reliability of the tasks themselves. "The question is not **what** the trainer should say to the child in order to lead him or her to the solution….but **how** much should be said…in order for him or her to reach the criterion" [original emphasis] (*ibid*., pp. 95-96). Guthke's research is discussed under the European research in the field. The results assessed do not emanate from a large literature base and thus the authors were limited in their conclusions (owing to the varied languages in which this research was available). Nevertheless, it was found that in studies which investigated the associations between the measures of learning potential and other psychological variables, learning potential was not affected significantly by environmental aspects (such as parental support), and tended to correlate with creativity[17] which in turn tended to reduce the importance of non-cognitive factors on performance. There is no consensus about this research as the authors state that findings from different researchers are at times contradictory.

Restructuring the test situation

Budoff's target population has included children from disadvantaged backgrounds who perform poorly and concerns the educable mentally retarded[18] (acknowledging that there are people for whom dynamic testing is not applicable). In his attempt to aid disadvantaged learners, he makes an effort similar to that of Feuerstein, but owing to the obvious lack of standardisation in Feuerstein's approach, Budoff's approach seeks to correct for this (at least partially) by standardising the test situation. However, with the studies conducted in this approach, results are limited in their predictive information and it has been found that correlations have been obtained in studies in which sample sizes differed, and ages as well as the nature of the institutionalised retardation differed. Little attention has been given to the nature of the coaching given during administration, according to the literature review[19] (Grigorenko & Sternberg , 1998).

Grigorenko and Sternberg (1998) highlight the fact that Carlson and Wiedl were the first to bring personality variables into dynamic testing and that personality factors seemed to play the same role in static as well as dynamic testing, but that perhaps the most important finding is the significant decreases in errors following verbalisation and feedback during administration. Administration is standardised and group administered which does not adhere to the strict test-teach-test

17 See section 2.9.6.

18 Of note is that Feuerstein and colleagues have worked with retarded performers and not retarded individuals (Grigorenko & Sternberg, 1998) and there is a great difference between the two. The performance may be construed as retarded but the individual's intellectual capacity is not. This is analogous to Budoff's differentiation between gainers and non-gainers, the former being incorrectly classified as retarded but in fact are not.

19 See Carlson and Wiedl's testing-the-limits approach in which type of coaching is explicitly explained, section 2.9.3 .

methodology.[20] The conclusions reached by Grigorenko and Sternberg for this particular approach are quite encouraging, since they state that "…[their] approach explored one dimension that is important for dynamic testing: the impact of instruction and feedback" (1998, p. 99). It is a fact that information obtained in this manner can aid in remediation programmes[21] (emphasised in the work of Feuerstein, for instance).

Training of a single cognitive function

This particular grouping is not much emphasised in South African research literature[22] but deserves mention as the work of Swanson is cited later (see section 4.4). The test developed (and commercially available) seeks to assess processing potential and makes use of the information processing approach in a dynamic manner which Swanson contends is similar to Feuerstein's cognitive modifiability (Grigorenko & Sternberg, 1998). Although the authors state that the test developed by Swanson is a psychometrically robust one, findings have located themselves in the grey area which requires further testing and proof.

An issue touched on a number of times in this survey of South African research is the dual nature of dynamic assessment, the value it possesses as a modifying paradigm and as an evaluation paradigm. The two are often separated, which is not truly the ideal situation, considering that mediation and remediation go hand-in-hand in this approach. However, owing mainly to costs, the latter aspect is more often than not left out of the equation. It is this very issue that Grigorenko and Sternberg investigate. They cite Embretson[23] (1987) who delineates three main goals of dynamic assessment:

1. To provide a better estimate of an ability
2. To assess the new ability
3. To improve upon efficiency.

Essentially, it is the goal that all three should take place within any one assessment administration, but practicalities and costs often result in mediation taking place without remediation. This is of concern, and justifiably so, since pragmatic issues cannot be ignored. The article of Grigorenko and Sternberg (1998) has received much attention in this section because of its comprehensive survey on the topic of dynamic assessment as it is currently researched overseas. However, much of what they state is available in texts already perused in this chapter and will not be given attention again.

Campione in Daniels (1996) and Campione (1989) consider three dimensions according to which dynamic assessment can be defined or areas in which it can be characterised; by looking at the **focus,** in other words looking at the various ways in which potential for change can be addressed, looking at the nature of the **interaction**, in other words the nature of interaction between the examiner and examinee and lastly investigating the **target** of assessment by looking at either general or domain-specific skills.

Lidz (1992a) outlines what she perceives to be five stable research findings (gathered from numerous studies) within the domain of dynamic assessment research, namely that,

1. Mediation results in improved performance in varied contexts and samples

20 Which then is not necessarily a definitive criteria for dynamic assessment even though it is often thought to be, especially in South African research.

21 This issue of remediation is a crucial component of dynamic assessment, a component which is often not paid adequate attention in South African research and Laughon states that "[I]f dynamic assessment is to establish itself as a more useful measurement paradigm, the function of remediation is likely the most important contribution that could be achieved by this new approach" (1990, Data regarding remediation, para. 1).

22 The work of Taylor (1999) and the CogLab programme emphasises cognition.

23 Hamers, Hessels and Pennings (1996) also cite these as the main goals of dynamic assessment.

2. The practice effect[24] alone does not result in this improvement
3. Two of the most powerful mediation "devices" are verbalisation and elaborated feedback (see Carlson & Wiedl section 2.9.3[25])
4. Mediation seems to be more beneficial to lower performing/functioning students
5. When process dimensions are used (dynamic tests), as opposed to static tests, IQ is "outpredicted" (p. 125).

Lidz's main findings are those that are spread across the board and are thus general statements, whereas the conclusions made by Grigorenko and Sternberg (1998) are specific to certain approaches within dynamic assessment.

2.4 The biological basis of learning - a brief interlude

Vygotsky, the father of what can be considered dynamic assessment, acknowledged the role played by biology, having stated that "[t]he history of the development of the higher psychological functions is impossible without a study of their prehistory, their biological roots, and their organic disposition" (1978, p. 46). The biological position within dynamic assessment has not been the main focus either locally or overseas, yet it is receiving more and more attention here as an area which has much to offer the field of psychology. It is for this reason that a brief digression is included.

Biological theories of learning have much to offer the field of dynamic assessment, since dynamic assessment measures and investigates the nature of change, specifically cognitive change. As the present circumstances permit, a large proportion of psychologists do not work with the biological sub-system in human behaviour and it is the contention of this author, that physiological brain functioning needs to be investigated as a supportive methodology for dynamic assessment research. Although the works surveyed in this study do not concentrate on the biological aspects of functioning,[26] it is thought prudent to include it here, possibly to serve as an initialising attempt to highlight this area.

2.4.1 Neural basis as environmental concern

Modification of learning abilities is intimately bound up with the biological modification of the system. It is this very intertwined and interdependent nature of intelligent beings, which results in the singularly linear measurement process becoming less valid than previously thought. The environment not only plays a part in aiding the development of latent abilities (or not in some cases), but also plays a central role, as is noted by Quartz and Sejnowski (1997) who suggest that the "…cortex has so evolved so as to maximize the capacity of environmental structure to shape its structure and function through constructive learning" (p. 537). The study of the anatomy of the brain was, prior to the latter part of the twentieth century, an area studied exclusively by anatomists, and remained within the realm of biological studies. However, this has since changed and the study of the brain has now moved into the focus of both biological and psychological sciences (Kandel & Hawkins, 1992; Kandel & Squire, 2000).

Dynamic assessment asserts, on the whole, that change through mediation is possible, a change that can be seen from both clinical and diagnostic tests such as Feuerstein's Learning Potential Assessment Device (LPAD).[27] However, it must be noted that the visible change is itself a reflection of an internal (physiological) change within the brain. Frisby and Braden (1992) agree with Eysenck (1988) (as cited in Frisby & Braden) that if cognitive modifiability as a concept is to be investigated, then it has to be looked at from three overlapping or interrelated concepts, namely, biological intelligence, psychometric intelligence and social intelligence. They add that in order to assess change

24 Klauer (in Hamers et al., 1993) states that the practice effect is of major concern in dynamic testing owing to the possible threat to test validity.

25 The research of Carlson and Wiedl receives high acclaim internationally and it is anomalous, therefore, that South African researchers have not made more use of this approach.

26 In fact there is a substantial paucity of physiological studies conducted within the dynamic assessment field in South Africa.

27 See section 2.9.1 for a description of the LPAD.

on the physiological level it is a necessary requirement that the system be observed through the study of brain tissue (neurons), physiological measures (evoked brain potentials) and indirectly via chronometric tasks (speed of information processing in reaction-time experiments).[28] They state that modifiability as defined and used by Feuerstein in his approach cannot be equated to the modifiability of actual physiological changes such as structural brain changes. In order for this to take place, extended and intensive intervention is deemed necessary. They also assign the role played by biology in the mediatory process to a distal and not proximal cause[29] of intellectual performance.

Hence the domain of neuroscience becomes all the more important and valid in the study of dynamic assessment. "Within this new framework, the scope of neuroscience ranges from genes to cognition, from molecules to mind" (Kandel & Squire, 2000, p. 1113). In this context, the emphasis on neuroscience and the relation between this field and that of dynamic assessment are relevant. A major tenet, which is important for both fields, is that the brain and thus behaviour is capable of change. Its plasticity over time contradicts the long held notion of stability throughout a lifetime. Cautious words from Budoff highlight the need to view biological agents of change carefully, since intelligence is not entirely a biological phenomenon and he warns that "…attempts to identify measures of general ability that are biologically based are doomed to failure because the construct of intelligence …cannot be considered independently of sociocultural influence" (in Lidz, 1987, p. 54).

Dynamic assessment research in South Africa, as far as can be ascertained, has confined itself to the visible changes[30] that occur during and after mediatory experiences. A field which may offer assistance in trying to understand more fully how the changes come about, is the field of neuroscience. It is important to ask what the internal workings of the brain, before, during and after extensive mediation and learning experiences are. "It is widely believed that the sensory and motor cortices mature early in life and thereafter have a fixed organisation and connectivity. However, it is now clear that these cortices can be **reshaped by experience**"[31] [own emphasis] (Kandel & Squire, 2000, p. 1119). Changes occur at minute levels of brain functioning with neurons undergoing changes as a result of learning, changes that can be studied on cellular and molecular bases (Hinton, 1992; Kandel & Hawkins, 1992).

Sustained learning as opposed to short-term training has different effects on the brain, and as is seen in the discussion on dynamic assessment in Chapter 4, mediation is very often fleeting and temporary, resulting in what is perceived to be ineffectual functioning. Sustained learning, however, more often results in better performances across the spectrum of mental activities.[32] The reason for this can be partially explained by the functions of individual neurons within the brain. Different expressions of growth are seen during longer periods of sustained learning as opposed to short bursts of learning (Kandel & Hawkins, 1992). This has important implications for researchers in the field of dynamic

28 Although it is deemed too detailed for in-depth discussion in this study, it must be pointed out that the authors continue to highlight the role played by physiological structures in structural modifiability, such as biochemical and electrical changes that result from learning, maturation or damage to the nervous system, chromosomal aberrations and organic influences.

29 According to Feuerstein, Rand, Hoffman and Miller (1980), distal factors include genetic factors, level of stimulation from the environment, emotional balance between the child and parent, as well as socio-economic status of the child. Proximal aspects are the immediate lack or insufficient exposure to mediated learning experience (MLE), which results in retarded performance. However Frisby and Braden state that this definition is rather tenuous at best "…finding that a cognitive deficiency was caused by a lack of MLE does not mean that the deficiency was caused by a lack of MLE" (1992, Logical analyses of mediated learning experience (MLE) theory para. 2.)

30 Here, visible changes refer to quantitative and qualitative improvements of scores, usually an improvement in scores obtained in pre- and posttest evaluations.

31 One can substitute "experience" for "mediation" or "learning".

32 Here, reference is made to the notion of "transfer" which is frequently viewed as a measure of success within dynamic assessment procedures. If one is able to transfer rules from one set of exercises to another set of similar exercises, one is said to have transferred the learning involved from one to the other. Lidz (1987) believes that the further the transfer, "…the more desirable is the training program *[sic]* which facilitates or promotes it" (p. x).

assessment, since it is argued that educators and psychologists should not only study behavioural levels of learning but also understand the deeper levels of brain functioning[33] (at least at a rudimentary level). Often, mediation programmes are to blame for not facilitating enough change. However, it is argued that the length of time spent on mediation should be extended and it is not necessarily the content which needs changing. Learning involves continual changes in the brain and the stability model of the brain (and thus the tests which are employed to test for this stability[34]) are no longer adequate explanations of learning (which in turn necessitates new models of testing[35]).

In order to grasp the nature of the changes which learners undergo, researchers in specialised fields and disciplines need to investigate these changes.[36] "Descriptions at both levels of explanation, the biological and the cognitive, would then be required in an integrated level of explanation" (Quartz & Sejnowski, 1997, p.537). Lev Vygotsky, perhaps the most influential advocate of alternative forms of assessment, believes that the activity in the brain and psychological functions are related and studied experimental psychology in combination[37] with neurology and physiology (Cole, John-Steiner, Scribner & Souberman, 1978). "If cognitive and neural processes really do interact then an added degree of complexity arises in analyzing development, for learning may induce large changes in the structures involved in learning. This complicates matters because now **learning can alter what can be subsequently learned**" [own emphasis] (Quartz & Sejnowski, 1997, p.537). Vygotsky (1978) maintains that learning causes or at least results in mental development.

A very important concept, established though it may be, is the idea of metacognitive[38] processes underlying learning and the ability to learn. Metacognitve processes may well be more developed in older individuals, seeing that experience and time have allowed the opportunity of exploration, elucidation as well as reflection on individual thought processes. Skuy (1997) states that the concept of metacognition links up with concepts such as mediation and perspective-taking and that the combination of metacognition (of one's thinking and emotions) can help to diminish egocentric and ethnocentric thinking. He elaborates: "Metacognition…could also enhance sensitivity to multicultural similarities and differences, and provide the tools needed for transcending narrow cultural definitions and prejudices" (p. 133). According to Craig (1996), Strohm-Kichener (1983) classifies metacognition as the second level of knowing, with the process of knowing itself being divided into three levels, namely, "…basic cognitive operations such as memory or first level cognition, second level of cognition or metacognition (knowing about knowing) and epistemic[39] cognition or the rules of tasks" (p. 52). What has been referred to as dynamic learning processes and non-linear situations is echoed once again by the term that Quartz and Sejnowski use when they refer to "nonstationary learners", "…the nonstationary learner is thus one in which learning causes large scale changes to its basic mechanisms" (1997, p.537).

---

33 An example of such as approach is the work of Das, who has developed the "Das Model" which "…conforms to Luria's conception of the brain and the hypothesised three functional units that are fundamental to cognitive functioning and information processing" (Cormier, Carlson & Das, 1990, p. 438). These three functions include areas of the brain such as the reticular formation, occipital-parietal lobes and the frontal lobes.

34 Psychometric tests are being referred to here.

35 Dynamic tests are being referred to here.

36 Kandel and Squire (2000) do admit though, that this endeavour is only beginning and has a long way to go.

37 Vygotsky clearly states that "[t]he history of the development of the higher psychological functions is impossible without a study of their prehistory, their **biological roots and their organic disposition**" [own emphasis] (Vygotsky, 1978, p. 46).

38 Metacognition as defined by De Villiers (1996) is the change in learning that results from further learning. Feuerstein's Instrumental Enrichment Programme is an example of a metacognitive programme aimed at elucidating specific cognitive deficits within certain domains of learning.

39 This third level or epistemology has to do with the theory of knowledge (The Penguin Dictionary of Philosophy, 2000). It may often be considered that metacognition is the outermost level of knowing, but in truth, knowledge of one's own knowing must precede a theory of knowledge.

Quartz and Sejnowski explore a manner by which to measure representational change in an organism (notably the neural basis of cognitive development), and maintain that:

> two themes emerge from finding a structural measure of representational complexity: 1) development is a progressive increase in the structures underlying representational complexity, and 2) this increase depends on interaction with a structured environment to guide development. These form the basis of **neural constructivism** (1997, p.538).

As the authors also point out, "…this name reflects the Piagetian view that there is an active interaction between the developing system and the environment in which it is embedded" (pp. 538-539). Piaget's conception of the biological development of the system as well as Vygotsky's notion of external (social) mediation are really expressions of the idea of neural constructivism.

The propensity for the young physical brain to undergo changes in neural structure when learning takes place is similar to the changes that take place within the adult brain. Potential thus may only arise when the system is ready to take on more complex tasks. Vygotsky (1978) maintains that the zone of proximal development signifies the area of learning potential, when a child is aided by a more competent peer or adult in completing a task that hitherto may not have been completed without this aid. This potential skill with guided assistance may be a behavioural representation of a possible physiological readiness to perform the task at a later time. Even though Quartz and Sejnowski's article centres on the neural adaptability of the brain during postnatal development (dendrites and axons, in particular) the same principles may be extended to the young adult brain, in keeping with the idea of Vygotsky's zone of proximal development. If, as it will be shown, Feuerstein's Learning Potential Assessment Device (LPAD) is an assessment tool aimed at intervention, this necessarily indicates that intervention will work with certain cognitive deficits and as a result will attempt to **alter or change** these deficits. The LPAD can be used with "…learning-disabled children, adolescents and adults" (Jensen & Feuerstein in Lidz, 1987, p. 401)". If this is the case, whether the adults are learning-disabled or not, the fact remains that adult brains can undergo change.

When considering development, one cannot turn a blind eye to physiological processes which govern much of a human being's development. "Learning is a dynamic interaction between a changing structured environment and neural mechanisms. The neural machinery is extensively shaped by activity stemming from the environment, while its intrinsic properties also constrain this modulation and play an indispensable role in shaping the resulting structures" (Quartz & Sejnowski, 1997, p. 539). In fact Haywood, Tzuriel and Vaught state that neuropsychological assessment is very similar to dynamic assessment and that it rests on the assumption that "…intelligence, although largely genetically determined, may not be accessible for a variety of reasons" (in Haywood & Tzuriel, 1992, p. 54). Neuropsychologists choose to look for potential and conditions that will promote this search for potential as opposed to strict reliance on static modes of assessment which assesses the processes people use to solve items rather than assessing the correctness of the answers.

When assessing the use and benefits of cognitive development programmes aimed at allowing individuals to at least partially fulfil their potential, one must take cognizance of the very nature of the physical brain and the role that structure and neural dynamics play in cognition. No psychometric device can afford to ignore the organic basis of learning and although this area of organic study is not the focus of this investigation, it must be noted that physiological factors must be considered when understanding the global context in which psychometric testing takes place. The often mentioned dyad of individual-environment, by its very nature has to encompass the organic basis on the individual's functioning. Nature-nurture issues are continuously raised in the literature and seek not to confound the issue of learning potential but to allow recourse to study those factors which invariably influence the individual and his/her cognitive development. The individual is an open dynamic system, an organisation that cannot easily be alluded to as a result of cause and effect. The assessment approach of learning potential must be seen as an open process, seeing that "…development can only occur within dissipative (open) dynamical *[sic]* systems" (Voss in Valsiner & Voss, 1996, p.23).

According to General Systems Theory (Hanson 1995) (much akin in this regard to the ecological theory of nested structures of Bronfenbrenner [Duffy & Wong, 1996]), sub-systems are necessarily subsumed in greater systems, "[a] system can be defined as any two or more parts that are related,

such that change in any one part changes all parts" (Hanson, 1995, p.27). Adaptive self-organisation occurs in ever-changing environments and just as the plasticity of "…insect populations and ethnic groups…[may be put on a high level of plasticity]…when facing an ever-changing environment" (Voss in Valsiner & Voss, 1996, p.25) so too does the brain undergo adaptive changes to deal with emerging situations. This novelty of change can be likened to potential which resides in individuals who, never presented with the opportunity or challenge to adapt and change, may never do so.

It is this untapped change, or potential that needs to be assessed. "[T]he change of a system is a function of the system itself, [and] is also an expression of the self-organisation of the system…self-organisation means that there exist forces inherent in the system that act on structural and functional properties of the system without a direct reference to the environment" (Voss in Valsiner & Voss, 1996, p.25). This clearly underlies the need to envision a holistic view of what it means to learn and what potential may reside in individuals begs not only a systematic investigation of psychology to understand more fully what potential is and how it can be measured but also seeks to take cognizance of other aspects of how this study can be investigated. Voss in Valsiner and Voss (1996) makes mention of the fact that psychological systems are often construed as hypothetical in nature, and that there is available a considerable body of research on " …synergetics of the brain, and a large proportion of it addresses the problem of learning in the context of connectionist models of brain functioning" (p. 27).

Learning potential may prove to be a murky territory when attempting to render it meaningful using only psychological constructs.[40] These psychological constructs are necessarily relegated to the realm of rational thoughts and supposed logical arguments based on social and philosophical notions. The potential to learn can be viewed simultaneously from a psycho-social point of view as well as from a physiological view point. The concept of learning potential however has not always received favourable acclaim (as has been encountered by Jensen's (1980) view of potential, as "…a metaphysical rather than a strictly scientific notion"). Jensen further highlights his apparent frustration with a concept that when looked at purely from a psycho-social angle, cannot amount to much by way or research, "…conditional probability statements based on empirical research simply do not include or require the notions of capacity or potential" (p.243). Arguably, Jensen prefers to deal with the idea of genotypical "reaction ranges" rather than potential. This genotypic reaction range "…indicates the performance values that can be expected for a given genotype under the range of environmental conditions that actually exists in the population" (p.243). Jensen concludes then that there are no operational means as such for estimating potential. Perhaps Jensen and other thinkers along this line would be persuaded more from the side of physiological evidence for change (or learning potential) as opposed to the theoretical principles of change.

From a purely pragmatic stance, it is not difficult to understand Jensen's statement that "in principle, we have no way of knowing *a priori* what level an individual's performance might attain under unspecified environmental conditions or even under some novel specified conditions that have never existed before" (p243). The key notion in Jensen's argument is that the myriad influences that impinge on each individual are too numerous and complex to, in any measurable sense, estimate or conclude what potential may reside within. However, the argument may be taken a step further, when deploying controls and limited contextual environments in which learning capacity or potential can be assessed. There is no reason, why under controlled conditions and limited areas of applications that potential cannot be assessed. The very nature of Feuerstein's Instrumental Enrichment Programme[41] and Learning Potential Assessment Device, affords researchers the opportunities to assess heterogeneous populations, culturally deprived children, learning disabled children, adults, and also retarded people. Environmental influences act in on these various cases are numerous and indeed

40 However, it is precisely from this angle that almost all research is applied in South Africa within the sphere of education and psychological assessment.

41 The aim of this study is to investigate dynamic assessment relating to assessment and learning and not to focus on the state of the second branch of dynamic assessment, namely; enrichment programmes exclusively, and although research has been conducted in this field in South Africa (for example Coosner, 1999) it will not receive attention here.

complex, yet an attempt to aid these individuals' potentials can be made. However validity and reliability studies in more settings in South Africa need to take place.

Changeux and Konishi (1985) emphasise the larger view of learning in terms of the organism's capacity to learn which manifests on several fronts, namely that "…learning is studied at four levels, 1) the organism as a whole, 2) ensembles or populations of neurons, 3) individual nerve cells and 4) that of their constituent macromolecules" (p.6). Merzenich (in Changeux & Konishi, 1985) maintains that "…there are dynamic processes by which the details of somatosensory cortical representations are established, maintained and altered by **use**" (p.345). The key issue here is the fact that physiological change is noted when use is made of certain structures in the brain. Habituation is one example of the ability of the brain to "change" to environmental stimuli. "Electrical signalling in neuronal circuits depends on the history of **previous use and on concomitant activity** in extrinsic neuronal pathways. Changes in synaptic transmission intrinsic to circuits account for behavioral *[sic]* phenomena such as habituation" [own emphasis] (Menzel & Bicker in Changeux & Konishi, 1985, p.433). The fundamental premise on which the notion of potential resides is the capacity of the brain to change (brain plasticity), physiologically, neurochemically and structurally, having undergone "learning" in any respect. Dynamic assessment seeks to **change** aspects of learning, hence dynamic assessment and the physiological basis of learning need to become integrated in order to provide an even more comprehensive overview of assessment as the two are complementary.

On a deeper level, Von der Malsburg (in Chageux & Konishi, 1985) states that "…synaptic plasticity may be regarded as the basis of brain organisation" (p.411) and this idea of change in the functional aspects of the brain has as its beginning the work of D.O. Hebb, who as early as 1949 stated that the brain should be conceived of as a dynamic system. This type of plasticity suggests that the brain undergoes changes during and after learning. "Preliminary observations suggest that in invertebrates, as well, (in addition to) enriched and deprived environments **alter** considerably the **structural** features of the neurons" [own emphasis] (Menzel & Bicker in Changeux & Konishi, 1985, p.461).

The study of learning potential should not be the sole preoccupation of psychologists but a joint collaboration of neuropsychologists and neurophysiologists:

> The neuroscience of higher cognitive processes is only beginning. For neuroscience to address the most challenging problems confronting the behavioral *[sic]*and biological sciences, we will need to continue to search for new molecular and cellular approaches and use them in conjunction with systems from neuroscience and psychological science. In this way we will best be able to relate molecular events and specific changes within neural circuits to mental processes such as perception, memory, thought, and possibly consciousness itself (Kandel & Squire, 2000, p. 1120).

As a psychological construct, Jensen's (1980) opinion on the non-validity of measuring potential may be understood in the light of purely psychological underpinnings, but as can be seen, (Merzenich in Changeux & Konishi, 1985) change can be physiologically measured and change may be construed as one of the cornerstones of potential. The fact that the lately researched ideas (for instance, Selkoe, 1992) that the human brain continues to develop and change during the course of life, attests to the usefulness and functionality of learning potential assessments, more so today than in the past. The fact that learning potential need not only look towards children and adolescents but includes adults, is also a factor making the study of learning potential exciting. Sporadic growth is the hallmark of children's and adolescents' cognitive development, but this in no way precludes study on adult potential. During the 1985 Dahlem workshop on the neural and molecular bases of learning, it was understood that the premise of the meeting "…was that the origins of complex behaviours can be understood on a cellular level" (Merzenich in Changeux & Konishi, 1985, p.352).

According to Von der Malsburg (in Changeux & Konishi, 1985), ontogenesis is taken to refer to the part of structural genesis of the mind and brain which in turn is shaped by each individual's mental history. Phylogenesis is likewise taken to render the initial state realisable "with the help of information stored in the genes" (p.414). "The genetic brain is certainly not a **tabula rasa**. On the other hand it is to be regarded as a weakness of most classical theories that for every function there is a specialised connectivity pattern invoked, and that whenever it is difficult to explain it by ontogenetic mechanisms,

one invokes genetic determination, i.e. one makes phylogeny responsible for it" (Von der Malsburg in Changeux & Konishi, 1985, p.421). As much as individuals cannot escape their genetic inheritance, aspects thought not to be expressed in these individuals may well come to the fore under amenable circumstances (referring to the psycho-social environment which itself entails a host of other nested environments). Potential, thus, is not what an individual can ever achieve, but what he/she might be able to achieve given the right circumstances. It is this very notion that needs to be explored at length within the South African context.

2.5 To what exactly does dynamic assessment refer?

There are various definitions available which attempt to capture the essence of dynamic assessment. It would appear that the term "dynamic assessment" is a blanket term covering a myriad of possible definitions, all of which are in some way related to one another.

As will be shown in Chapter 4, this array of definitions in South Africa[42] is not limited to theory but is very much part of practice as well. Looking for definitive criteria is at least a start in the attempt to define the field. Lidz (in Flanagan et al.,1997) offers various definitions based on the historical progression within the field and states that it can be defined according to the

- Relationship between examiner-examinee
- Nature of intervention
- Model within which the assessment is carried out, typically in a test-teach-test format
- Modification of learner as opposed to assessment and categorisation of learner only
- Content on which assessments are based, typically academic content as opposed to non-academic content.

Delclos, Vye, Burns, Bransford and Hasselbring (in Haywood & Tzuriel, 1992) highlight two aspects which they state as being the most defining features of dynamic assessment, namely; that dynamic assessment attempts to change the components of a task in such a way that the learner understands what is required and secondly to experiment with the different approaches available in teaching the learner how to complete tasks. The most often cited clear-cut definition of dynamic assessment is that it usually follows a sequence of a test (usually serving as a pre-test) followed by some sort of mediation and then followed by another test (usually serving as a posttest) (Budoff in Lidz, 1987a, 1987b; Campione in Daniels, 1996; Elkonin, Foxcroft, Roodt & Astbury in Foxcroft & Roodt 2001; Hamers & Resing in Hamers, et al., 1993; Haywood & Wigenfeld, 1992; Haywood, Tzuriel & Vaught in Haywood &Tzuriel, 1992; Lidz & Pena, 1996; Lidz in Flanagan et al, 1997; Lidz in Lidz, 1987). However limiting this definition may be, it is one of the key features and most universally recognisable element of dynamic assessment. This is not to say that all methodologies which follow this format are necessarily dynamic in nature. Lidz also lists learner metacognitive processes and responsiveness to intervention as another defining feature of dynamic assessment (in Flanagan et al.,1997). In essence, the dissatisfaction with intelligence tests to test for learning capacity has spurred on the efforts of dynamic assessment researchers. "The theory of dynamic assessment raises anew the issue of the relationship between intelligence and learning ability…" (Lidz, 1987, p. 25). The question now raised is what exactly the differences are between intelligence and learning potential or is learning potential just another facet of intelligence? To date, conventional tests have not been very successful at investigating learning potential as much as they have been successful at investigating what has been learnt within certain contexts (i.e. school learning). Standardised test results frequently have little or no impact on instruction which follows as inferences about designing such instruction programmes cannot be made from the test results alone (Bejar, 1984).

42 Delclos et al., (in Haywood & Tzuriel, 1992) make mention of the fact that dynamic assessment may refer to many things, that it may be used in different ways and that results need to be interpreted accordingly. This sentiment emanates from a substantial perusal of overseas literature and it is hardly surprising that there is a lack of consensus about the field in this country as well.

2.6 Dynamic assessment and conventional psychometrics - uneasy alliance or mutual beneficiaries?

It is argued (Embretson in Lidz, 1987; Feuerstein, Rand, Jensen, Kaniel & Tzuriel in Lidz, 1987; Lidz in Flanagan et al.,1997; Sewell in Lidz, 1987) that conventional psychometrics and the comparatively newly emerged field of dynamic assessment are diametrically opposed conceptually at loggerheads and theoretically at odds. There is another view, however, which is tempered by more regard for both points of view, and that is a view which encourages co-operative engagements in measurement. Both approaches can be used together as complementary sources of information,[43] "[d]ynamic assessment is not intended as a replacement of current approaches, but as an addition to currently available procedures" (Lidz, 1987, p. 4). But first a brief look at the dissatisfaction with conventional psychometric testing and the increasing moves made towards alternative assessments such as dynamic assessment is necessary. Guthke (in Hamers, Sijtsma & Ruijssenaars, 1993a) lists the major trends in test theory, trends reflecting a focus on findings of experimental and developmental psychology in order to ground theory more fully in what practice has revealed and to make the field of testing more grounded in theory than has hitherto been the case. Indicators of intelligence other than pure theoretical conjecture are increasingly becoming important as ecological validity is playing a more prominent role. The field of biology offers a rich area of research on which measures of cognitive functioning can now build. A more holistic approach to assessment now includes personality variables and other non-cognitive factors and individually attuned testing is receiving more attention such as computer adaptive[44] testing which is dynamically adjusted to the level of the testee. Lastly, but most importantly, is the move towards a more dynamic manner of assessment away from static measures of potential.

Prior emphasis on measurement (conventional psychometrics) and later emphasis on training (dynamic assessment) are key differentiating features between the two approaches (Hamers & Resing in Hamers et al., 1993). In contrast to Minick's (in Lidz, 1987) Vygotskian interpretation of dynamic assessment, Hamers and Resing (in Hamers et al., 1993) state that the key feature of a learning potential test is that although it possesses the psychometric properties of a regular test, it differs in administration, with an inclusion of a training phase. Conventional psychometrics is interchangeably used and in many instances is synonymous with concepts like classical test theory, traditional test theory and true score theory. The latter concept has at this stage more bearing on the history of measurement theory. The problem with true score theory (which is a statistical measure of what score may have been attained on a classical test of indefinite length) (Ittenbach et al., in Flanagan et al., 1997) is that these scores are specific to the particular test taken and the testees taking the test which makes for relative results.

Daniel (1997) states that one of the main reasons for the psychometric model's success is in part due to the long history of empirical findings which "…gives it a type of robustness that more atheoretically driven models do not enjoy to the same degree" (p. 1043) and also that this model is quite flexible. However, this is not to negate the sentiments of many researchers who feel that this model has yet to demonstrate its "…practical applications and benefits of abilities in educational, occupational and clinical settings" (*ibid*.). Of course, it goes without saying that if a model seeks to overthrow the psychometric model or at least seeks to complement the model, it will be expected not only to vindicate disadvantages of this model but to address these issues itself effectively and replace them with adequate substitutions. The nagging issue with dynamic assessment is the fact that many dynamic batteries do not or cannot address aspects such as reliability and validity. The disillusionment with psychometric tests, is done away with for various reasons when scientific accuracy (as far as can

43 This is a view which is espoused by many South African researchers in the dynamic assessment field. Often citing one or the other approach as insufficient in detailing information about certain cognitive functions, an approach which makes use of both sides is usually perceived to be the best alternative.

44 Computer adaptive testing is built on the rationale that all things being equal and taking into consideration guessing, the best item to use is that item which will yield a 50/50 chance of being answered correctly. Students with differing abilities can also be assessed using Item Response Theory, seeing as the students need not write different tests (Bejar, 1984).

be obtained within the intelligence measurement field) is required. Haywood and Wigenfeld (1992) state that dynamic assessment procedures are more often than not carried out in a repeated-measures design and that when they are administered according to standardised instructions and conditions "…they provide baseline information, that is, describe [a] subjects' unassisted performance or [ZPD] …in this respect dynamic assessment **shares all the advantages of static assessment** [own emphasis] (1992, How interactive assessment can be used to get knowledge that is otherwise elusive, para. 1).

Bransford, Delclos, Vye, Burns and Hasselbring (in Lidz, 1987) state that the dissatisfaction with traditional tests most often lies with the fact that these tests deal with products of learning and do not pay any attention to the process of learning itself. Moreover, the learners' responses to the tasks within the tests are not assessed and lastly, traditional tests offer no insights as to how to plan possible strategies for intervention. These experts add that seeing that a similar baseline of functioning cannot be established for all learners (due to prior inequalities), reliance on prior learning as a possible indication of future performance (predicting future performance) will not suffice as a suitable measure; and in order to overcome this, reliance on present learning is the key to addressing potential. It is at least a more equalising process.

Craig notes that the psychometric paradigm differs from the dynamic paradigm in that, "[the] conventional psychometric [paradigm] tests that which has already developed/matured/been learned, while a dynamic framework allow[s] for the assessment of that which is in the process of maturing" (Craig, personal communication, March 6, 2001). When hypothesising about the exact nature of learning potential tests, it would seem that there is no agreed upon design or possible outline of such as test. "Searching for a test of this nature is equivalent to searching for a genotype. It is a theoretical abstraction not an empirical reality" (Miller, 1991, p.1). Coosner (1999) states that the defining feature of dynamic assessment is its attempt to "…assess processes and [continuous teaching] within the assessment" (p.6). This is perhaps the most important defining feature of dynamic assessment. All too often studies considered dynamic define themselves as such based on the "pre-test/post-test" design. The defining difference between static and dynamic assessments according to Lipson is "…the interaction and complementarity of the descriptive and remediation components of dynamic assessment that make it such a rich and useful tool" (1992, p. 7).

Lidz states that the interests of most dynamic assessment developers lie with identifying the areas of application of cognitive functions within learning situations as opposed to testing intelligence *per se* (in Flanagan et al., 1997). This process-based approach to assessment versus a product-based approach (in which the emphasis is on the end result) is a hallmark of dynamic assessment. "Standard tests analyze the student's current level of performance but provide no direct evidence regarding the processes that may have operated of failed to operate to bring about that performance…they provide at best a partial view of the testee's status" (Campione & Brown in Lidz, 1987, p. 85). It is also the contention of Haywood and Wigenfield that "…examiner inferences about deficient cognitive functions and the amount of mediation required to remediate these deficiencies constitute the most important information derived from a dynamic assessment" (1992, Psychometric properties of dynamic assessment: research and problems, para. 1).

As diagnostically and theoretically superior as dynamic assessment may be, Embretson (in Lidz, 1987) states that psychometrists are "less than enthusiastic" about this paradigm of assessment owing to a number of reasons:

- Attempts at modification or mediation may result in scores being viewed with scepticism, since what the scores now reveal is not what they initially may have revealed, hence decreasing the validity of the results
- Learning ability (potential) is seen by some as an unworkable construct
- Change measurement seems to offer more problems than solutions (previous reliance on classical test theory when measuring change is being supplanted by item response theory).

Furthermore, Lidz (1992a) adds that what is perceived by psychometrics as annoying and interfering is in fact the research domain of dynamic assessment and that it is relatively easy to generate data about conventional issues such as validity and reliability but more difficult to generate information about issues such as quality, meaning and process. This statement essentially encapsulates the

"agendas" of both types of assessments. Both methods of assessment contain distinct advantages and disadvantages.

2.7 Dynamic assessment as a means to enrichment

The essence of a dynamic approach to assessment is its endeavours to enrich those whom it tests. It does not seek merely to assess but seeks to promote growth and development, especially in cognitive areas which require developmental assistance. As is often the case though, many intervention strategies which propound dynamic principles do not embark on enrichment programmes, but merely assess in a dynamic manner. This does not fulfil the tenets of dynamic assessment. "…assessment needs **to link diagnosis with treatment** [and] that the outcome of assessment should be an array of interventions with potential for direct application to instruction…" [own emphasis] (Lidz, 1987, p. 3). The aim, then, is in short to modify learners, and teach them to "learn to learn". As is evidenced in Chapter 4, initiatives within dynamic assessment which encompass this aspect of enrichment are often cut short or disbanded due to timing and costs.

2.8 The influence of Lev Vygotsky and Reuven Feuerstein - precursors to the study of dynamic assessment

As is well known within the dynamic assessment field, both locally and internationally, the field is dominated by the thoughts[45] and theories of Lev Vygotsky and Reuven Feuerstein, emanating from Russia and Israel respectively. The works of both these theorists are cited in most studies including South African studies as being the forerunners of the dynamic movement. Dynamic assessment is synonymous with the theories, works and studies of these two researchers. Although the work of Vygotsky was not, at the time of his writing, acclaimed, it has since the 1960s received recognition for the contribution made towards education and assessment of learners. Feuerstein's contribution grew from working with culturally disadvantaged immigrants. No discourse within the dynamic assessment field is complete without reference to both Vygotsky and Feuerstein. The principle works of Vygotsky have been translated from the original Russian and provide insights into the ideas of education and remediation. Historical priority has been granted to Vygotsky as the founder of dynamic assessment, although as Grigorenko and Sternberg (1998) assert, simultaneous and independent developments within the field have occurred.

Minick (in Lidz, 1987) notes that unlike the work preceeding that of Feuerstein, dynamic assessment studies have tried to yield quantitative results in order to avoid the potential pitfalls of scant information regarding reliability and validity. He adds that Feuerstein specifically avoids a static baseline measure, as this may well impede the progress of the learner if scores are low or tasks cannot be performed adequately. In doing this, a more positive relationship between tester and testee is established, allowing the child to carry out tasks which reflect strengths and weaknesses and in order to provide more individual and specific kinds of assistance (which is most often not the case with other dynamic assessment measures). Two poles of the same continuum come to the fore (a continuum which is later seen to characterise the research in South Africa), a quantitative pole and a qualitative pole, both sacrificing the one for the other, or at least meeting mid-way in order to avoid the disadvantages of the other.

Vygotsky's ideas are more qualitative than quantitative and Minick contends that the work of Feuerstein is closer to that of Vygotsky more so than the work which is more closely aligned with quantification of learning potential. Minick also believes that the underlying or fundamental attribute of dynamic assessment which is central to all research that labels itself as dynamic is the notion that what this assessment wishes to tap is **the manner in which a child (learner) responds to assistance or help given by a more competent peer or adult.** Of particular note and of bearing here is the fact that Luria, who having worked with the concept of ZPD, paid close attention to the

45 The word "thoughts" is used owing to Vygotsky's, at times, vague explanations of his theory (Cole, John-Steiner, Scribner & Souberman, 1979) and also of Feuerstein's at times vague conceptualisations of ideas and concepts (Andrews, 1996).

degree to which benefit is derived from mediation and that an assessment measure taken after mediation and compared to a measure taken before, and investigating the degree to which the learner improves, remains the same or performs less well, is of particular importance. This need to emphasise the pre- and posttest nature of assessment is not the emphasis that Vygotsky placed on in the ZPD. **His emphasis lay with the interaction with the child and the nature of that interaction** (Minick in Lidz, 1987) and that pre-test assessments in combination with learning assessments are more powerful than when either are used alone (Day, Engelhardt, Maxwell & Bolig, 1997). This is of paramount importance today where it can be seen that almost all dynamic assessment research in South Africa has as its definitive dynamic feature, the reliance (almost exclusively so) on the pre-test, mediation and posttest methodology.

The shaky translation[46] from the Russian "blizhaishei" translates more appropriately as meaning "nearest" as opposed to "proximal" and the measure of ZPD is perhaps a measure of readiness more so than learning potential or efficiency (*ibid*.). In fact, Vygotsky has explained the notion of the ZPD as enabling a new formula to be developed, one in which "…good learning is that which is in advance of development" (1978, p. 89). The stress that is placed on collaborative activities in assessing cognitive processes is greater than the stress placed on quantifiable improvements or otherwise emanating from static measures of change, hence the role assigned to the examiner in the work of Feuerstein. The interactive nature of development which is paramount to Vygotsky is played forth in his theories of proximal development, considering that the child develops within a socio-cultural context.

Maturing functions[47] are the result of interaction and in order to measure maturing functions one needs to assess the child in interaction. Many pre-test/posttest studies do not at all emphasise this critical issue. Yet collaborative attempts themselves are determined by the level of maturation or development reached by the child at any one particular stage. The nature of ZPD would seem to be quite circular. Although the notion of ZPD is well-grounded in theory, with the implementation of the concept in both sociological and pedagogical research, experimental validation of this construct remains scarce (Grigorenko & Sternberg, 1998). Brown, Campione, Webber and McGilly encapsulate the notion of Vygotsky's ZPD as follows:

> [ZPD] marks ever-changing boundaries of competence within which a child can perform with and without aid. At the lower boundaries are those …cycles already compelled, a conservative estimate of the child's status. At the upper boundary are the estimates of just-emerging competencies that are actually created by interactions in a supportive context (in Gifford & O' Connor, 1993, p. 137).

Hedegaard (in Daniels, 1996) states that Vygotsky's notion of ZPD represents a connection between a general psychological theory of child development with a pedagogical perspective on instruction. "The underlying assumption behind the concept is that psychological development and instruction are

46 A major problem often encountered with the works of Vygotsky at least when his work was initially coming to the fore was the poor translations of much his work. An example is the idea of "mediation" which is a translation of the idea of "…the individual [who] actively modifies the stimulus situation as a part of the process of responding to it" (Cole et al., 1979, p.14) and not in the usual manner in which stimulus/ response is understood. Vygotsky is not to be construed as a stimulus-response theorist but interposes mediation between the stimulus and response (Vygotsky, 1978). Also translations of Vygotsky do not always contain the arguments as Vygotsky may have seen them and in some editions of his works and papers, certain arguments are left out (Minick in Lidz, 1987).

47 Vygotsky highlighted the role played by speech as a tool and stated, "[t]he child begins to perceive the world not only through his eyes but also through his speech. As a result the immediacy of 'natural' perception is supplanted by a **complex mediated process**; as such, speech becomes an essential part of the child's cognitive development" [own emphasis] (Vygotsky, 1978, p.32). Internal speech is an example of internalisation, with its own code and language system which is unique to each individual. Internalisation requires that what is to be learned has to be understood and thus the transfer of external codes (language) into internal codes is conducted and in so doing modifying the external or original language (Das & Conway in Haywood & Tzuriel, 1992).

socially imbedded…" (p. 171). Society and social relations are of paramount importance when considering the idea behind ZPD.

One of the main differences between Vygotsky's and Feuerstein's approach is that, as expressed practically, Feuerstein seeks to modify deficient cognitive structures within the learner and does not reach for the zone of proximal development as expressed by Vygotsky. Both however seek to assess general skills from a clinical assessment process (Zolezzi, 1995). Nearest development, the readiness to proceed with the next stage of development (as viewed by Vygotsky) in social collaboration is the emphasis on ZPD as opposed to the emphasis placed on modification of processes with Feuerstein. Both views value the role played by competent others in helping the child/learner but the subtle placing of directing change is noticeable.

2.8.1 Vygotsky's ZPD, a rival for Piaget's notion of biological readiness

Vygotsky viewed Piaget as one of the very few psychologists who recognised that spontaneous and non-spontaneous concepts had different origins and goals. Piaget researched the cognitive development of children, noting that concepts such as everyday objects within the reference framework of children developed spontaneously, such as "food", "dog", "house" and the like. Scientific concepts which have to be imparted to the child by a teacher or parent, such as "political" and "mathematical" are clustered together as non-spontaneous. However, the difference between Vygotsky and Piaget is their respective understandings about how these concepts arise and how they interact with one another. Vygotsky pointed out three mistakes that in his opinion Piaget had made about these issues, namely that;

- Spontaneous concepts arise naturally whereas scientific concepts have to be imparted to the child
- There exists a barrier between the two types of concepts because of this
- The spontaneous and scientific concepts are antagonistic in the sense of the child's spontaneous concepts eventually being supplanted by the scientific or non-spontaneous concepts.

(Minick in Lidz, 1987)

South African research (Moll, 1989) has been devoted to identifying the essential differences between the theories of Piaget and those of Vygotsky. Piaget was interested in the production of knowledge and was mindful of the limited interpretations of correct answers given to questions in static tests. He was more concerned with what could be gleaned from a child's cognitive functioning based on incorrect answers rather than correct answers (Lloyd, 1995). Piaget's cognitive developmental approach is briefly characterised by the following ideas and notions:

1. Complex notions such as thought, memory and language development are best understood when investigated from the period at which they manifest themselves
2. Each new stage of development in a child's mental world is one of sudden leaps and bounds and is not a slow process
3. Children continually build their own worlds and are active participants not passive bystanders; they construct their worlds
4. The path of development is one of interaction and not innate development; growth is evidenced in an interactive environment with others
5. Change is brought about by a constant search for equilibrium and a person is said to be in continuous conflict.

(Lloyd, 1995)

Vygotsky's theory emphasised the role of social factors, an aspect Piaget has been criticised for not attending to in his theory. Vygotsky stressed the role of instruction and the learning that occurs through this interactive instruction (role of environment) whereas Piaget stressed adaptation to the environment (the role of the individual). Unassisted learning (Piaget) versus assisted learning (Vygotsky) and the role accorded the individual and environment then are major differentiating aspects of both developmentalists' theories:

> Vygotsky supported the view that social interaction and instruction require appropriate levels of development to be successful, but he rejected the notion that instruction must await the development of all the mental functions that are required for the independent performance of the tasks that provide the content for instruction (Minick in Lidz, 1987, p. 128).

In essence, Vygotsky sums up his idea of what he refers to as the "zone of proximal development" as a notion in which " 'good learning' is that which is in advance of development" (1978, p. 89). Moll (1989) offers a lucid and comprehensive summary of the key ideas of both Vygotsky and Piaget and is summarised as follows

- The theoretical ontology of Piaget is based on biological structures which constitute cognition, whereas Vygotsky's theoretical underpinning is that social relations constitute cognition
- The parameters of the mind are for Piaget structural whereas they are functional in Vygotsky's theory
- The motor (or driving force) of cognitive growth is the need for equilibrium for Piaget but the need for mediation in Vygotsky's theory
- The conception of teaching is viewed as taking place from the inside-out in Piaget's view in contrast with the view of Vygotsky for whom the process is reversed and takes place from the outside-in.

Daniels (in Daniels, 1996) agrees with the view taken by Bidell (1992) as cited in Daniels that the two should not be cast as opposing views and that the strengths of both be drawn together. He states that differences between the two theorists "…are to be found in the extent to which the social dimension of development is refined and also how **relations** between the social and the personal are conceptualised" (p. 16). From this brief look at both these developmental psychologists' views of learning and developing, one concludes that the two theories are not opposed to such a degree that they cannot be used in collaboration with one another. Points of divergence as well as of convergence make it such that the differences between Vygotsky and Feuerstein are akin to the difference between Vygotsky and Piaget. The two theories can most certainly work hand-in-hand in future dynamic assessment research.

2.8.2 Non-cognitive factors within dynamic assessment

Conservative predictions of behavioural factors can be made based on performances on learning potential tests as well as predictions on criterion scores such as academic performance and the like. The scope of learning potential tests within the dynamic assessment field seems to be flexible in this regard. Although not intended as a study dealing with personality factors *per se*, Meijer and Elshout (2001) hypothesised that "…two factors, namely mathematical ability and test anxiety should suffice to explain the discrepancies between … performance on mathematics learning tests and conventional mathematics tests" (p. 105-106). If it is true that anxiety bias plays a large role in the difference between static and dynamic tests scores, then Vygotsky's zone of proximal development could be said to be not necessarily valid. It was found that Vygotsky's zone of proximal development as a construct did indeed provide a measure of prediction in the Meijer and Elshout (2001) study encompassing conventional and learning potential tests. The ZPD can thus not be discredited as a construct. However, it was also been found that test anxiety does not explain the amount of predictability as was previously thought, but that lack of self confidence did indeed play a role (one of three factors of the concept of test anxiety which was revealed after factor analysis).

Meijer and Elshout (2001) conclude that even though lack of self confidence (part of the construct of test anxiety) "appears to be an important constituent of test anxiety… the results indicate that the observed differences between a conventional, static testing procedure and an experimental, dynamic testing procedure for mathematics cannot be explained sufficiently by a differential bias towards test anxiety" (pp. 93-94). The reason why these results have been included in this discussion on personality variables is to highlight the importance of non-cognitive factors in the field of dynamic assessment, especially in South Africa, where the majority of students' results may be said to be influenced by these non-cognitive factors. These sentiments are echoed by Huysamen (1996), "[t]hroughout the world it has been found that groups formed in terms of some or other biographic or demographic variable perform differently on tests of cognitive ability" (p.201).

The Meijer and Elshout results indicate the role played by lack of self-confidence (even though the role played by this factor was not as great as anticipated by the authors). This underscores the need to look more closely at how dynamic assessment can be harnessed in South Africa. Those students

who are more anxious in test situations are also more able to benefit from dynamic assessment as opposed to the more conventional test situation. Meijer (in Hamers, Sijtsma & Ruijssenaars, 1993) states that since anxiety leads to lessening of test scores, those aided during the test will be more able to complete questions correctly (a distinct form of dynamic assessment) and thus "[t]he hypothesis is that the predictive validity of test scores that are weighted according to the amount of help a subject needs to solve the test items, is higher than the predictive validity of test scores that only reflect the independent achievement of a subject" (p. 343). Meijer states that merely making available help in a test situation "…can alleviate tension as experienced by subjects so that concentration can be directed at the task instead of being disturbed by worrisome thoughts about failing…"(1993, p.360). This furthers the need to make use of dynamic assessment, as such factors, which would usually impinge negatively on test performance, cannot always be overcome effectively.

2.9 Various theorists and their contribution to the field of dynamic assessment

Daniel (1997) classifies dynamic assessment into two subgroups based on the administration procedures used, namely, "… clinical, nonstandardised intervention by the examiner to reveal the cognitive processes in which the [aim is] to improve … cognitive processes" (p. 1041) and secondly, approaches which are standardised and make use of objective measures. The former approach is typically characterised by Feuerstein's Learning Potential Assessment Device, for example, and the latter being typically characterised by the work of Campione and Brown for instance.

Dynamic assessment as a term is indeed vague and difficult to define but this is hardly surprising given the number of definitions and practices by which it is known and characterised, as already discussed. There are numerous approaches within the field of dynamic assessment; only the broad defining and most well-known are discussed here. Coosner (1999), supported by Bransford et al*,* in Lidz (1987), highlights the similarities and differences between the varied approaches within dynamic assessment (see below) and states that each approach attempts to link instruction with assessment. There are, however, significant differences in the manner in which they do so as well as the emphasis they place on certain aspects such as:

- Theoretical orientation
- Purpose of assessment (identification, classification, prediction, informing instruction)
- Types of skills assessed (general or domain specific)
- Types of tasks employed
- Nature of interactions between the examiner and the learner
- Quantity of empirical support.

2.9.1 Feuerstein's Learning Potential Assessment Device and mediational assessment

Perhaps the most widely used dynamic assessment method in research in South Africa today is the Learning Potential Assessment Device (LPAD) as developed by Feuerstein in collaboration with Rand and Hoffman (1979) and this is also considered to be one of the most important and influential theoretical bases for dynamic assessment (Lidz in Flanagan et al*.,* 1997). Essentially the LPAD (mentioned earlier in the study) is concerned with structural cognitive modifiability (SCM), the notion that transformation in a person's intellect may deviate quite significantly from the expected course of development through the aid of learning through teaching. In fact, Feuerstein, Feuerstein and Gross (1997) state that SCM is as much a generator of LPAD as it is a product. Cognitive functioning in turn is dependent on a learner's prior mediated learning experience (MLE) (Jensen & Feuerstein in Lidz, 1987). Lack of appropriate mediated learning experience early in life may well result in passivity, reduction in modifiability as well as non-cognitive factors such as personality and motivation being affected adversely (Hamers & Resing in Hamers et al., 1993). Furthermore, the LPAD is particularly suited to the South African context where cultural differences and disadvantaged backgrounds (socio-economic as well as educational) play a major role in how learners perform on conventional intelligence tests. Feuerstein seeks to liberate human potential which is partially an expression of mistrust of static testing. Measuring instruments tap limited types of cognitive functioning and the issues of increasing reliability and validity seek to stabilise intelligence measures thus not allowing for the notion of change which, if detected in a static instrument, is regarded with scepticism. Hamers and Resing (1993) state that Feuerstein distinguishes between two forms of learning - by direct exposure

to learning and mediated learning, the former relying on a continuous process of learning by trial and error and the latter in which learning is mediated by a competent adult or peer (in Hamers et al.).

The LPAD and the movement from which it springs seeks to address issues such as predictability, human change, measurement, labelling and the invasive influence of sociological functions of classifications (Feuerstein et al., 1979) with humans seen to be open systems that are susceptible to cognitive changes brought about by structural modifiability (Jensen & Feuerstein in Lidz, 1987). Instead of reflecting inequality, static tests are said to create inequality by perpetuating stereotypical results in such test batteries. The target population for whom the LPAD would be most effective is culturally disadvantaged learners and/or retarded learners, a group for whom static measurement lacks appropriate measures of assessment (Feuerstein in Cronbach & Drenth, 1972). Administering the LPAD on efficient and high performing learners would be superfluous (Feuerstein et al*.,* in Lidz, 1987).

The LPAD represents not only a new shift in testing theory but in fact a radical departure from tenets espoused by classical test theory and is considered by some to reflect a paradigmatic shift akin to Kuhn's ideas of scientific revolutions. The LPAD is seen as fulfilling the three criteria set forth by Kuhn as a new paradigm, namely, disenchantment with current psychometric testing, social pressure from dissatisfied groups who may have been "victims" of testing, and the need for a new paradigm, which is considered to be free of the inconsistencies of the previous paradigms[48] (Feuerstein et al., in Flanagan et al.,1997). The theoretical basis of the LPAD as an assessment tool sprung from the myriad questions with which psychologists were faced when using classical test theory concepts that were found to be inadequate in explaining much of intelligence and behaviour. Issues, such as whether or not psychometrics test traits or states, whether characteristics are acquired or innate, whether these traits/states are fixed or malleable, and whether low scores on psychometric tests are really representative of an individual's adaptive behaviour (Feuerstein et al. in Flanagan et al., 1997), were some of the aspects put to psychometric practitioners.

The various test instruments used in the LPAD battery can be distinguished according to four main characteristics, namely,

- The structure, nature and modality of the instruments which differ from other traditional tests and seek to investigate and evaluate changes and not yield scores
- The tasks which are able to detect changes
- The assessment procedure which follows a test-mediate-test format
- Microchanges in behaviour that can be detected as opposed to only detecting macrochanges.

(Feuerstein et al, in Flanagan et al., 1997; Feuerstein, Rand & Hoffman, 1979)

Buchel and Scharnhorst (in Hamers et al., 1993) state that the tests in the LPAD battery are intended to illicit from the testee a broad range of cognitive functions making use of higher mental processes, tasks that are not reliant on prior knowledge and which reflect modifiability in testees by detecting microchanges. The LPAD consists of a number of sub-tests which assess various cognitive functions. A variety of these sub-tests is used in the South African[49] research literature and warrant brief descriptions:

1. The organisation of dots - which consists of connecting dots with lines in order for shapes to be formed and is a test of general intellectual ability
2. Progressive Matrices - which uses original versions of the Raven's Progressive Matrices and follows a mediate-test procedure. Mediation is given to ensure motivation
3. Representational Stencil Design Test - in which the testee has to identify and order a number of stencils that have been superimposed on one another to present different patterns. Testing follows mediation

48 Of course this does not mean that the LPAD does not carry with it its own inconsistencies which it invariably does.

49 Numbers 1, 2, 4, 5 and 11 are the most frequently used in the South African research surveyed here when the LPAD is administered.

4. Numerical Progressions - ascending and descending number series for which rules have to be worked out in order to progress to the next number in the series
5. Complex Figure Drawing Test - in which a drawing is copied and later tested from memory. Testing takes place before and after mediation
6. Positional Learning Test - which assesses organisation principles in practice using a memory task, and level of mediation differs on the need to aid the learner
7. Associative Recall - in which visual and associative memory is assessed
8. Test of Verbal Abstracting - a verbal similarities test which is similar to those found in standard static tests with training (if need be, this can be administered in graduated form) being separated from testing
9. The Plateaux Test - spatial organisation is assessed as well as the ability to learn from repeated exposure
10. Word Memory Test - which assesses the extent to which organisational principles are applied in a test of free recall of words
11. Organiser - assesses the extent to which information can be elaborated upon, based on the manner in which hypotheses are formulated and tested.

(Buchel & Scharnhorst in Hamers et al., 1993; Feuerstein, 1979)

The identification of deficient cognitive functions are revealed by the LPAD, and the cause of these deficient functions is ascribed to lack of prior mediated learning experience. Cognitive tasks yield areas of functioning which would otherwise not be assessed and the cognitive "map" of Feuerstein includes the assessment of mental acts according to seven parameters, namely: the content, the modality of presentation, the required phases (such as the in-put, elaboration and out-put phase), the operations required to complete the tasks, the level of abstraction, level of complexity, and level of efficiency (*ibid.;* Tzuriel & Klein in Lidz, 1987). The group administration of the LPAD can be successfully implemented and may even serve as a preliminary assessment prior to individual testing (Feuerstein et al., 1979; Rand & Kaniel in Lidz, 1987). Group administration does not (and cannot) replace individual testing and is used to establish a baseline of functioning, but interpretations as to performance on the group administration are not as straightforward as with individual administration. High performance on the group administered LPAD reflects high performance but low performance may be due to the nature of the administration and not to the functioning of the individual, as the LPAD is intended for individual assessment. Furthermore, Rand and Kaniel (in Lidz, 1987) warn that group testing of the LPAD should not be used as the only instrument for decision-making in selection and placement situations. Frisby and Braden (1992), however, are not entirely convinced of the usefulness and indeed theoretical underpinnings of the group administered LPAD, as they see the group administration moving more into the domain of conventional psychometrics, "…DA [dynamic assessment] supporters need to do a better job of articulating why group DA is needed and how it related philosophically and theoretically (if at all) to individually administered DA. Otherwise, group DA will appear to be gradually conforming to the assumptions that underlie traditional static tests" (Logical analyses of mediated learning experience (MLE) theory, para. 7).

2.9.1.1 Another view of dynamic assessment with reference to the LPAD

However, not all researchers are ready to accept the results from LPAD studies and very important contributions are brought up, in particular, by Frisby and Braden (1992) who assess Feuerstein's LPAD in terms of semantic, logical and empirical results and their arguments are sobering. They state very early on in their critique that "[I]t is concluded that evidence casting Feuerstein's approach to dynamic assessment as a serious competitor to 'traditional' assessment is not compelling" (Historical context in North America, para. 9). Frisby and Braden assess the work done in the United States which makes use of the LPAD and Feuerstein's mediatory concepts and ideas. They applaud the intentions behind the approach but state that due to the lack of empirical support, the ideal notions and ideas can only be relegated to the realm of philosophy. They maintain that concepts such a structural cognitive modifiability cannot hold sway in a system which does not fully explicate the meaning of "potential". Haywood and Tzuriel (in Haywood & Tzuriel, 1992) are in fact in agreement with Frisby and Braden when they state that empirical evidence in the field of dynamic assessment in general "…has not been blessed either with great breadth [n]or with great volume" (p. 505). In this context, potential is seen as some aspect which has not yet developed and Frisby and Braden (1992) argue convincingly that modifying something which has not yet been formed is near impossible. Moreover,

they maintain that Feuerstein makes use of Piagetian concepts which do not fit into the scheme of the former's approach towards assessment, as Piaget spoke about enduring changes that occur within the individual, and cognitive modifiability attests to deep structural changes, changes that are not in fact "deep" at all. They support this argument by citing Spitz (1986) who found that most compensatory education efforts either show gains which level out or in fact "wash out in follow-up testing" (Semantic analysis of selected terms, para. 14).

Frisby and Braden objectively critique dynamic assessment from an informed point of view and their intentions are not to discredit the new manner of assessment but rather to bring to it a sense of scientific rigour by noting the many contradictions and methodological errors that pervade this field. Measurement of processes as opposed to measurement of products is a key differentiating feature of dynamic assessment, but they state that however much processes are sought after, the fact remains that what is measured at the end of a test session is product-based, as answers have to be scored in some manner or other, assigned a value or rated on a scale. "The extent to which an assessment procedure can be called 'dynamic' is a question of degree, not kind" *(ibid.,* Semantic analyses of selected terms, para. 20). Frisby and Braden delineate three main areas that dynamic assessment supporters need to take a closer at, namely,

- The principles of MLE as opposed to similar principles used in parent-training and techniques embedded in other types of special instruction need to be differentiated. MLE needs to be clearly differentiated from the many varied techniques available
- The mechanisms used in MLE that influence cognitive functioning and the move from the abstract to the concrete need to be specified
- The need to justify why MLE should be a preferred theory of assessment as opposed to many other explanations available.

Although most texts are in general supportive of the results yielded by dynamic assessment studies, the well researched critique of Frisby and Braden cannot be ignored and their objections to certain issues within this field are justifiable. Their article should spurn the further development of dynamic assessment and not hinder it in any manner. Tzuriel (1992) however debates the sentiments and arguments proffered by Frisby and Braden by highlighting seven faulty areas which he finds in their article, namely, the incorrect application of a paradigm when critiquing Feuerstein's approach, the lack of an historical context for the development of dynamic assessment, failure to mention the philosophical approaches behind both paradigms, misinterpretation of the structural nature of dynamic assessment, ignoring recent research on the validity and reliability of dynamic assessment approaches and lastly incorrectly implying that static tests measure cognitive processes as personality-motivational factors. One cannot ignore what Tzuriel has stated yet Frisby and Braden deserve acknowledgement in their efforts to at least keep dynamic assessment research "on its toes" so to speak. In an effort to remain as scientific as possible in the investigation into the utility of dynamic assessment, the field needs to be scrutinised and debated, as this is the way in which any field progresses.

2.9.2 Budoff and the test-train-test assessment

Budoff's review of studies which were conducted in the late 1960s and early 1970s (in Lidz, 1987) indicates the potential usefulness of employing learning potential tests in conjunction with personality variables. According to various findings in Budoff's review, personality variables (such as self-concept, level of aspiration, success striving vs. failure-avoidance, frustration, task dissonance, impulse activity, self-perception, social interests and familial relations) are more often positively correlated with scores from a variety of learning potential assessments studies Budoff and colleagues developed in order to test this assumption than not:

> Given the hypothesis that the more able LP [Learning Potential] student is educationally not mentally, retarded, and that the improved response following training reflects increased cognitive ability, we hypothesised that the more able students by the learning potential criterion should be more able to manage and control their behavior *[sic]*, cope more effectively with stress, report better adjustment, and understand and plan more realistically for their future than the less able LP nongainers (Budoff in Lidz, 1987a, p.67).

Budoff's early work in this field can be traced to the reading of a paper by Luria (who was a collaborator of Vygotsky in the 1920s and 1930s) who discussed the notion of ZPD (Brown et al., in Gifford & O' Connor, 1993; Minick in Lidz, 1987). Budoff's contention that educational opportunities which have not been afforded children and learners from either poorer circumstances or from different cultures, often resulted in such poor test results that these learners were assigned the label of retarded or learners of low intelligence. Static tests are not able to discern any potential that may reside within these individuals. Budoff's approach is mainly characterised by the concept of intelligence which stresses instructability or the ability to profit from learning experiences which are directly related to the task being assessed (Brown et al.). Two types of mentally retarded learners can be identified according to Budoff, those who do not profit from training and those who do (Harrison, Singer, Budoff & Folman, 1972). The former group can also not adapt to circumstances outside the classroom context whereas the latter evidence adaptability (Hamers & Resing in Hamers et al., 1993).

Budoff furthermore differentiates between high and low learning potential scorers, noting that high scorers and gainers can be classified into the former group and nongainers are classifiable into the latter group (Budoff, Meskin & Harrison, 1971). In a study conducted by Budoff and colleagues, it was found that during a nonverbal procedure for assessing learning potential, high scorers (those evidencing initially high scores on the criterion [knowledge of electricity]) as well and gainers (those receiving initially low scores but yielding significantly better results after mediation) learnt more about a specific subject in the classroom as opposed to the nongainers (those failing to benefit from mediation attempts and who also scored low on the initial tests). The results show that previously diagnosed mentally retarded persons were in fact classifiable as educationally retarded. This points to prior learning experiences and factors not necessarily inherent within these individuals.

Budoff's work is centred more so on empirical findings rather than theory (Haywood & Wigenfeld, 1992), and as early as 1964, Budoff was able to show that educable mentally retarded teenagers, when given a sufficient training and practice in certain skills were able to perform better on these tasks than those in which they had not received training. He concluded that obtaining an estimate of potential might "…be a useful supplementary technique to predict adult status" (Budoff & Friedman, 1964, p. 434; Folman & Budoff, 1971). Essentially the work of Budoff and his colleagues is based on the premise that some educable mentally retarded learners are more capable of learning than static IQ tests would usually reveal; and in the attempt to assess these individuals, general intelligence in the construct investigated and in general this aspect (or "g") is not related to the more verbally attuned static measures but rather to non-verbal measures of intelligence (Laughon, 1990).

Budoff's approach is one of standardised assessment of general skills as opposed to the clinical approach towards assessment of general skills used by Vygotsky and Feuerstein (Zolezzi, 1995). Perhaps the most widely implemented manner of dynamic assessment in South Africa is this standardised assessment of learners as evidenced by Budoff, making use of conventional tests in a dynamic manner. In his research over the years, Budoff has made use of the Kohs Block Designs, the Raven's Progressive Matrices, the Series Learning Potential Test and the Picture-Word Game (Budoff in Lidz, 1987b). Budoff states that research with the Raven's has shown that group training was as effective with black students as it was with white students regardless of socio-political background. Budoff also contends that posttest scores will predict retarded learners' abilities far better than conventional IQ scores or in fact the pre-test score (which is merely a static measure). The introduction of training into the assessment procedure results in a more pleasant assessment session and that merely "…expanding on the salient features of a problem resulted in considerably enhanced problem solving among these low-functioning, substantially mentally retarded adolescents and young adults" (Budoff in Lidz, 1987b, p. 190).

The standardised yet dynamic assessment approach of Budoff allows for a more economical approach to assessment and also allows for meaningful interpretations of the posttest score. He has also made use of the "train-within-test" method and states that there are a variety of training-based assessment models. Training forms part of the assessment with neither a pre-test score nor training session as such available, and, as Hamers et al. (1996) observe, this avoids the problematic issue of change measurement. Essentially, Budoff has tried to show that some so-called mentally retarded children may have been wrongly diagnosed and this can be shown if and when these children perform at similar levels to that of average children on tests such as the Wechsler or Binet if they are given a

short period of training. Average performance from these children then is evidence that intelligence can be defined as the ability to profit from experience, in this case learning experience on a task (Budoff in Lidz, 1987a). However, it remains the task of the educator to develop this "hidden" potential fully through a comprehensive education programme.

2.9.3 Carlson and Wiedl and testing-the-limits assessment

Testing-the-limits, a procedure in which a variety of aid is given, such as standard instruction, verbalisation during and/or after the test, simple or elaborate feedback as well as a combination of feedback and verbalisation. is used to assess learning potential (Hamers & Resing, in Hamers et al., 1993). A variety of testing-the-limits measures is used in current dynamic assessment research within this approach. Carlson and Wiedl (1978)[50] for instance employed various such approaches when administering the Raven's Coloured Progressive Matrices to children between the ages of eight and twelve. They found that performance increased significantly according to the test version, testing conditions and format of testing (repeated measures). Testing conditions included verbalisation and feedback which were most effective in increasing performance on tasks, and these testing conditions were so designed as to allow for comparisons between variables. They entailed **standardised instruction** after instructions were given (condition one); **verbalisation during and after a solution**, during which the child would describe the main pattern before identifying an answer and when an answer was chosen, would describe why the particular answer was chosen (condition two); **verbalisation of answer and decisions made are described after solutions** (condition three); **simple feedback** in which children are merely informed as to whether their decisions were correct or incorrect (condition four); **elaborate feedback** during which children, in addition to the simple feedback are given reasons as to why their choice was correct or not by having the principles of the task pointed out (condition five); a combination of conditions two and five including **verbalisation before and after** solutions are arrived at in addition to **elaborated feedback** (condition six). These conditions are the hall-mark of Carlson and Wiedl's testing-the-limits research in this field and their approach is noted as having stronger links to empiricism than to theory (Haywood & Wigenfeld, 1992).

Early studies evidenced that assessment procedures involving either elaborated feedback or verbalisation led to higher performances on the Raven's Progressive Matrices and also more qualitative information such as test anxiety being reduced and less negative orientation to the test situation when dynamic assessment techniques were applied (Bethge, Carlson & Wiedl, 1982). In an attempt to gauge the construct validity and predictive validity of how these conditions may have affected the test results, and to explore the interactions between these variables and the different testing procedures, measures of intelligence, personality and criterion scores - such as school achievement - were taken. The sample size of this study was N = 108 and the sample was split into second and fourth grade pupils, who exhibited learning difficulties. A major aim of the study was to determine the effectiveness of the conditions on children with learning difficulties. The results from the study investigating children with no learning problems were very encouraging, with conditions two, five and six leading to higher levels of performance than conditions one, three and four. The study yielded positive results regarding improvement in performance after the various conditions were implemented, and "…elaborated feedback offered by the experimenter (condition five), coupled with the children's verbal description of the task prior to solution and explanation of choice after solution (condition six), provide the quality and quantity of information that develops appropriate analytic strategies for abstract reasoning, **regardless of whether the child is designated through IQ measurement as one with learning difficulties or as 'normal'**"[own emphasis] (Carlson & Wiedl, 1978, p. 563). The efficacy of applying certain testing-the-limits procedures has been shown. Carlson and Wiedl state furthermore that that their approach is directly applicable for use by psychological practitioners, as it avoids problems related to measurement of change and "...it allows for direct analysis of how and to what extent sources of individual differences on putatively noncognitive variables affect test performance" (in Haywood & Tzuriel, 1992, p. 170).

50 Although this text is dated, it is considered a primary text in the testing-the-limits approach within dynamic assessment and is included here for this very reason.

The approach of Carlson and Wiedl is not limited to the above-mentioned six conditions but can be integrated with other approaches within the psychological measurement domain, such as with the information processing approach. In 1990, Cormier et al. investigated the relationships of dynamic assessment involving verbalisation, planning ability and cognitive performance. They combined elements of differing developments by "…applying a specific model of dynamic assessment [testing-the-limits] to the measurement of mental abilities derived from a particular neuroanatomical model of cognitive functioning and information processing" (*ibid.,* p. 438). This cognitive model is known as the Das model.[51] Positive results show that poor planners can perform significantly better under the verbalisation condition (where they verbalise what steps they are taking to solve a problem) and this tends to compensate for the poor planning abilities of children to such an extent that they even outperform good planners in some instances. Verbalisation has been affirmed as a method that works in aiding children in their task and planning forms an important part of intelligence (a component not measured by static tests). The study thus serves two purposes, researching the role of verbalisation and planning.

Carlson and Wiedl, Guthke as well as other Dutch researchers (Hamers, Hessels, Sijtsma and Rujssenaars) within the field represent the German school of dynamic assessment (Brown et al., in Gifford & O' Connor, 1993). This approach of testing-the-limits does not feature to any measurable degree in the South African literature and research in dynamic assessment.

2.9.4 Campione and Brown and graduated prompting assessment

Campione and Brown[52] (in Lidz, 1987); Hamers & Resing (in Hamers et al., 1993) and Laughon (1990) state that their approach to assessment and instruction is heavily influenced by Vygotsky and neo-Vygotskian approaches in determining possible reasons for the academic delay of school children, yet it does not represent Vygotsky's views unchanged but is also a mix of their own theories of cognition and instruction. Several features underlie their alternative approach to dealing with static tests, namely,

- Assessment should evaluate processes underlying particular performances
- Assessment should take place within a specific domain rather than align itself with a general assessment of intelligence (this will lead to instruction-specific attempts at remediation)
- Assessment should itself be a dynamic process, in other words, re-diagnosis should be routinely implemented as change is occurring throughout assessment and once-off assessment and diagnoses negates the core underlying feature of dynamic assessment.

Working within domain-specific set-ups, Campione and Brown have concluded that readiness to perform in a specific domain is a reflection of learners' learning potential and transfer efficiencies (which are stressed as an indication of learning potential [Hamers & Resing in Hamers et al., 1993]). Methods of assessment which yield positive results are those in which the testing situation is concerned with "…structured intervention involving complex social interaction" (*ibid*., p. 89). Essentially, the assessment procedure starts with an evaluation of a learner's general competence, which is then followed by collaborative interactive work sessions ("a mini-learning environment") where learners complete tasks on their own and; if help is needed, a series of hints are provided by competent adults, initial hints being of a general nature with successive hints becoming more and more select and specific with end hints providing the answer outright. This, the authors contend, allows for a measure of a minimum number of hints to be given and in so doing allows the examiner to assess learning efficiency. The attainment of the learning criterion (when two successive problems are solved without the aid of hints) reveals the level of learning potential.

The most striking difference between this approach and that of Vygotsky's notion of the zone of proximal development is not how much improvement is brought about by the hints but how much aid is

51 Too complex and detailed for what is concerned here.

52 Campione and Brown along with other colleagues such as Ferrara refer to themselves as the Illinois Group (Brown et al., in Gifford & O' Connor, 1993).

needed to proceed to the next task in order to bring on a specified amount of learning (Campione & Brown, in Lidz, 1987). In this regard Haywood and Wigenfeld (1992) state that the Campione and Brown approach is grounded more in theory than in practice and that the emphasis in Campione and Brown's approach is domain-specificity[53] varying between the clinical and standardised approach (Zolezzi, 1995). The transfer tasks are also provided with hints and this measure of hints serves as a dynamic manner of assessment. In other words, one is able to assess change based on the number of hints required during initial testing and then to assess the number of hints required after this session is complete. The authors are not interested in the nature of the tasks needed to be solved, but really with the way in which testees deal with related sections of the overall problems.

This type of assessment is further characterised by the nature of the task as opposed to being child (learner) centred. Hints progressively yield more information until an answer is given and as such are standardised and not individually attuned and are fixed in a specific sequence. This at least provides for measures of reliability so allowing psychometric measures entry into dynamic assessment. Although the Campione and Brown have stated their attempts at modifying some of their existing procedures, they acknowledge that a "trade-off" has occurred, in terms of sacrificing richer information for psychometric defensibility. In their varied studies, these authors have not discounted the influence of personality factors and have indicated that testees of different ages relate to hints in different ways, younger testees welcoming hints while older testees attempt to avoid them. In studies conducted by Delclos et al*.,* (in Haywood & Tzuriel, 1992), it was found that the choice of graduated prompting versus mediation depended very much on the purpose of the assessment, with mediation being suited to discovering information about different methods of instruction for different children on an individual basis, whereas graduated prompting might be better suited for classification purposes.

2.9.5 The continuum of assessment services

Another approach to dynamic assessment, not unique yet original as it combines graduated prompting with mediation, is the continuum of assessment model (Jitendra & Kameenui, 1993). It incorporates the assessment procedures of Feuerstein with those of Campione and Brown. In this regard, dynamic assessment can be seen to be a flexible and integrated manner of testing, continually changing and adapting to new circumstances. The model involves screening learners by individually administering an intelligence test and those whose scores are average or above do not receive assessment. Those who evidence IQ scores lower than one standard deviation below average are then given graduated prompting during testing. Graduated promoting is used as this approach more closely parallels what occurs in a classroom learning situation. Learners who perform poorly on the static measures yet evidence high scores in the dynamic measures may reflect poor learning conditions or lack of an enriched environment (Vye, Burns, Delclos & Bransford, in Lidz, 1987).

Graduated prompting is also used because of the standardised nature of the process of prompting, it being properly scripted (which means that an administrator need not be a professional). In addition, learning speed can be ascertained in this manner and it is also a more cost-effective method of assessment. Children who are responsive to the prompting and score above criterion levels are given no more assessment, yet those falling below the criterion scores are administered mediation, in keeping with the tenets advocated by Feuerstein. The purpose of the mediation is to find out where the fundamental cognitive deficits lie and how best to provide instruction to the child. The mediation, although not as extended as the prescribed manner of assessment (Feuerstein, 1979), is used in an abbreviated form. Three tasks are carried out within the mediation attempts, namely, familiarisation of tasks, instruction on rules and feedback. Vye et al*.,* (in Lidz, 1987) maintain that the consistency of results obtained from their measures and the results obtained from static measures over a period of

53 Guthke and Wingenfeld (in Haywood & Tzuriel, 1992) state that the issue of domain-specificity as measured by learning potential tests has been debated. They pose the question of whether learning potential assessment can assess for domain-specific attributes. There has only in the recent past (last decade) been a move away from the emphasis on general abilities and a move towards specific abilities and the testing of these abilities. They state that domain-specific tests will predict learning success better so than domain-unspecific tests.

time lead to the conclusion that static-dynamic correlations are stable and thus point to the reliability of both types of assessment. Both static and dynamic measures yield different predictors of success, yet they do so in a very consistent fashion. The two-fold nature of this model allows for different emphasis to be placed on various factors. For instance, Vye et al, maintain that the graduated prompts are a "…procedure by virtue of its scripted nature [which] seems ideally suited for prediction….[whereas] [t]he mediation procedure…may prove to be better suited as a diagnostic-prescriptive device in view of what appears to be a relatively greater emphasis on contingency in instruction and metacognitive skills" (in Lidz, 1987, p. 340).

This particular approach is an example of how successfully different dynamic assessment methods can be used with ease. Perhaps it is advantageous in that it highlights and makes use of and combines positive aspects expressed in different approaches and in so doing minimises the disadvantages of the one with the advantages of the other. The field of dynamic assessment is itself dynamic and ever-changing, with clear directions to follow yet flexible enough to withstand amalgamations of various approaches.

2.9.6 Guthke and the learning ability test concept

Three types of tests developed by Guthke can be classified according to the type of training used as well as the duration of the testing procedure:

- The "classical" test in which a pre-test followed by standardised training and a posttest are administered over a period of several days. Training can be individualised, group administered or programmed instruction can be given
- A shortened version of the above, in which the testing procedure takes place in one test session with training administered during the test. Assistance is given to the learner and the effects of this assistance are measured by answers on parallel items. Once the criterion stage is reached (when the learner is unable to proceed with the tasks) a score is obtained
- The interval test which is characterised by long-term training. However, this is time-consuming and it is also difficult to pry apart the effects of training and other classroom learning activities.

In essence, Guthke's approach is characterised by a pre-test, training and posttest methodology. Attention is given to impulse control and feedback (Hamers & Resing in Hamers et al., 1993). Guthke (in Hamers et al., 1993b) states as his and his colleagues' main intention with the learning ability test, the need to "…combine the advantages of assessment during a training phase with the advantages of psychometric models" (p. 43). A blend of Feuersteinian emphasis on individual learning and psychometric comparability underlie the rationale behind these tests and they are seen to be more theoretically than practically based (Haywood & Wigenfeld, 1992). Guthke's work has also grown from the theory of activity which itself is a flowering and continued development and criticism of the work of Vygotsky (Guthke, 1982). Learning itself - not only the results of learning - is emphasised in this theory of activity (Guthke in Hamers et al., 1993b). The main findings from the three types of learning potential tests mentioned above, include:

- Increased performance on tests which is evident after training has taken place
- Gains made during the training phase which cannot be predicted sufficiently well from the pre-test score (underlying this finding is the notion that static tests cannot assess potential, as the pre-test functions as a static test)
- Valid and reliable static tests which were used with a training phase but posttest (learning potential) validity increased significantly when the administration took place within small adaptive teaching groups
- Learning potential tests which yielded a higher factor validity than static measures, were less sensitive to environmental factors, discriminated less between different school classes in the same grade than static tests and correlated higher with creativity.

In Guthke's early research, the main concerns were with issues such as the problems associated with validity and measuring change in learning potential as well as the theoretical foundation of learning potential. However, the future of intelligence research will, according to Guthke, rely on the combination of test construction, task and process analysis based on research principles within the

learning potential approach. Future diagnostic tests will follow the format of a short-term learning potential test guided by the theory of item construction with key features such as:

- Content validity, which unlike classical test theory not only looks at the indices of item difficulty but also at the cognitive operations needed to solve the items
- A structured process of tasks, from less complex to more complex task demands
- Learning that is brought about during the testing procedure and not in a testing phase, with a subsequent build-up of feedback
- Factors that determine the course of the learning process.

(Guthke in Hamers et al., 1993b)

A number of tests which investigate domain-specific tasks have been developed by Hamers and Ruijssenaars and have been inspired by the work of Guthke. Different forms of training are used within the tests such as positive feedback and strategies. The training sessions are standardised and brief, but qualitative behaviour analyses is also conducted on the learners.

2.9.7 Curriculum-based dynamic assessment

The curriculum-based dynamic assessment approach proposed by Lidz (1991) (cited in Jitendra & Kameenui, 1993) appears to provide an intermediate solution to the some of the limitations addressed thus far. The assessment includes the dimensions of describing the learner, the examiner, and the task. The approach incorporates the test-teach-retest format and the mediational approach to the intervention phase. In addition, although it includes tasks selected from the learner's curriculum, the approach differs from diagnostic teaching in its focus on cognitive processing (Jitendra & Kameenui, 1993, General guidelines for special educators and diagnosticians, para. 5). Furthermore Jitendra and Kameenui highlight the need for dynamic assessment to be closely aligned with instruction and instructional goals. "To improve student outcomes, a supportive framework for effective dynamic assessment is needed that is guided by the alignment of performance objectives, curriculum activities and test items" (*ibid*., para. 4).

This approach emphasises learning as contingent upon direct experience and is focused on domain-specific tasks, as the needs of the curriculum require; and it has been shown that teachers rank performances as generally higher when using this approach as opposed to making use of static assessment tools (Lidz, 1991 in Jitendra & Kameenui, 1993). Although they do not term interactive assessment as curriculum-based, Brown et al*.,* (in Gifford & O' Connor, 1993) delineate the differences between interactive assessment and traditional assessment with new perspectives concentrating on the following

- Generality versus domain specificity - in which learners are seen to exhibit different propensities towards different areas of cognition and are not considered to possess a general ("g") intelligence indicative of all forms of intelligence
- Stability versus malleability - in which intelligence and learning are seen to change across the life-span
- Isolated versus socially imbedded learning - which perhaps has more prominence within dynamic assessment, since the intervention process is so strongly advocated as integral to dynamic assessment. This view emphasises the need for teachers and students to work in co-operation and the need for guided learning to take place as opposed to instructional learning
- Prediction versus prescription - commonly referred to when differences between static and dynamic assessment are being discussed. Static measures seek to predict future success whereas dynamic assessment seeks to prescribe and aid in remediation efforts which are based on prior assessments. In this regard interactive assessment is proactive and not reactive.

Haywood, Tzuriel and Vaught (in Haywood & Tzuriel, 1992) state that curriculum-based assessment is a response to the dissatisfaction with traditional modes of student performance and that it is consistent with the conceptual base of dynamic assessment due to the following six aspects which they have in common;

- Learning is viewed as modifiable
- Performance is not necessarily understood to indicate potential
- Emphasis is placed on the modification of the processes of learning

- Results are compared with those in the same sample[54]
- Intervention is part of the assessment.

The two are thus suited and should be used in a open-system (such as the new Outcomes Based Education system) within the classroom.

2.10 Assessment versus test

More often than not the terminology used to describe dynamic assessment reflects a move away from psychometric terminology, but this is not always the case because of practical issues. "Assessment" is usually tied to "dynamic" as "test" is to "static". However, the two are often found to be interchangeable. Haywood, Tzuriel and Vaught (in Haywood & Tzuriel, 1992) use the word "interactive" interchangeably with "dynamic", which follows from the transactional nature of dynamic assessment.

2.11 Dynamic assessment and learning potential

Typically "[m]easurement of learning potential generally focuses on disadvantaged groups" (De Beer, 2000, p. 26). According to De Beer (2000), "[d]ynamic cognitive assessment, also known as learning potential measurement, is based on the view that cognitive processes are highly modifiable" (p. 72). Here, De Beer asserts that the two are synonymous and uses the terms interchangeably. "Learning potential is **what** is measured, while dynamic assessment is the **way** in which it is measured" [own emphasis] (p. 76). Furthermore, she states that the defining feature of dynamic assessment is the test-teach-test format, and although there are a myriad of approaches available, most are based on interpretations of Vygotsky in some or other way. Boeyens (1989a) takes issue with the criticisms levelled at learning potential tests enumerating five principles that should be adhered to within this field if it is to be at all implementable and practical, namely,

1. It is to be cost and time efficient
2. It should be based on a methodology that differs from clinical approaches
3. Items are to be drawn from an homogeneous pool
4. The pre- and posttests should be "analogous" so that improvement from pre- to posttest can be assessed with accuracy
5. The pre- and posttest scores should be highly reliable so as to ensure the reliability of the difference score.

These criteria are applicable to both learning potential and dynamic assessment. Learning potential usually forms part of test instruments and measures are referred to as learning potential assessment devices, housed within the paradigm of dynamic assessment. Nevertheless, the two terms, although slightly different, are usually classed as synonymous. For the purposes of this study, however, dynamic assessment is accorded the status of paradigm, while learning potential concerns the instruments used within the approach.

2.12 Issues at stake within the field of dynamic assessment

Jitendra and Kameenui (1993) sum up the disadvantageous issues evidenced within the dynamic approach as follows

- Construct fuzziness (impinging on issues of validity)
- Procedural spuriousness (impinging on issues of reliability)
- Instructional aloofness (impinging on issues reliability)
- Instrument inadequacy (impinging on issues validity)
- Labour intensiveness (an all pervasive negative factor in dynamic assessment research).

54 This, then, cannot be said to be either criterion nor norm-referenced.

There are a number of issues in this field which merit attention, but a few are more contentious than others and warrant further discussion namely, reliability and validity, differential item functioning and item response theory, costs and timing and difference (change) scores. These issues are also posed in the literature pertaining to this field and so receive brief attention below.

2.12.1 Reliability and validity

Referring to the Teach-Test-Teach (TTT) programme run at the university of Natal from 1986 - 1995, Craig states that the two different goals of psychometric and dynamic assessment should be noted: "…a direct comparison would be odd. Reliability and validity issues have, however, been tracked … for purposes of refining our interventions and tracking student performances over their careers" (personal communication, March 6, 2001). Hence, even within the domain of a strictly dynamic set-up, the costs of not assessing validity and reliability within these two frameworks far outweigh the benefits derived from such endeavours.

In essence, these somewhat contentious issues are necessary but not always deemed as forming part of the paradigm from which dynamic assessment springs. This is quite a complex issue, which has in part been attended to in the work of De Beer (2000), who manages to combine the needs of the one school of thought with those of an apparently opposed school of thought successfully. Whether or not purists from either camp (psychometric and dynamic) would agree on the methods and techniques used in the above-mentioned research remains to be seen. A compromise, it seems, may have to be agreed upon. "Using it [dynamic assessment] ought to be dictated by the problem at hand. Psychometric tests have many advantages, but **cannot** assess/measure that which has, as yet, **not** developed/matured/been learned. Dynamic assessment does not work in practice when it is used without grasping what it is about, i.e. introducing a **process of learning for monitoring/assessment**" [own emphasis] (A. Craig, personal communication, March 6, 2001). This is perhaps the key "ingredient" when trying to define dynamic assessment, which is not merely a process in which a pre-test session is followed by a posttest session, with some sort of mediation taking place in the middle.

Sternberg (1991), whilst referring to Howard Gardner's systems approach to intelligence (the so-called multiple intelligences), states that "[t]hose who are advocating this type of approach need to demonstrate the psychometric soundness of their instruments" (p. 266). One cannot help applying this sentiment to dynamic assessment as well, a sentiment which has already been voiced on numerous occasions, least of all by Sternberg himself (Grigorenko & Sternberg, 1998). The critique levelled at psychometrics is that "…attempts to measure stable performance are bound to produce spurious information in a context where education deliberately intervenes to change individuals' levels of preparedness" (Griesel, 2000, p.6). Griesel also asserts that what may be happening to static measurement is akin to a paradigm shift, a shift which is necessitated by the current paradigm's ineffectiveness in dealing with the problem on hand. This shift entails a move away from static once-off testing to one aligned with processes and continued learning which themselves change the learner.

Lidz (in Flanagan et al., 1997) delineates the varied reasons why an unease exists with practitioners advocating a more psychometric manner of testing, and foremost among them are reliability and validity. Test-retest reliability is not applicable for instance when working with dynamic assessment, since by implication stability of test scores are an indication of stable intelligence scores, a feature of static measurement which runs counter to the notion of change which is what dynamic assessment seeks to capture. Test-retest reliability will thus be low. Interobserver agreement would perhaps be the best alternative to the traditional measure of reliability. Inventing new ways of dealing with these thorny issues, states Lidz, is quite a challenge. "…the further that validities on standard psychometric measures are pushed up, the more stable are the characteristics measured and the less relevant they become for aiding the search for ways of meaningful intervention" (Feuerstein, et al., 1979, p. 24) This in essence, negates the point of dynamically assessing a person. Griesel quotes Feuerstein (1981) as stating that "[t]he notion of reliability or accuracy thus rests on very different assumptions - and relates to the extent to which the **education process intervenes to change the individual's performance**" [own emphasis] (in Griesel, 2000, p.8). Lipson (1992) states an all-too-often quoted criticism levelled at dynamic assessment, that the construct validity of instruments such as Feuerstein's LPAD's interventions serve to modify scores and can be regarded as coaching. Savell et al. (1986) note the difficulty in evaluating the alleged claims of dynamic assessment techniques, because in most cases

these techniques have not been empirically tested or authenticated by researchers other than the ones who originally developed the techniques (in Jitendra & Kameenui, 1993). If dynamic assessment is to take hold within established institutions in South Africa the issues of reliability and validity need to be taken very seriously. Differential item functioning and item response theory are partial solutions to the problem.

2.12.2 Differential item functioning and item response theory - partial solutions to the questions of reliability and validity

Differential item functioning (DIF) is used to detect bias when comparing similar ability levels across different cultural and/or language groups. It is of particular importance that ability levels are similar, as low and high performers will automatically achieve differing results, which have no bearing on their respective cultures and/or languages (Kanjee in Foxcroft & Roodt, 2001). An item is said to be biased if individuals who are of the same ability but from different groups (culture and/or language) exhibit widely disparate results when the probabilities state that they should score comparably. Item response theory is one method of investigating DIF.

As the Employment Equity Act (see 2.13) states that tests developed and implemented in South Africa have to be shown to be both reliable and valid, dynamic test instruments need to comply even if these criteria are not relevant to dynamic assessment. In order to legitimize dynamic assessment, the solution has been to integrate measurement theory concepts with those of dynamic assessment. Item response theory (IRT) is one such partial solution and has been referred to as initiating a "quiet revolution … in test theory" (Embretson & Reise, 2000). IRT is uniquely placed to solve a number of issues within dynamic assessment (see section 4.3.19 for a locally developed dynamic assessment test which makes use of IRT in order to assess the amount of bias [differential item functioning] in the test). Kanjee (in Foxcroft & Roodt, 2001) lists five distinct advantages for the use of IRT in tests (not specifically referring to dynamic assessment, however):

1. The availability of information at the item level which makes it possible to select items with high discrimination values
2. Sample invariance, referring to the fact that both test-taker and item are calculated independently from the sample from whom the data was obtained (the "true" value of an item remains consistent within differing samples)
3. Ability scores are independent of items chosen, as including or excluding items does not necessitate the recalculation of the entire measure
4. IRT has been said to have found its niche in the computerised age more so than in any previous age due to the advent of computer adaptive testing allowing scores to be made available instantly
5. Graphical representation of information makes for easier analyses.

A limitation to the use of IRT is the large sample sized needed to complete such studies as well as the complex mathematical models used to derive specific formulations and answers.[55] Nevertheless IRT is construed as an " …advanced test theory approach that can satisfactorily handle many of the instrument problems that are dealt with more primitively by CTT [classical test theory]" (Sijtsma in Hamers et al., 1993a, p.117; cf. Sijtsma in Hamers et al., 1993b). Sijtsma takes the reader through the main features of CTT highlighting the role played by reliability and validity, with validity playing a more important role than reliability but that test construction's main aim is item reliability which is usually higher when inter-item correlations are high and if there are a sufficient number of items. Perhaps the most important aspect in CTT is investigating the degree to which test performance is affected by error sources, thus the two types of scores, error and true scores. Sources of error may be evident across independent replications of a test and it is the task of CTT "…to determine the degree to which a test result obtained by a specific testee can be repeated" (*ibid*.).

IRT is illustrated by the item characteristic curve (ICC), which is itself a reflection of the probability of a testee giving a correct answer to an item which is a function of both the testee's ability and characteristic of the item (Ittenbach et al., in Flanagan et al., 1997). Sijtsma points out two properties

55 Software packages are, however, now used which greatly aid the researcher making use of IRT.

of the ICC; firstly, the higher the ability, the higher the probability of success on the item that measures the ability and secondly, the ICC curve remains relatively constant for low and high abilities and in these areas differences in abilities do not affect probabilities of success, yet it is between these low and high ability areas that even slight changes in ability strongly affect success probabilities. It is this area that, if reached by the testee, is said to indicate mastery of the item at hand (Sijtsma in Hamers et al., 1993a). The ICC is described according to three defining characteristics:[56]

- A pseudo-chance level - which describes the probability of success for extremely low ability level individuals (which for multiple choice items, for instance, is positive but for open-ended questions the probability level tends towards zero)
- Item difficulty - which indicates how the ICC is located on the ability scale; the higher the level of ability needed to answer an item correctly, the higher the location of the parameter. An ICC parameter location to the left would indicate higher chances of success for a range of abilities and also would indicate that the item is easier
- Discrimination power - which is related to the slope of the ICC. The steeper the slope the better the discrimination between high and low ability level individuals.

The development of tests using IRT is more complex (owing to the nature of the mathematics involved) than tests developed with CTT. IRT involves longer periods of time for data analysis and is restrictive in the number of items that may be rejected from a test battery. Apart from these disadvantages, other advantages of IRT include its manner of scaling, as it ties scores to the metric of the underlying trait, provides a standard error which accompanies each estimate of item difficulty and the proficiency of the testee. Standard error is determined by assessing the variability in a testee's behaviour, variation introduced by the examiner and variability within items (Ittenbach et al., in Flanagan et al., 1997). Embretson and Reise (2000) assess the "rules" of measurement and refer to the old (but hardly defunct) rules of measurement and some new rules of measurement (made available through IRT) and can be briefly summarised as follows

- The standard error of measurement not only differs across scores but can be generalised across populations
- Shorter tests can be more reliable
- When test difficulty levels differ between individuals, test scores across multiple forms can still be compared
- Unrepresentative samples can now be used without the worry of bias entering the equation
- Change scores can be compared in a meaningful way even if initial scores differ
- Item features can be reconciled with psychometric properties.

These rules evidence the different set of principles to which IRT adheres and also show that many CTT principles are no longer necessary.

Alternative measures of reliability and validity are investigated in the educational domain and although not specifically referring to dynamic assessment *per se*, Gipps's (1994) ideas are seen to be aligned with those of dynamic assessment. Gipps states that reliability and validity are still important dimensions of test development "…acting essentially as quality assurance devices…but for educational assessment…we need other indicators of quality…" (p. 172). She lists these other indicators:

- Curriculum fidelity - in which the curriculum is well specified and covered broadly
- Comparability - in which the approach to assessment is consistent and the criteria of assessment are understood by all, with the same rubric used by all testers
- Dependability - which emanates from curriculum fidelity
- Public credibility[57] - in which the public needs to be assured that testing remains comparable (from one assessor to the next) and consistent (from one test to the next)

56 See section 4.3.19.1 in which De Beer (2000) makes use of IRT in her South African dynamic assessment instrument.

57 Perhaps of particular importance in South Africa, where public apathy towards testing in the past resulted in dissatisfaction with testing, but has since spurred the testing movement forward. This issue of public credibility is not often emphasised, as it may be mistakenly assumed that the broader public have no knowledge or interest in the matter. It is the public, however, that needs to be informed.

- Context description - in order for transfers from one context to another be carried out, it is imperative that information about contexts is made available
- Equity - which allows multiple opportunities for students in an assessment programme.

What Gipps has tried to show is that alternative forms of reliability and validity are available and that assessment need not be housed under the psychometric umbrella alone:

> If we move teacher assessment and performance assessment too closely towards standardization in order to satisfy traditional reliability, we are in danger of throwing out the baby with the bath water. **The search for objectivity in assessment is not only futile but can be destructive**; the alternative criteria proposed will allow us to ensure quality in educational assessment without resorting to highly standardized and narrow testing procedures with all that this implies for teaching and learning…we need to stop supporting the notion of assessment as an objective activity or exact science [own emphasis] (Gipps, 1994, pp. 174 - 175).

2.12.3 Costs and timing

The most often cited criticism levelled at dynamic assessment is the time taken to administer the assessment as well as the costs involved in doing so (Boeyens, 1989a; Coosner 1999; Vye et al*.,* in Lidz, 1987). South Africa can ill-afford a costly and timely endeavour into assessment, yet this need not be the case as has been evidenced by the studies of Coosner (1999) and De Beer (2000). A compromise, it would seem, would be the best solution to this dilemma. On the one hand, the tenets of dynamic assessment preclude a so-called quick-fix answer, yet one cannot ignore the practicalities of reality. "The need is even more urgent for a dynamic, affordable, **educator-friendly** assessment strategy for educators" [own emphasis] (Coosner, 1999, p.11). Flexible use of interventions and approaches, such as evidenced by the continuum of assessment model, standardised test-teach-test models are efforts at curbing costs and timing and it is for this reason that most approaches have come into existence. The TTT programme, as will be discussed in Chapter 4, was one initiative that was disbanded partially owing to costs and timing. Similar efforts at cutting time and costs involve the administration of the LPAD in group format, for instance (Andrews, 1996).

2.12.4 Difference scores

Early South African research in this field welcomed the new and innovative ways of conducting dynamic assessment. However, the main concern was and still is the extent to which dynamic assessment can withstand scientific scrutiny. The difference score between pre- and posttest assessment is a typical point in case. Differences scores cannot simply be interpreted as a indicating "potential", for this in itself eschews the ideals of dynamic assessment (when taken in its purist form, that of extensive individual mediation and re-mediation). As early as 1970 (and possibly prior to that time) Cronbach and Furby (1970) recommended a number of alternatives towards solving the issue of change measurement without the need to estimate the change scores for individuals.

During the 1950s, international research relied almost exclusively on what was referred to as "gain scores". Lidz (1987) cites Vernon (1954) as identifying some problems inherent in looking only at the pre-post test score differences, "…for example, the large standard deviations that led to unreliability for small groups (under 100), and the lack of uniformity of measuring units across research projects" (p. 8).

Boeyens (1989a) states, with regard to this issue, that tests which profess to test a variety of domains (unlike domain-specific tasks) will have lower reliabilities and thus higher measures of error. "The reliability of a difference score is reduced by the error in the pre-test and the posttest scores …the reliability of the difference scores is always lower than the reliability of the pre-test and the posttest scores" (p. 38). In essence, even though both the pre- and posttest scores may have acceptable levels of reliability, this in no way attests to the high reliability of the difference score. This cumulative effect of error measurement most assuredly has an effect on the difference score. Early studies merely subtracted pre-test from posttest score and one has to consider the appropriateness of these actions. Of note is Boeyens' firm comment on the matter of reliability and the manner in which it is

treated by researchers in the field. This criticism is one that needs to be looked at and not merely to be paid lip-service.

Embretson (in Lidz, 1987) explores five types of dynamic scoring which reflect different approaches by dynamic assessment, namely,

1. Simple gain - which is simply the posttest score minus the pre-test score.[58] However, the greater the pre- and posttest measures of error are the greater the error of the gain score and the fact that gain scores do not always correlate with initial status (as stated above). This is where the problem of the ceiling effect emerges, with lower pre-test scorers "outperforming" high pre-test scorers in the posttest merely due to their larger gain scores
2. Residualised gain scores - in which simply put, predictive validity of scores is not dependent on pre-test scores since one is dealing with linear regression coefficients for both the pre- and posttest results and gain on the posttest is not predicted from results on the pre-test, hence gain is not linearly predicted from the pre-test. This is said to have the fewest disadvantages but still does not resolve the issues of reliability and inequality of the score units used in the pre- and posttests
3. Posttest scores - the rationale here is that if it is change that is to be measured, surely the pre-test score is irrelevant due to its static nature, since nothing has yet changed. However, considerations such as the constancy of traits measured between the pre- and posttest, norms remaining the same from pre- to posttest assessments and prediction of scores being valid after intervention are still thorny issues to be resolved with this type of scoring
4. Scores reflecting both initial performance and change - here simple gain (as in the first approach) is added to the initial score on the pre-test. However this becomes viable as a method only if the simple gain is added to the initial score in a multiple regression equation to predict criterion variables[59]
5. The course of the intervention - in which the intervention process itself is scored (such as quantifying the number of hints given to a testee as used by Campione & Brown). The extent to which the number of hints provide a different index to that of the pre-test score arises.

These varied approaches all provide distinct advantages and a myriad of disadvantages. Classical test theory on the other hand encounters three difficulties with the change scores, namely,

- If differences between individuals are to be assessed, it must be shown that the raw scores are represented on an interval level so that incremental increases across individuals are comparable to increases in ability
- Classical test theory provides no means of investigating gain scores if the norms for both pre- and posttest are continually changing. Change can only be measured if a degree of correlation exists between the two tests and if changes are evident from pre- to posttest then norms themselves are shifting and hence there is no way to interpret change scores
- Unreliability of error measurement makes it such that different score levels cannot be interpreted correctly, because classical test theory assumes that error measurement is constant across score levels. This has proven not to be the case where extreme items (either very easy or difficult) are measured with less reliability than normally distributed items and this presents a problem "…for gain scores because the extreme levels can be expected to **change more between alternative forms** solely on the basis of measurement error arising from too few appropriate items" [own emphasis] (Embretson in Lidz, 1987, p.155).

2.13 Assessment in South Africa - how far have we come?

Psychological assessment in South Africa has come a long way since the first assessments were carried out in the 1920s and with the new dispensation, assessment is now under the spotlight

58 Simple gain is a type of measure that was employed very often in the early dynamic assessment research in South Africa up to early 1990s.

59 Embretson adds another variation to this approach, namely using the predicted posttest score along with the residualised gain (second approach) and their sum being the posttest score (in Lidz, 1987).

regarding biases of many kinds. As with many aspects of research and development in South Africa, often reliance on non-South African ways of doing things has led to dissatisfaction with end-results, notably within the field of psychometric testing. As Huysamen (1980) states "[i]n many respects South African tests and testing practices may be regarded as the direct offspring of corresponding developments in the United States of America, Britain and Western Europe…some [are] simply standardized adaptations of overseas tests" (p. 9; Foxcroft & Roodt, 2001), yet he acknowledges the novel tests that have in the past been devised for use on illiterate and "unacculturated"[60] testees. Of interest is that group tests were administered in South African before individually administered tests, owing to the popularity of group tests in the United States and several researchers became engaged in the development of group tests for school learners during the 1920s. However, it was not until 1930 that the South African Group Test of Intelligence was published and group testing flourished.

During and after the second World-War Dr Simon Biesheuvel was appointed as Commanding Officer of the Aptitude Testing Section of the South African Air Force and it was owing to the success of this unit that the National Institute for Personnel Research was established (NIPR). It was here too that Biesheuvel devised a number of classification tests for all race groups (Huysamen, 1980).

Foxcroft, Roodt and Abrahams (in Foxcroft & Roodt, 2001) delineate three major trends that characterised early developments in psychological assessment in South Africa, namely,

- The fact that tests were only standardised on and for whites
- The use of biased norms on groups other than those for whom the test was intended
- Hasty and biased conclusions reached about test results which served to perpetuate the cycle of misuse of these test results that sought to reaffirm skewed notions of certain racial groups.

This situation was further compounded by the political issues of the day, a time in which job reservation and testing of illiterate people with unstandardised test instruments was the norm and by fulfilling this agenda many people were most likely misclassified to their discredit. The Human Sciences Research Council (HSRC) was the result of an amalgamation between the National Institute of Personnel Research (NIPR) and the Institute for Psychological and Edumetric Research (IPER).[61] Unfortunately today many tools of assessment are still used, tools which make use of norms not appropriate for the sample which is being tested and as Foxcroft et al*.,* in Foxcroft and Roodt (2001) note, statements cautioning against the interpretation of results when such tools are used merely served to "[ease] practitioners' consciences and lulled them into a sense that they were doing the best they could with the few tools at their disposal" (p. 26). During a stint of research conducted in the late 1980s and early 1990s, various researchers concluded that many tests were unsuitable for use with groups other than the groups on whom they were normed and subsequently a growing anti-test movement swept across South Africa and it has been stipulated that tests are prohibited unless they adhere to the following three criteria:

- The test must be scientifically proven to be a valid and reliable test[62]
- It must be applied fairly to all employees
- A test may not be biased to any group (race or gender).

(Adapted from the Employment Equity Act No 55 of 1998 [Section 8] Foxcroft et al*.,* in Foxcroft & Roodt, 2001.)

The anti-test movement may have resulted in a positive renewed look at assessment in South Africa. One of the later developments to take place within psychological assessment in South Africa was the

60 One must take cognizance of the time in which Huysamen was writing this and the issue as to who exactly is unacculturated is a thorny one, but one can assume that he is referring to the previously disadvantaged population.

61 Measures to do with education and clinical practice were developed by this institution as opposed to the NIPR which concentrated on measures used in industry (Foxcroft & Roodt, 2001).

62 Which does seem to pose problems for dynamic assessment, as a Feuersteinian and Vygotskian approach for instance, does not align itself with classical test theory concepts. Therefore, other alternatives have to be found or a new form of dynamic reliability/validity has to be developed.

development and use of dynamic assessment instruments, both developed and normed locally as well as taken from overseas. It must be noted that although dynamic assessment lends itself more fully to an unbiased approach, in its non-verbal status and reliance on mediation, this does not necessarily mean that overseas dynamic instruments are immediately useful in the South African context. As is shown later in this study (for instance, see section 4.3.11.3) certain items from the LPAD have to be rewritten for use locally. This survey deals with this local dynamic assessment research, how it originated and grew and the status it has achieved today. Local dynamic assessment research was implemented in the early 1980s, followed by a few seminal works (Boeyens, 1989a, 1989b; Gaydon, 1988; Henley, 1989; Hoffenberg, 1988; Murray, 1988; Shochet, 1986) in the form of dissertations in the latter half of the 1980s. A trickling of research in this field emanated from a variety of institutions, with the bulk of research emanating from a select few institutions. The context of these studies was invariably situated within tertiary academic selection and admission programmes, which aimed at assessing and selecting potentially low-risk students. Other research includes assessment of previously disadvantaged learners within the secondary school context and special education programmes. Some research has been conducted in other contexts in South Africa. The path of progress followed by dynamic assessment research is not an exponential one, it being characterised by steady and fast-paced growth, followed intermittently by periods of little research. In the past few years, research efforts have waned but this is not necessarily a reflection of the whole picture.

The need for future dynamic assessment research speaks for itself. As both the advantages and disadvantages have been highlighted, it is more accurate to state that the need for dynamic assessment research far outweighs any negative side-effects that it may yield. This statement can be further supported when the data is investigated in Chapter 4. The method of data capture will now be explored.

CHAPTER 3 Methodology

"Dynamic assessment to a significant extent represents a clinician's dream and a psychometrician's nightmare."

- Carol Lidz

3.1 Introduction

This research study collates information that has been collected from literature sources and, where necessary, additional information gleaned from informal interviews conducted both telephonically, and electronically. It is deemed necessary to elucidate exactly how the information was obtained, since this has been the main method of data collection.

The manner in which the information was sought and compiled was systematic and as complete an investigation as time and resources would allow. The method of data collection, elaborated below, entailed library searches, literature surveys and information gathering via structured discussions and interviews. The purpose of this study was to ascertain the amount of work either already completed or in the process of being completed in the field of dynamic assessment in South Africa. Further to this and perhaps more importantly is an analysis of the content of the information gathered. Critical appraisals, summaries as well as recommendations based on the gathered information can therefore be construed as the main tasks accruing from the data collated for this study.

Broadly speaking, an attempt was made to answer two main questions namely, how much has been researched in this field and what is the nature of the findings? In order to progress in any field in as scientific a manner as possible, it is imperative to start at the beginning, which in this case is an overview of the literature available. In some of the studies investigated here, the details of how the researcher procured literature sources is not often divulged, as it is often assumed (among others Andrews, 1996; Henley, 1989; Lipson, 1992) that these sources are common knowledge and are often utilised by most researchers in the field.[1] It is surprising how often this assumption is not the case. Naturally, it is not possible to assess all the available databases and information sources that may potentially exist. However, an attempt has been made to treat available material as thoroughly and comprehensively as possible within the given time constraints.

3.2 Topic delineation and preliminary sources gained from the literature study

Upon the choice of topic, namely to survey the status of dynamic assessment in South Africa, the first leg of research encompassed a familiarisation process, after which narrowing of information was sought by means of locating specific sources of information. In reviewing the literature on dynamic assessment in South Africa, it has been found that dynamic assessment research and application comes to the fore in two main areas, namely, dynamic assessment emphasising the testing procedure for either **categorical and selection purposes** (selection and placement of students in tertiary education or placement of employees into various job categories) or, for **remedial purposes**, the former being the main emphasis in this study. The latter direction of research is usually of a more qualitative nature and one such study is surveyed here.[2] However, owing to the overwhelming number of studies which have emphasised the former attributes, it was decided to follow this main stream approach.

1 Of course this study emphasises the literature gathered and the manner in which it is gathered, an emphasis which is not necessarily as important in other works.

2 See Coosner (1999) section 4.3.17.

3.2.1 Databases

3.2.1.1 International Databases

The nature of the research study is such that both the **ERIC**[3] (The Educational Resources Information Centre) and **PSYCINFO** (previously known as PSYCLIT[4]) databases were searched. The overlap in sources was evident but not extensive, owing to the nature of the study (working within the realms of education and psychology). Interestingly, although access to the latest sources on PSYCINFO was not available to the researcher at the start of the research process, the ERIC database yielded a substantial amount of information, information that when searched for at a later date on the PSYCINFO database, did not yield that much more by way of new research. This may be an indication of how quickly the field is progressing.

One must not forget to note, that, as with all research one is at the mercy of varied uncontrollable factors, one being (in this particular instance) the select criteria which different databases impose on different research articles' potential submission to their bibliographic indexing. A cursory glance at most topics housed within these various databases yields a very small percentage of South African literature "hits" (the number of times a record appears after a query is submitted). This is understandable to an extent, as South Africa does not publish as voluminous an output in the field as do the United States and the United Kingdom.

Nevertheless, ERIC, PSYCINFO as well as **EBSCO**[5] were used to find information on the topic, both locally as well as overseas. EBSCO makes available the full text version of articles found in accredited journals. Undoubtedly there are numerous other databases available, but the nature of information needed was of a local and not international nature. These databases are all available at the University of Pretoria's Academic Information Service.

3.2.1.2 National Databases

Local South African databases played a more prominent role in finding literature sources. The National Research Foundation's (NRF) various databases (such as **Nexus**[6]) enabled the researcher to trace the most recent theses and studies either completed or current, and which were, or are in part, sponsored by this organisation.

It is apt at this point to note that it is entirely possible (and in fact occurred on no less than three occasions) that although research is being conducted, failure to submit theses and reports to respective universities does occur, consequently not allowing for any indexing of the research to take place. Also, if research is being conducted but not being funded, it is likely that it will not be disseminated, a fact which militates against the purpose of scientific research.

---

3 "ERIC is a national information system designed to provide ready access to an extensive body of education-related literature. Established in 1966, ERIC is supported by the U.S. Department of Education's Office of Educational Research and Improvement and is administered by the National Library of Education (NLE)" (www.eric.ed.gov/about/about.htm).

4 "The main PsycINFO database contains more than 1.5 million references to psychological literature from 1887–present, from journal articles, books, book chapters, technical reports, and dissertations" (www.apa.org/psycinfo/about).

5 "…a collection of specially designed, comprehensive full text databases, available online via EBSCO host" (www.epnet.com/index.html).

6 "The NEXUS Database System is a collection of databases focused on the humanities and social sciences in South Africa. It includes a database of: current and completed research projects; research organisations; professional associations; researchers; forthcoming conferences; research methodology courses and teachers; and periodicals submission requirements" (www.hsrc.ac.za/databases.html).

SABINET[7] perhaps played the largest role in facilitating the accumulation of sources. All articles currently available in South Africa are indexed in this database. The Nexus and SABINET information did not always yield the same information, however, no database is without fault but this study's almost exclusive reliance on these databases made it imperative that they function as they are intended to. The Human Sciences Research Council's (HSRC) online library as well as internal catalogue was also made use of on many occasions. In many instances, the various databases complemented one another; in other instances some included sources found nowhere else.

3.2.2 Keywords used in the various database searches

The following keywords were used to search the various databases:

- *dynamic*
- *learning*
- *potential*
- *interactive*
- *assessment*
- *dynamic* AND *assessment*
- *learning* AND *potential*
- *interactive* AND *assessment*
- *dynamic assessment* AND *South Africa*
- *learning potential* AND *South Africa*
- *interactive assessment* AND *South Africa*
- *static* AND *tests*
- *conventional* AND *tests*
- *static* AND *assessment*
- *psychological assessment*
- *disadvantaged* AND *learners* AND *South Africa*
- *psychological assessment* AND *South Africa*
- *Feuerstein*
- *Feuerstein* AND *South Africa*
- *Vygotsky*
- *Vygotsky* AND *South Africa*
- *Teach-Test-Teach*
- *Test-teach-test*
- *LPAD*
- *learning potential assessment device*
- *learning potential assessment device* AND *South Africa*
- *mediated learning experience*
- *cognitive structural modifiability*
- *developmental assessment*

3.2.3 Survey of contacts

As stated earlier, this study's main methodological focus almost entirely rests with bibliographic sources, yet this proved more valuable when complemented with additional information gathered from various institutions.

7 "Sabinet Online enables information users to obtain relevant information according to their specific information needs from … international and South African databases" (www.sabinet.ac.za).

The following universities were initially contacted either via electronic mail or via telephone.[8] Those failing to respond to electronic mail (for whatever reasons) were contacted telephonically, and on many occasions were contacted two or three times. Responses and information gleaned from these informal interviews are dealt with in Chapter 4. The reason for contacting tertiary education institutions was that much of the information gleaned from the literature pointed to the use of dynamic assessment in tertiary education.

The domain of dynamic assessment is not exclusively relegated to higher education, but its field of application (in many other countries the world over) is centred in this area. Almost all of the contact details were located on the internet - University and Technicon web page information - and these details were generally accurate and led to successful contacts. A number of alternative paths were followed in order to gain access to the various web pages, notably South African search engines. It is appropriate at this point to note that although information on the present topic was indeed available on the worldwide web and the internet, peer reviewed journals are often the main source of information. As information on the internet is not always peer reviewed and, if reviewed, entails a cost factor, limited use, as detailed earlier in this chapter, was made of internet sources, despite their usefulness.

3.2.3.1 Technicons contacted

The following Technicons were contacted sucessfully:

- Technicon Orange Free State
- Pretoria Technicon
- Port Elizabeth Technicon
- Natal Technicon
- Witwatersrand Technicon
- Cape Technicon
- Technicon Northern Gauteng
- Peninsula Technicon
- Technicon SA
- Johannesburg College of Education
- Border Technicon
- ML Sultan Technicon
- Mangosutho Technicon
- Vaal Triangle Technicon
- Eastern Cape Technicon

3.2.3.2 Universities contacted

The following Universities were contacted successfully:

- University of Zululand[9]
- University of the North
- University of Durban-Westville
- University of Stellenbosch
- University of Natal
- University of Cape Town

8 In most instances, tertiary institutions did not respond to the electronic mail questions, and thus had to be contacted via telephone. In some instances, both methods were employed at different times and different informants were spoken to (which also allowed the researcher to compare what others had said within the same institution). In a few instances, personal visits were made to enhance the information gathered. Telephone conversations tended to elicit the best information but this was not exclusively the case. Of note is the fact that almost all contact details for the various institutions were located on the various web pages, which were mostly accurate in their contact details.

9 See review of article by Sibaya, Hlongwane, and Makunga (1996), section 4.4.

- University of Port Elizabeth
- Rand Afrikaans University[10]
- University of the Witwatersrand
- University of Pretoria
- University of the Orange Free State
- University of the North West
- University of South Africa
- VISTA University
- University of the Western Cape
- Rhodes University
- Potchefstroom University for Christian Higher Education
- University of the Transkei
- University of Venda
- University of Fort Hare[11]
- Medical University of South Africa

3.3 Consolidation of information gathered

The results of the interviews (whether telephonic, electronic or personal) are discussed in Chapter 4. In all instances, for both Technicons and Universities, the contact persons were those either involved directly or indirectly in the institutions' selections and admissions departments. Follow-up calls were made in certain instances to ensure accuracy of information obtained. In a few instances, the follow-up calls yielded information that was not always similar to the original information gathered. As call-backs to all institutions were not always possible because of cost and time constraints, the information recorded is, of necessity, subjective.

The process of collecting data, whether literature (which was on many occasions suggested by the institutions themselves, information that was not always available on the literature databases), or information gathered from the institutions themselves, continued throughout the research year. It is important to point out that owing to the nature of the research endeavour, waiting for material to become available could also be prohibitive regarding timing and, as such, a cut-off point (date) was established, after which further information was not considered in the study. Details of unsuccessful attempts to gain access to some relevant studies, are given in APPENDIX 1.

Chapter 4, the focal chapter, provides a detailed analysis of the various studies investigated.

10 This study refers to this university as "Rand Afrikaans University" even though it has been seen as "Rand Academic University". At the time of writing the former was the official name of the university.

11 After several call-backs, this was unfortunately the only tertiary institution for whom reliable information could not be sought. It would seem from the numerous contacts that this university does not employ dynamic assessment tests in any form, but the author cannot state this unequivocally nor without any substantial evidence.

CHAPTER 4 South African research into dynamic assessment as alternative or complementary assessment procedures: Results

"IQ begins to look more like a rather arcane ability, at which only peculiar people would (and would wish to) excel, than a crucial quality for living."

- Guy Claxton

4.1 Introduction

This chapter investigates the findings of various studies conducted within the field of dynamic assessment in South Africa, ranging from the early and mid -1980s to the present. This chapter will be divided into three main sections, Part A dealing with Master's dissertations and Doctoral theses sources obtained in the literature search, Part B dealing with findings from published articles and Part C detailing findings from informal interviews with various tertiary institutions.[1] Part C, will also in some instances be complemented with further documented research findings encountered in these informal interviews.

Each study is discussed in its entirety and for the sake of clarity and consistency, the format follows the same structure, according to four main headings:

- **An introduction** - which briefly introduces the study and its main objectives
- **Research design** - which illustrates the design that each researcher used in order to complete the study
- **Main findings** - which details the findings and statistical data that each study has to offer
- **Conclusions and recommendations** - which summarizes the conclusion and recommendations that were reached by the author of the respective studies.

During the course of this chapter, some comment will be delivered where it is warranted and seems pertinent usually by way of footnotes and cross references are indicated for the sake of clarity and ease of reading. However, these comments are not pervasive, but serve to illustrate, where necessary, inconsistencies and anomalous findings in the various texts.

4.2 Main findings

Before the individual studies are presented, brief consideration of confusion of terminology and misinformation about the concept of dynamic assessment, ambiguous use of the word "potential" and the effects on the status of dynamic assessment as well as misperceptions of the role of dynamic assessment in South Africa is given.

4.2.1 Confusion of terminology and misinformation about the concept of dynamic assessment

Based on the findings of the literature survey as well as the informal interviewing conducted, it is clear that "dynamic assessment" deserves careful definition. Although the term is defined in the literature, dynamic assessment, appears to refer to a form of testing rather than to a type of test. There are a number of ways in which dynamic assessment is currently understood. It is -

- A procedure which includes a pre-test and posttest design. The learning or mediation in between is not strictly defined and the straightforward difference between pre- and posttest scores is an indication of "potential"
- A form of assessment which is aimed at disadvantaged students only and is culture-fair, but takes too long and is far too costly to implement

1 This demarcation into three parts is done for ease of reading and in no way reflects quality of work contained in any of the three parts. For the sake of clarity and continuity, numbering will be continuous and will not start anew with each part.

- A form of assessment which merely makes use of conventional testing instruments and implements them in a dynamic manner
- A form of assessment which comprises only mediation and individual attention and is wholly qualitative and thus not suitable for larger samples
- A form of assessment, (consisting of two branches, that of mediation as well as assessment) which emphasises previous deficiencies in learning opportunities, allowing all learners to participate as far as possible on a similar level. The bulk of South African research in this field places emphasis on the assessment aspect, often not considering mediation and/or remediation at all or, if so, paying only scant attention to it.[2]

Often the concept of dynamic assessment is understood, but only after explanations are offered as to its varied meanings. It is unfortunately perceived by some as unsuitable in South Africa at the moment (for instance Engelbrecht, 1999; L. Irvine, personal communication, August 30, 2001[3]). There is a rich heritage of research in this field in South Africa and although much has yet to be investigated in terms of reliability, validity, application and utilisation of findings, the results surveyed in this study are valuable in any study commencing research in this field.

4.2.2 Ambiguous use of the word "potential" and the effects on the status of dynamic assessment

Test batteries are available in South Africa which contain within their titles the word "potential" apparently conveys the meanings inherent in the terms "ability" and "aptitude". These three terms are not, however, synonymous. They represent different aspects of intellectual functioning and measurement; the term "potential" represents a concept or meaning pertinent to a completely different paradigm of measurement or assessment, whereas "ability" and "aptitude" are taken from a product-based approach[4] to assessment, "potential" is taken from a process-orientated approach to assessment.

It is not the aim of the present study to highlight all test batteries which make use of the imprecise terminology, but this issue becomes an even more inflamed one, when persons involved in testing at various institutions indiscriminately make use of the word "potential" when they are clearly referring to "ability" and/or "aptitude". Often all three terms are used in the same sentence, intimating that all three can be used interchangeably. This is a looming indication that, if dynamic assessment is to gain a stronger foothold in as varied a context as possible, the first task is to familiarise practitioners with the differences between such terms. (Chapter 2 has dealt with the differences between these concepts, but the semantics is mentioned here because of the overwhelming misuse of the terms).

4.2.3 Misperceptions of the role of dynamic assessment in South Africa

When the term dynamic assessment is mentioned to persons involved in admissions, selections and testing in South Africa, those who are not too familiar with the notion often refer to this method of assessment as one in which the applicable field is **disadvantaged learners only**. Although this is the main field of application in South Africa and abroad, this is by no means the sole field of application.

2 In this regard the reader's attention is drawn to the Teach-Test-Teach (TTT) programme implemented at the University of Natal (see section 4.5.2.1).

3 Here reference is made to the Teach-Test-Teach programme offered up till 1995 at the University of Natal but is no longer implemented.

4 See section 2.6 on the difference between process- and product-based approaches towards assessment.

Dynamic assessment is not only applicable to educationally disadvantaged learners, but to gifted learners[5] (in disadvantaged circumstances), and to average and above-average learners who are struggling with one or more components of cognitive functioning (and these problem areas need not necessarily reflect serious deficiencies). Culturally disadvantaged children, who have not received adequate mediated learning experience at home or who find themselves in a different culture, are also learners who can and do benefit from dynamic assessment measures. **Dynamic assessment is not only related to slow learners or disadvantaged learners, but does represent a main research focal area within the field**.

Dynamic assessment is not **the** alternative to static testing, but is made use of in a **complementary** manner, for additional value-laden information that would otherwise be unobtainable through the exclusive use of static measures. Huysamen (1997) states that "…the fact that applicants have been exposed to educational systems of differing quality does **not** necessarily provide grounds for abandoning psychological testing as a component of admission procedures" [own emphasis] (p. 67). Furthermore, he states that according to some research in dynamic assessment, results of some studies do not warrant the abolition of aptitude testing.[6] It would do well to heed this statement even though dynamic assessment offers a viable alternative, for it (static testing) is able to offer value-added information, in addition to what is already available. Dynamic assessment is also not a method through which standards are lowered in order to accommodate larger numbers of disadvantaged students, but merely seeks to identify those disadvantaged students who harbour the potential to perform well within certain contexts, identifying learners who would most likely be ignored if not assessed in this manner. However, it is also not a panacea for the many problems facing measurement in South Africa. Admittedly, this method of measurement has much to overcome but is a method that can only offer greater advantages than hitherto offered by static tests alone.

PART A

4.3 Results of literature studies - sources gained from the literature databases

The following section investigates, in detail, the results of various studies within the field of dynamic assessment in South Africa. The results, conclusions and recommendations are findings that each author has made and are not offered as opinions in this study. However, where it is deemed necessary, comments are provided in footnotes. These are included so as not to hamper the flow of text but are fruitful in highlighting aspects that may otherwise be overlooked. Each study is dealt with separately but is cross-referenced with other research studies so to allow easy reference to other similar points made in other studies. Regarding the various tests used in the studies that are surveyed here, after the initial description of each test, the test thereafter is only named and not described again. This is to ensure ease of reading and to avoid unnecessary repetition which might disrupt the flow of the argument.

5 Recent overseas research includes Lidz and Macrine (2001) who demonstrate that dynamic assessment is able to contribute to the identification of gifted minority children; Bolig and Day (1993) who state that dynamic assessment can be used to assess gifted minority/poor children as well as Borland and Wright (1994) who make use of dynamic assessment procedures as part of a programme to identify economically disadvantaged potentially gifted nursery school children.

6 Here Huysamen is in fact quite correct, as he also points out that many of these static aptitude tests are made use of in many dynamic assessment programmes. "…such tests formed an integral part of the procedure for determining students' so-called modifiability scores…" (1997, p. 67).

4.3.1 Manifest and potential performance in advantaged and disadvantaged students - Doctoral study conducted by I.M. Shochet (1986)

4.3.1.1 Introduction

The study completed by Shochet in 1986 reflects the pioneering endeavour in the field of dynamic assessment in South Africa, in terms of its usage of the Feuerstein Learning Potential Assessment Device (LPAD) (group administration) for the purposes of tertiary selection procedures at the University of the Witwatersrand. Owing to the then (apartheid era) unsuitable measures taken to select university students for admission to the Arts faculty, Shochet's novel idea of utilising dynamic assessment proved to validate his two hypotheses of student selection and admission procedures based on the results of two sets of assessment procedures shaped by LPAD philosophy. "The study set out to find alternative predictors of university success other than the school matriculation results" (Shochet, 1986, p.xvii).

Manifest functioning of individuals as measured conventionally by static achievement tests is only valid for those advantaged students who were afforded the "luxury" of an education under the then white government. Disadvantaged students (those who were educated under the Department of Education and Training, the Department of Indian education as well as the Department of Coloured education) were seen to be unfairly discriminated against and hence not offered the same opportunity to develop their potential at a tertiary institution.

4.3.1.2 Research design

Figurally the design of the study can be seen as follows:

Table 1

Summary of the Primary and Secondary subject, Predictor and Criterion variables (Shochet, 1986, p. 201)

Subject Variables	**Predictor Variables**	**Criterion Variables**
Primary Subject Variable 1. Level of disadvantage a) Advantaged b) Disadvantaged obtained	1. Matric*[sic]* rating 2. Deductive reasoning traditional 3. Deductive reasoning enriched 4. Deductive reasoning difference 5. Pattern relations traditional 6. Pattern relations enriched 7. Pattern relations difference	**Primary criterion** Number of credits **Secondary criterion** Average mark
Secondary Subject Variable 1. Sex 2. Age 3. Attendance at a special programme a) Attenders b) Non-attenders		

The two hypotheses tested:

- HA1 - Advantaged and disadvantaged students will have different predictors correlating significantly with the criterion of university success.

- HA2 - The prediction of university success will be significantly enhanced through the introduction of the enriched testing conditions as conceptualised by Feuerstein and operationaliased for the purpose of the study.
    - HA2 - Sub-hypothesis - the prediction of university success will be significantly enhanced for disadvantaged students through the enriched testing conditions.

The two predictor tests chosen for this study consisted of the Deductive reasoning test[7] (which is a "…test of deductive syllogistic reasoning" (p. 156) and the Pattern Relations Test (a "…test of inductive reasoning and reasoning by analogy" (p. 160) both developed by the National Institute of Personnel Research (NIPR) in Johannesburg. These two conventional or static testing procedures were then altered in a dynamic manner - in other words in using the LPAD approach, changes were made to the construction of the tests and in the administration in the testing (Shochet, 1986). A few guiding principles were used when deciding on the use of these tests, namely, "generalisability" of the mediation received and as such the tests would be internally consistent so as to allow for potential transfer of learning from one test exercise to another; the tests were normed on the population undergoing testing; the tests had pre-established reliability coefficients and manifested face validity (Shochet, 1986).[8] Shochet employed the Feuerstein cognitive map as a guide in which to dynamically assess students using these two conventional tests.

In changing the examiner-examinee relationship to a teacher-pupil relationship[9] and also by introducing a training process into the assessment procedure, Shochet's testing procedure is classifiable as a dynamic assessment. As conventional tests are not valid predictors of university success for disadvantaged and more modifiable students, it is construed as unfair practice to select students based on these static premises. One needs to look for modifiable potential - potential that can be harnessed at tertiary institutions made evident by assessment which entails mediation and measurement of potential. This is the crux of Shochet's argument. Academic selection procedures are by their nature predictive and in a rather ironic manner, the Feuersteinian approach denies the notion of predictability since it does not imply an unchangeable and consistent pattern of cognitive abilities. Shochet shows, however, that prediction of potential cannot be classified in the same way. The predictors for the study were the results of the two conventional tests (DRT and PRT) as well as matriculation marks and the criterion variables were the marks received at the end of the first academic year at the university (average marks and number of credits received). However, school performance remained a strong predictor in his study, correlating strongly with first year success for advantaged students but not at all for disadvantaged students. Hence the need arises to look for alternative selection procedures (see Table 1).

By enriching the testing procedure for each of these tests, Shochet was able to implement the Feuersteinian concept of mediation and enrichment, in other words the test-coach-test method of appraisal. Both the advantaged and disadvantaged groups were given the traditional form of the tests as well as the enriched version of the tests. Predictions based on both sets of results were made. Furthermore, the group was divided into modifiable and less modifiable students. The two tests were "enriched" according to the methods prescribed by Feuerstein. The study thus encompassed three main variables; subject variables (advantaged and disadvantaged students), predictor variables (traditional and enriched tests and the resulting difference between them [modifiability] as well as matriculation results) and criterion variables (first year academic marks). The sample consisted of N=156 students registered for the first time in the first year of the Faculty of Arts (1984). These were divided into advantaged and disadvantaged students. The mean age was 19.56 years ranging from 17-31 years of age and was reflective of the student representation of the those taking courses in the

7 See section 4.4 for more information on the PRT and DRT by Shochet (1994).

8 Boeyens, whilst commenting on reliability issues and himself having read Shochet's study, states that Shochet "[u]nfortunately …fails to report the standard deviations obtained on the PRT both when administered as a pre-test and when administered as a posttest" (1989a, p. 41).

9 The teacher-pupil relationship assumes a more active role from both parties. In this regard, Lidz (1987) states that both are active in this newly defined relationship.

Arts faculty. Students were tested twice on the both the Deductive Reasoning and Pattern Relations tests, with the experimental group receiving enriched training on the tests:

> The first time the test was administered in its traditional form in order to establish the baseline manifest intellectual functioning and on the second attempt (following immediately after the first) the students were provided with teaching, mediation and enrichment during the testing in order to assess their potential intellectual functioning (Shochet, 1986, p.150).

Testing the hypotheses

The statistical methods used in the analyses of this data was comprised of simple correlations as well as regression analyses, which is the most common method employed when dealing with predictive studies. Regression analyses was carried out separately for advantaged and disadvantaged students, and both main and interaction effects were examined. Significance levels were set at 0.05 and were based on two-tailed assumptions.

Regarding the first hypothesis, to test for a main effect, a stepwise multiple regression was conducted. Firstly, predictors were included (as discussed above) as well as the multiple R^2. Secondly, a stepwise regression of Feuersteinian variables was selected to see whether they significantly improved on the prediction. If any of these Feuersteinian variables significantly improve the model then the first hypothesis can be supported. F tests were carried out to ascertain the level of significance for the incremental increases in multiple R^2. Interestingly the study also yielded significant moderator variables which "…refers to a variable that influences the predictive effectiveness of a predictor variable (Saunders [1956] in Shochet, 1986, p.225).

Shochet's study grouped the subject variables (the students) into the four groups to which they were initially assigned based on how they were recruited for the study, but conducting Sheffé tests yielded no significant differences in the means of the groups, which is positive since they differed in number from group to group.

4.3.1.3 Main findings

A t-test for paired differences between the traditional and enriched conditions for the Deductive Reasoning Test as well as the Pattern Relations Test was significant at the 0.001 level. This was significant "…not only for the sample as a whole but also separately for advantaged and disadvantaged students" (Shochet, 1986, p.232).

As mentioned previously, simple correlations were determined, serving as preliminary statistical procedures before multiple regression was determined. When Pearson Product Moment correlations were computed for the advantaged and disadvantaged students, none of the predictor test variables correlated with the criteria for the advantaged students. The only significant correlation for the advantaged group arose from the relations between matriculation results and university success, explaining 30 per cent of the variance of students' end of year results. As Shochet has mentioned, this finding is not surprising as matriculation results predict quite well for advantaged students. However, when Pearson Product Moment correlations were run for disadvantaged students, the matriculation results no longer showed any significance as a predictor of success at university.

The traditional administration of the Deductive Reasoning Test, however, was seen to correlate negatively with success in the disadvantaged students. Hence the first hypothesis is supported by the results, namely, that advantaged and disadvantaged students will have different predictors correlating significantly with the criterion of university success. Furthermore to see just how strong or weak the differences in the correlations between the two groups were, Fisher's $\underline{z}$ transformation was done. Based on the matriculation score as predictor, there was a significant difference at the 0.05 level between the groups as well as for the degree of modifiability on the Deductive Reasoning Test (i.e. the difference in scores between traditional and enriched conditions on the test).

In order to accept the second hypothesis, it has to be proven that the Feuersteinian measure, taken when conducting the enriched versions of the tests, would significantly improve upon the predictability

of the model. Because there was a difference between the advantaged and disadvantaged groups in terms of which predictors were found useful, stepwise multiple regression was conducted for both groups once again. Results showed that none of the Feuersteinian measures enhanced the prediction of the model for advantaged students. Thus the second hypothesis cannot be affirmed for the advantaged students by a main effect. The results were also negative in that no Feuersteinian measure was shown to enhance the model for disadvantaged students significantly. The picture looks bleak[10] until that is, Shochet investigates for moderator effects (discussed above) for both the advantaged and disadvantaged groups. It is possible that the second hypothesis is indeed supported by a moderator effect. Interpretation of Shochet's results is supported by De Villiers (1999), and as elaborated below, "…some difficulties surround the interpretation of this finding [i.e. the modifiability of students as moderator variable] because of the low reliability of the difference scores" (Boeyens [1989a] in De Villiers, 1999, p.123). Moderated multiple regressions[11] were conducted to ascertain whether students' modifiability (the difference in scores between the traditional and enriched conditions) in anyway moderated the Feuersteinian measures.

Once again, there were no significant results pertaining to the advantaged group, and it would seem that this group's performance can be predicted to a fair extent based on matriculation results alone. For the disadvantaged group, there is a significant finding for the moderator effect (modifiability) with R^2 being significant at the 0.05 level but only for the Deductive Reasoning Test. This moderator effect was found to be significant, the incremental increase in R^2 being significant in an F test. As the second hypothesis can now be affirmed so too can the sub-hypothesis attached to it, namely that the prediction of university success will be significantly enhanced for disadvantaged students through the enriched testing conditions.

In order to further clarify the results of the moderated stepwise regression, Shochet conducted subgrouping analysis as an alternative. The disadvantaged students were subgrouped according to their difference score (modifiability score) on the Deductive Reasoning Test. Those falling above the mean were classified as more modifiable and those falling below were classified as less modifiable. The subgrouping analysis allowed Shochet to pinpoint exactly where the differences were significant. The results showed that there was a clear moderator effect for the less modifiable students as opposed to the more modifiable students. Shochet's statistical investigation into the results appears thorough in that after subgrouping was conducted, Fisher transformations were once again conducted to see whether there were any significant differences between the subgroups' correlations. The results showed that there were significant differences. "Statistical artefacts" were also accounted for, including factors such as a ceiling effect. The fact that no moderator effect was found for the Pattern Relations Test made it such that the chance of encountering an anomalous finding may have been possible (Shochet, 1986).

However, the fact that two statistical procedures were carried out (the moderated stepwise regression as well as subgrouping) can lay that issue to rest. In order to make sure that a moderator effect had indeed taken effect in the Deductive Reasoning Test administered in its traditional form and was not simply a fluke, Shochet (as described above) ran moderated multiple regressions and subgrouped the findings. However, he also decided to check to see whether or not the Pattern Relations Test difference score (which tells one more about the enrichment) acted as a moderator on the Deductive Reasoning Test score. Shochet ran another moderated multiple regression analysis for the disadvantaged students and found that "…the modifiability on the Pattern Relations Test…is also moderating the relationship between [the Deductive Reasoning Test in traditional form] and university success" (Shochet, 1986, p.256). He concludes that "…the moderator effect of [the Deductive Reasoning Test administered in traditional form] was not an anomalous result and that similar trends

10 Boeyens comments on Shochet's findings stating that "…the mean difference score…obtained by Shochet is… too low to be interpretable" (1989a, p. 43). Shochet makes use of conventional test instruments in a dynamic manner as stated above, however, Boeyens states that there "is so much noise" in his data that one cannot really say whether learning took place or not and concludes that conventional instruments are not reliable enough for learning potential assessment.

11 See section 4.4 for more information on this statistical method which Shochet also employs in a 1994 study on cognitive modifiability as moderator effect in predicting academic success.

can be found when [the Pattern Relations Test administered in traditional form] is used as the moderator" (Shochet, 1986, p. 257). Hence, the sum of the difference scores (i.e. the degree to which students were modified) was added as a greater measure of modifiability, which in turn was computed using another moderated multiple regression analysis. The results further supported the second hypothesis. A succinct summary by Shochet after all moderated analyses were run stated that:

> [t]he moderator effect of the construct of modifiability on [the Deductive Reasoning Test administered in the traditional form] for the disadvantaged students, means that the more modifiable the disadvantaged student, the less will be his/her predictability on the basis of the traditional intelligence test of [the Deductive Reasoning Test administered in the traditional form] (and visa versa) (Shochet, 1986, p. 260).

4.3.1.4 Conclusion and recommendations

The conclusions of the Shochet study are a positive step for the continued use of dynamic assessment, more specifically dynamic assessment used as potential selection tests at tertiary institutions although the study's sample and population was focused on students wishing to fulfil a degree in the Faculty of Arts. School results do not have predictive validity for disadvantaged students but do have predictive validity for advantaged students. Most important was that enriched testing conditions will significantly aid in selecting disadvantaged students, especially those students who are classified as being modifiable.[12] An important point made in the study was that none of the Feuersteinian measures actually had an effect in a direct manner but were effective only as moderator variables. This lack of "directness" of Feuersteinian assessment is not an uncommon criticism.

Shochet does make the important distinction between measurement and assessment, when he points out that his study was an involvement of assessment and not really measurement, as "measurement" is usually paired with "psychometric", and "assessment" is usually paired with "dynamic". Learning potential cannot be measured but assessed; it goes against the fundamental tenets of Feuersteinian motions to use the term measurement as this translates into a static concept of intelligence, the very idea from which movement seeks to distance itself. In other words, manifest prediction (the old static way of measurement) should be abandoned as it is clearly unsuitable for a large section of this specific population group, and that learning potential should be the main focus of selection procedures. It must be noted, however, that this study was conducted in 1985/1986 and as such the political times were very different from what they are today. However, the fact that many previous disadvantaged communities still exist is in no doubt.

4.3.2 Effectiveness of the learning potential assessment device with high achieving adolescents from an advantaged community - Master's study conducted by S.R. Hoffenberg (1988)

4.3.2.1 Introduction

Hoffenberg's (1988) findings with the Learning Potential Assessment Device (LPAD) conducted with high achieving adolescents from a disadvantaged community did not, in general, reveal positive results.[13] Mediation was not effective in improving performance on the LPAD tasks nor on the transfer measures:

> Of particular importance in terms of the broad aim of the study, that is, to examine the usefulness of the LPAD identifying giftedness among the Black disadvantaged community, was the finding that mediation did not raise the intellectual status of the subjects to the extent

12 These and similar findings in other studies are in keeping with oversees findings as highlighted by Lidz (in Flanagan et al. [1997]).

13 See the results of Gaydon (1988) (section 4.3.3) in this regard.

> that they would be regarded as gifted when compared to the norms of the general population (p. 41).

The realisation of a lack of learning potential tests in South Africa in order to identify potential is further underscored by the concurrent lack of tests for gifted children in disadvantaged communities. All too often the perception of disadvantaged learners is housed within the notion of this group forming an homogeneous pool of students who are all in turn much the same regarding level of intelligence and potential. The habit of classifying any disadvantaged group as homogenous is an error which can lead to difficulties later in research, especially when one has to differentiate between these members on certain variables. Not all disadvantaged students perform similarly just as advantaged learners in middle class schools are not on the same level of achievement and Hoffenberg's study offers a refreshing look at identifying high achievers in disadvantaged communities.

The failure to find any significant differences between control and experimental groups was thought to be caused by the ineffectiveness of the mediation process (which itself has been seen as a detrimental aspect in the research of Boeyens (1989a, 1989b) and Tayler (1996). Hoffenberg reasoned that since certain culture-fair tests are norm based, scores of the minority (usually African students) tested would inevitably fall below the scores of majority (usually white students) learners. One way of overcoming this was to use the dynamic testing approach of Feuerstein's LPAD. Also mentioned was the programme of assessment, diagnosis and instruction (PADI) developed by Johnson, Starnes, Gregory and Blaylock (1985) "…which combines the use of culture-fair tests and the potential enhancing approach" (in Hoffenberg, 1988, p.4).

Hoffenberg (1988) explores the less often researched area of high achieving students, making use of Feuerstein's (LPAD) method. Although dynamic assessment still has a long way to go in proving its effectiveness, especially in South Africa (with contentious issues such as costs and timing), Hoffenberg asserts that "…dynamic assessment techniques which are designed to measure potential rather than manifest ability, show much promise in this regard" (1988, p.ii).

Hoffenberg's motivation for the study was the paucity of research in the field concerning identification of students who are gifted and not necessarily identified as such and the need to develop alternative measures of identification of gifted students. The aim of the study was to ascertain the value of using the LPAD device in screening for potentially gifted students in high school. The continuing emphasis in research on finding and treating deficits within the black schooling community in a way neglects those students who are not deficient in any way and hence does not offer them the opportunity to develop to their full potential. The question is whether the LPAD is a useful instrument in this endeavour. Hoffenberg discusses a number of tests which have purported to be "culture-fair" but which really only employ different norms for different groups, which merely emphasises the difference in norms for each group, replaying the notion of differences between the races.

Previous measures were based exclusively on student performance on tests and achievement in the upper 5 -15 % range, which seems a very constricted means of judging giftedness, especially since, at the time of the study, most schools had multiple systems by which to rate and place students. Many schools rated students differently, based on different assessment techniques. Another major issue was the fact that the socio-economic status of students played such an important and impinging role on their performances that to merely accept results would be unfair, seeing that some had better opportunities than others. Instead of studying manifest results, Hoffenberg chose to study the degree of change that students underwent after mediation, this result serving as a better indicator (and a fairer one) of giftedness.

The two hypotheses of the study were thus:

(1) It is expected that mediation on the various tasks of the LPAD would be effective in improving the performance of subjects on these tasks, relative to the performance of subjects who did not receive such mediation, in other words "…[f]ollowing mediation, the performance of the experimental group will be significantly better than that of the control group on certain LPAD tasks…" (Hoffenberg, 1988, p.11); and
(2) It is expected that effects of mediation would generalise to performance on other, comparable measures, in other words, "…[f]ollowing mediation on the LPAD, the performance of the

experimental group will be significantly better than that of the control group on the respective transfer measures…" (*ibid*.).

4.3.2.2 Research design

One hundred pre-selected black adolescents from Soweto and Alexandra who attended a gifted student programme were divided into experimental and control groups where the experimental group was administered dynamic testing. (Hoffenberg refers to the technique as "testing" and not as "assessment".) The sample was more or less evenly divided among male and female with the mean age of 14 years ranging from 12 to 19 and ranging educationally from Std Six to Std Ten. The mean IQ of the group (derived from the New South African Group Test - HSRC, 1985), was 90.7, ranging from 74 - 116. A minor problem was the use of this test as it was normed on the white South African population, but owing to the lack of an appropriate test for this sample, this instrument had to be used to indicate IQ levels. Hoffenberg also gathered pertinent data on the socio-economic background of the students, including their parent's education and their own living conditions so as to gather as much information of the level of disadvantage as possible in order to use this information as possible factors of influence in test results.

Hoffenberg's dynamic testing refers to the mediation that the experimental group received on four of the LPAD tasks, and the results compared both groups' performance on the LPAD tasks as well as on "…measures comparable to the LPAD tasks" (*ibid*.). The three LPAD tasks used were the Organiser (A) (a verbal sub-test with a numerical component); Verbal Analogies Test (VAT, a verbal sub-test, which is a measure of analogical thinking); and Set Variations I and II (a non-verbal sub-test based on Raven's Progressive Matrices, which is a sub-test of non-verbal analogical thinking). The three comparable test measures, which acted as transfer measures, consisted of the Organiser (B); Similarities subtest as well as the Raven's Standard Progressive Matrices. "Each of these [independent] measures was selected to assess a comparable ability to its LPAD counterpart, and thus served as a measure of transfer of learning from the LPAD task" (Hoffenberg, 1988, p.14). The tasks of comparable measures are thus measures of transfer and are referred to as such.

Hoffenberg's design made use of the Solomon design and looked as follows:

Table 2

Summary of the sample (Hoffenberg, 1988)

	Pretested		Unpre-tested	
	Experimental	Control	Experimental	Control
Group	I	II	III	IV
Number of subjects	24	23	28	25

According to Neuman (1997, p.187), " [a] researcher may believe that the pretest measure has an influence on the treatment or dependent variable. A pretest can sometimes sensitize subjects to the treatment or improve their performance on the posttest". In order to ensure that the results truly are a reflection of the mediation administered to the experimental group and are not attributable to the fact that subjects become sensitised to mediation (in the pre-test), this factor (of sensitisation playing an unwanted role) is eliminated. No differences were found between the pre- and posttest groups on age, IQ, socio-economics standards (SES) or academic achievement, after conducting both Chi-square analyses and ANOVA's.

Groups I and II were administered the transfer measures tasks before and after mediation, whereas groups III and IV were given an educational video to watch. Both groups were divided into smaller groups so as to receive the LPAD tasks, the experimental group receiving the mediation but not the control group. The groups consisted of a small number of students so as to allow for more effective mediation.

Each group of students (both experimental and control) was headed by a psychology student, trained in administration with and without mediation. These psychology students were trained in small groups supervised by post-graduate psychology students. The "critical distinguishing" factor in the experimental and control groups was the fact that the former group received mediation. "[They] were taught the cognitive strategies involved in solving the particular task in line with the procedures described by Feuerstein et al. (1983)" (Hoffenberg, 1988, p.20). The independent measures allowed the researcher to see whether transfer had indeed taken place after the completion of mediation. The control groups who received the LPAD tasks were merely shown what to do and were not coached in the sense of receiving true mediation. Performance on the transfer measures as well as on the LPAD served as the dependent variable.

An ANCOVA was conducted for the pre-tested groups in order to study the effects of mediation on the transfer scores as well as on scores in the LPAD tasks and for the unpre-tested groups, where the pre-test served as the covariate, an ANOVA was conducted to "…examine the effects of mediation on the LPAD on subsequent performance on the LPAD tasks and their respective transfer measures" (Hoffenberg, 1988, p.22).

4.3.2.3 Main findings

In assessing the above, one would be able to ascertain the effectiveness of this approach in identifying potentially gifted students in South Africa and the feasibility of such an effort.

Firstly there were no differences between the pre-tested and unpre-tested groups on performance results. "An ANOVA revealed that there were no significant differences between the scores of pre-tested and unpre-tested groups for both experimental and control halves" (Hoffenberg, 1988, p. 26). Within the pre-tested group, the experimental group performed consistently better than the pre-tested control group, however only the results on the Set Variations II yielded a significant result. As for the unpre-tested group, there were no significant differences at all for any of the tasks. However, of interest is the fact that the control group performed better than the experimental group on the Organiser.

It must be noted here that unlike many South African studies conducted in the field of dynamic assessment, most do not take cognizance of the possible effects of pre-testing on results[14] and the Hoffenberg study is an example where this is taken into account, and should be viewed as a contingency measure when conducting studies of a similar test - teach - test design which make use of both an experimental and control group. It thus avoids possible contamination of results. Analyses for the whole group were conducted on the basis that both groups were similar, in other words, no main effects of pre-testing was noted.

Results of the experimental and control groups were not that divergent. The only test in which the control group outperformed the experimental group was on the Organiser (A). However, even though there was a tendency for the experimental group to perform better than the control group, the only significant differences encountered were on the Set Variations II, which was significant at the 0.05 level. Thus the first hypothesis cannot be wholeheartedly supported because of the poor results. Only one significant measure was obtained for the experimental group, whereas the hypothesis stated that significant differences would be revealed on all four LPAD tasks.

The ANCOVA and ANOVA conducted for both the pre-tested and unpre-tested groups revealed no significant differences between the experimental and control groups even though the experimental group tended to achieve higher scores in each of the transfer measure tasks. The unpre-tested group, however, did reveal an interesting finding once again with the Organiser (B) in that the control group performed better than the experimental group. Hoffenberg once again ran an ANOVA to ensure that no main effect caused by the pre-test measure accounted for differences between the two groups. The result was the same. No significant differences were found, and so both the unpre-tested and

14 Pre-testing did in fact have an effect in the De Villiers (1999) section 4.3.18.3 study for example.

pre-tested groups were combined for analyses. Once again, no significant results were yielded when an ANOVA was conducted to highlight differences. There was once again only a tendency for the experimental group to achieve higher scores for the Raven's Progressive Matrices.

Hoffenberg analysed the results for the pre-tested groups, seeking differences between the pre- and posttest results on the transfer measures. (As already noted only the transfer measure tasks were administered both in the pre- and posttests.) There was a tendency towards improvement for both groups (even though it has already been established that this is not the result of a main effect caused by the mere fact that pre-testing took place). Hoffenberg notes that the pre-test results on the Organiser (B) left room for improvement, hence the marked (yet insignificant) improvement in results on the posttest. For the Raven's test, by contrast, there was little room left for improvement owing to the high scores initially achieved. Hoffenberg's rather disappointing conclusion is that "[m]ediation on the LPAD did not appear to have a significant effect on the posttest transfer measure scores" (p. 34). Hence the second hypothesis cannot be supported.

4.3.2.4 Conclusion and recommendations

Could these negative results be due to the fact that the LPAD was not initially designed for the use of diagnoses of gifted children? Hoffenberg nowhere states this as a possible factor but rather discusses other limiting factors of the study which may have contributed to the poor results. In this context, the term "poor" is used in the sense that neither hypotheses was supported to any significant degree, although the first hypothesis was partially supported. A variety of reasons are cited as to why the results are so poor. Hoffenberg states that perhaps the mediation provided for the Organiser (A and B) was inadequate,[15] and the improvement from pre- to posttest may have showed a delayed "effect of mediation" for the experimental group (p. 36).

As for the Verbal Analogies Test and Similarities Test, Hoffenberg states that confusion may have arisen because of the task demands on both of these tests, the Verbal Analogies requiring deductive processes whereas the Similarities requires inductive reasoning. However, it is questionable whether these are adequate reasons, since it should be assumed that both tasks should have been explained thoroughly. Regarding the Set Variations I and II, Hoffenberg states that owing to the relative lack of language conveyed and used in the tasks, students performed better, and "…it is likely that [they] experienced less of a disadvantage as a result of being mediated in English … (p. 38). Also two reasons put forward for the lack of increase in results on the Raven's test, are firstly, that the effect of mediation on the Set Variations did not generalise to the Raven's and secondly, that a "ceiling effect"[16] was reached by students, who performed well on the pre-test, so much so that there was little room left for improvement on the posttest.

Other reasons cited included:

(1) The Hawthorn effect, in that irrespective of mediation, students would improve from pre- to posttest and that participation alone may account for these slight improvements
(2) Limitations in the transfer effects (as discussed above)
(3) Limitations regarding the training of moderators. Hoffenberg mentions that the psychology students used in the study were not necessarily experienced enough in the ways of teaching or moderating and included those "…who received relatively little training in the LPAD" (p. 40); but she states on page 20 that "[t]he students were trained in administering tasks both with and without mediation. **In order to ensure that the students acquired the necessary skills, training was conducted in small groups**" [own emphasis].

15 Feuerstein emphasises this very aspect of poor administration when he states that "…the presence of negative examiner-examinee interactions [in this instance poor mediation] that contribute[s] to the masking of potentiality is usually ignored, and failure is attributed to the child's lack of capacity" (1979, p. 319).

16 The ceiling effect issue is difficult to answer because of the regression effect, in which the regression to the mean may lead to a negative status gain for low level performers who seem to gain more from the retest than high performers (Klauer in Hamers et al., 1993).

There seem to be some contradictions in this regard in the Hoffenberg study. The last three reasons cited by Hoffenberg are aspects mentioned frequently in studies of this nature, and that is the lack of time needed for adequate mediation, the lack of individualisation with students when administering the mediation and the lack of mediation in the language of the students. This is a major concern when it comes to applying dynamic testing in South Africa, aspects which have proved very difficult to surmount.

The final conclusion reached by Hoffenberg is that the mediation provided was inadequate (notwithstanding the already high achievement level of gifted students, which militated against further improvement). Lastly and most importantly (apart from the constraints imposed on this study [mentioned above] which may have affected the results to a greater degree than expected), Hoffenberg states that "…the findings of the study suggest that the LPAD is an unsuitable instrument to administer on a wide scale in that it is too costly in terms of time and manpower. This approach is therefore not recommended as a first line of testing for giftedness" (p. 43).

4.3.3 Predictors of performance of disadvantaged adolescents on the Soweto/Alexandra gifted child programme - Master's study conducted by V.P. Gaydon (1988)

4.3.3.1 Introduction

The research conducted by Gaydon (1988) is one of the earlier research studies in the field of dynamic assessment in South Africa. As with the study by Hoffenberg (also conducted in 1988) who studied the effects of mediation (by using the LPAD) on a sample of disadvantaged yet gifted students (as with Gaydon), the findings were not encouraging (see section 4.3.2 above for results on the Hoffenberg study).

It must be recalled that at this time in South Africa's history the Apartheid education system was still being implemented, with progressive centres such as the Soweto/Alexandra Gifted Child Programme attempting to better the future prospects of some gifted black students. The main aim of the study was to "…determine the relative effectiveness of both conventional and dynamic assessment techniques in predicting performance of culturally and socio-economically disadvantaged students [in this project] with the further aim of developing more effective selection procedures for the programme" (Gaydon, 1988, p. ii).

Citing the work of Hoffenberg (1988), Gaydon was already aware of the somewhat discouraging results of that study. However, Gaydon' s goals were not aimed solely at assessing the usefulness of the LPAD, but in determining the usefulness of both static and dynamic assessment instruments for this group of students. Poverty, lack of parental guidance, poor schooling as well as the socio-economic structure at the time were all factors militating against the effective schooling of many students. The attention paid to the ordinary black student was framed in these terms and the likelihood of the potentially gifted students receiving aid was scarce. Also, as with Hoffenberg's tests, the use of white-normed tests did not allow for the fair assessment of these students. Dynamic assessment would thus seek to redress this situation. Gaydon briefly reviews the literature on giftedness research and comes to the conclusion that the very concept proves difficult to define.

Gaydon points out that in the attempt to develop "culture-free"[17] tests, the Raven's Progressive Matrices Test can be cited as one such test, owing to the non-verbal nature of the items included. However, the sentiments of Gewer (1998), (see section 4.3.15), writing later and accessing newer information, states that the "…the Raven's Matrices is not[18] an accurate measure of intellectual ability

17 It is debatable as to whether such a test exists or will ever exist. The notion of "culture-free" is an ideal for which to strive. In fact Feuerstein et al (1979) state that "[s]ince it [the Raven's] is considered, perhaps fallaciously, to be a culture-free test and to require only a limited amount of verbal instruction or mediation, it is preferred by practitioners…" (p. 150).

18 It is unclear exactly why Gewer states that the Raven's is not an accurate measure, as he makes this statement based on other researchers' efforts in the field, cited in his dissertation.

within a cross-cultural context" (Gewer, 1998, p. 19). However, Gewer used the Raven's **Coloured** Progressive Matrices as part of his research. Even though a test battery may be said to be "culturally fair" or "culturally free", it does not necessarily mean that the tests are dynamic in nature. It is true to state that many dynamic tests (or at least conventional tests used in a dynamic manner) may be developed and used within the context of testing culturally disadvantaged students/learners, but this does not imply that all culture-fair tests are dynamic. In her discussion on the culture-free batteries available, Gaydon refers (among others) to the System of Multicultural Pluralistic Assessment (SOMPA) which she states has been used as an alternative approach to the assessment of individuals from disadvantaged communities. However, a shortcoming of this battery is that the potential of the learner is not tapped.[19]

As the LPAD has already received attention elsewhere in this study, it is only referred to as being used and is not discussed again in detail. As Gaydon was writing in the late eighties, reference is made to research in this field at that time, namely the research of Skuy and Shmukler (1987)[20] as well as Hoffenberg (1988).

The Programme

The criterion for entry into this programme at that time was the need for the students to obtain marks in the top 5 - 15% of any subject. An aptitude test was also written. As each DET school at the time had their own system in place, this criterion may seem a bit arbitrary. The only selection tests available for this programme were static tests based on prior learning. The aim of the study was to determine whether alternative measures could be found which would reasonably predict those students likely to succeed in the programme, without relying on prior learning (which has been considered invalid owing to the status of education at that time) and hence selecting these culturally and socio-economically disadvantaged students specifically for the Soweto Gifted Child Programme.

Gaydon correctly emphasises the need to make use of multiple measures of selection as is often the case, multiple tests are able to render a more accurate account of what the student is capable, and these multiple measures include eight different measures, notably: "The New South African Group Test; Ravens *[sic]* Standard Progressive Matrices; School performance; The Khatena-Torrance Creative Perception Inventory; The Piers-Harris Childrens' Self Concept Scale; and three instruments from Feuerstein's LPAD, and their associated transfer measures" (Gaydon, 1988, p. 13). The potential combination is important as different tests may yield different results when used together.

Gaydon briefly introduces the tests to the reader, and the statement that the Raven's is "…a culture-fair measure of intellectual ability" (p.14) is clearly at odds with the sentiments of Gewer who stated the Raven's Standard was not a culture-fair test and so made use of the Coloured Matrices. This is quite interesting to note and has been mentioned above. The Khatena-Torrance Scale, accoring to Gaydon, is a measure of a child's creative thinking which is supposedly culturally attuned. As self concept of black students correlated with academic achievement , Gaydon decided to make use of the Piers-Harris Self Concept Test as well as the Teacher Temperament Questionnaire.

4.3.3.2 Research design

The sample consisted of 165 black students who had attended the programme at DET schools. All subjects participated in at least one testing session held at the University of the Witwatersrand. 100 students completed all three LPAD tests and transfer measure tests but the sample was further reduced to 99 due to the absence of one student. Owing to the fact that Gaydon employed a

19 In fact findings by Taylor and Richards (1990) indicate that a sub-test from the SOMPA battery, namely the Estimated Learning Potential, resulted in low to moderate correlations with traditional IQ measures and they concluded that learning potential should not be considered a unitary construct based on these findings.

20 Not referenced in this study.

Solomon Four Group Design[21] (see also De Villiers, 1999 and Hoffenberg, 1988 who also made use of the Solomon design in order to control for the practice effect), the total number of subjects being tested on the independent variables amounted to 47. The mean age of the sample was 14 and ranged from 13-18. Gender groups were more or less evenly distributed and biographical information was also obtained from the students, such as living conditions and parents' educational status.

So-called baseline measures were obtained using the above-mentioned test instruments (all measures including school performance, except the LPAD sub-tests which was used to test for the effects of mediation and transfer). It must be noted, that the Raven's was utilised for two reasons, namely, the need to "…obtain a less biased measure of non-verbal reasoning and intellectual functioning than the New South African Group Test; and as a transfer measure to ascertain the transfer of learning for the LPAD Set Variations I and II tasks" (1988, p. 21). The New South African Group Test had at that stage not been normed on black children and was still in the validation test phase. This is interesting to note, as the test may not have been suitable at all. However, it is possible that no other option was then available.

Mediation formed part of the three LPAD tests administered to the students, namely the Organiser (a verbal test with a numerical component), the Set Variations I and II (used in numerous studies) as well as the Verbal Analogies Test (VAT). Mediation took place in all three subtests, in the form of lessons (for the Organiser), training analogies (Set Variations) as well as aiding and helping in strategising (VAT).

The transfer measures were found by making use of three independent measures based on the tasks already performed. No mediation was given on these tasks. An alternative form of the Organiser (known as Organiser B) was used to test for transfer for this subtest. Raven's Standard Progressive matrices was used as the transfer measure for the Set Variations and the WISC-R (Wechsler Intelligence Scale for Children - revised) and WAIS-R (Wechsler Adult Intelligence Scale - Revised) similarities tests were used to indicate transfer for the VAT.

Thus baseline measures were obtained, as well as mediation and transfer measures and criterion measures, thus measuring the "[p]erformance on the DET Gifted Child Programme, …[in] English, Mathematics, Science and Aggregate marks…" (Gaydon, 1988, p. 24).

Design

Students were randomly assigned to one of four groups, the pre-tested experimental and control group as well as the unpre-tested experimental and control groups. This allowed for the detection of practice, mediation and interaction effects of pretesting. Gaydon made use of the same sample that Hoffenberg (1988) used. (See Table 2 for the sample description.)

The LPAD tasks spanned three hours and comprised three sessions each. The use of the LPAD and the resultant time taken by the mediation is one of the primary reasons why this technique is not altogether popular. The students were placed into smaller groups so as to allow for greater interaction between the subjects and testers. This can be commended, as other studies making use of mediation mediated to larger groups and stated that had circumstances allowed for smaller mediational groups, that this would have been the preferred choice.

To enhance the control for practice effects, the control groups were exposed to training sessions for each test (testing and training but not mediation). Three sets of data are available, namely for the group as a whole, for the pre-tested sample and lastly for the unpre-tested sample.

21 With regard to the possible effects of testing, Klauer states that gains due to "…retesting vary according to the **time** span between the two tests, the **kind** of tests involved and the **pre**-test levels…" [own emphasis] (in Hamers et al., 1993, p.138).

4.3.3.3 Main findings

Whole sample

Pearson's Product Moment Correlation Coefficient was used in order to determine whether there was any significant relationship between the criterion variables and the independent variables. Those variables found to correlate significantly with the criterion variables were used in a stepwise multiple regression analysis in order to determine explained variance of the criterion variable accounted for by the variation in the independent variables. Only the aggregate marks were used as criterion in this analysis.

Pre-tested Sample

Pearson's Product Moment Correlation Coefficient was used for Groups I and II in order to compare them on the independent variables. This, once again, established whether there was any relationship between the criterion and the independent variables. A stepwise regression analysis was carried out for this sample as well.

Unpre-tested Sample

Pearson's Product Moment Correlation Coefficient was used for Group III and IV in order to compare them on the independent variables, and once again to establish whether there was any relationship between the criterion and the independent variables. A stepwise regression analysis was carried out for this sample as well.

4.3.3.4 Conclusion and recommendations

Based on the findings of the whole sample, Gaydon states that it would be safe to assume that the combination of VAT, LPAD, overall school performance, Organiser posttest score as well as the combined similarities posttest score would be the best and "…most useful and cost effective predictors of performance on the ….Programme" (p. 45). Of course, it must be recalled that these measures are validatory measures for academically superior children from a disadvantaged background. In other words, the sample is quite specific. School performance as a predictor is hardly surprising states Gaydon, as the programme's emphasis is on school related tasks. A point of significance is that the children construed here as superior in terms of academic achievement, are so within their community but may not be classified as such when compared with other communities. (As Gaydon notes, the New South African Group Test results "…fell within the average or below average range of intellectual functioning relative to the White South African population" [p. 46].) Of course the New South African Group Test had not been normed on the population used in this study.

School marks would be the first set of results that would be assessed when considering allowing students to continue with the programme; secondly, the VAT as well as LPAD Organiser test in conjunction with the Raven's Standard Progressive Matrices and Similarities subtests of the WISC-R and WAIS-R would be used when considering who to allow to continue. Gaydon states that although these tests account for up to 50% variance in the Aggregate marks achieved in the programme, it would mean that potentially 50% of prospective students would be missed. This would entail further research into the types of tests used. Once again the reader is confronted with the same argument as to why mediation may not have proved successful, namely that the mediation administered was not very good. This begs the question of effective training in the first place. It must be pointed out that if, in a study, one's objective is to ascertain the level of effectiveness of mediation, the mediation administered would have to be adequate. Yet, as can be seen in this study and in the Hoffenberg (1988) study, the mediators are not trained particularly well.

If one is to train mediators for future research, it should be done with greater concern for the end results. Gaydon states that:

> [s]everal factors militated against effectiveness of mediation. Firstly, the mediators themselves received only limited training given during the four sessions, …[they] were third year psychology students, and were therefore inexperienced as teachers in general, [and]…the mediation was provided to groups of subjects as opposed to individually (pp. 49-50).

She takes cognizance of the results of mediation in the work of Hoffenberg and acknowledges that in Hoffenberg's study "…mediation was not effective in significantly improving performance" (p. 49). Nevertheless, improvement from pre-test to posttest did yield improvements in task performance; and extended exposure to the tasks yielded better performance for those who did receive the added mediation as opposed to those who did not. The group administration of the LPAD is useful in that individual training is still at this point in time a burden and is costly. The main conclusion is that the tests used before the study in selecting potential students did not do as good a job before the introduction of the new tests of the LPAD.

4.3.4 Effectiveness of Feuerstein's learning potential assessment device in a South African context - Master's study conducted by D.A. Murray (1988)

4.3.4.1 Introduction

A study conducted during 1987-1988 using Feuerstein's Learning Potential Assessment Device in testing potential among Indian and Coloured teenagers in South Africa of high and low socio-economic status, concluded that the concept of mediated learning experience can provide a fertile area of research and practice in terms of assessment and enrichment of disadvantaged groups in South Africa (Murray, 1988). The study formed part of ongoing research at the University of the Witwatersrand at the time. Murray highlights the "ambiguous social identity" of the Coloured community in South Africa during the Apartheid years and the inclusion of this group in dynamic assessment studies attests to their marginal status within society.

4.3.4.2 Research design

In order to interpret the results of the LPAD, a set of independent tests was also used to measure the purported aspects measured by the LPAD. This served as a reference point, from which to more adequately judge the results of the study:

> To determine the extent of transfer of learning on the LPAD to comparable tasks, four independent, external measures were administered….(namely), Raven's (1958) Standard Progressive Matrices (RSPM), Equivalent Complex Figure (ECF), Coding subtest (COD) (Wechsler Intelligence Scale for Children – Revised, WISC-R) and Similarities Subtest (Sims, WISC-R) (Murray, 1988, p.18-19).

The experimental groups outperformed the respective control groups in terms of scores (Murray, 1988). An interesting finding in this study showed that the child's academic standing was important and a good indicator of how well he/she would perform in LPAD and transfer tests. Academically superior children benefitted more from exposure to the material than did those less superior academically performing children. One of the more striking statements made by Murray is that of the potential invalidation of the Raven's Standard Progressive Martrices (1958) (which was used in parallel with the LPAD in order to ascertain similarities and differences in test performance, assuming that the results would be valid and reliable – the RSPM being a validated and reliable source of non-verbal reasoning and of intelligence). "From another viewpoint, findings on the RPSM suggest that even culture-fair tests such as this are unlikely to be reliable or valid measures of potential and that, conversely, appropriate training on the measures, particularly when a mediational approach is used, will serve to better reflect individual potential" (Murray, 1988, p.33). The similar sentiments by Gewer (1998), who also has reservations about the RPSM being culture-fair, are noteworthy.

4.3.4.3 Main findings

The main finding of the study was that academic performance "…was found to be the most significant variable differentiating among subjects in this study" (Murray, 1988, p.32). "…if one considers that early provision of MLE [mediated learning experience] is important, a qualitative analysis of the LPAD results shows the lack of MLE is manifest in cognitive dysfunctions within the mental act" (Murray, 1988, p.34). The study also found that those students who performed better in both the control and experimental groups on the pre-test section, ultimately benefited more from the mediation experience than did those who scored less highly on the pre-test. In this regard Boeyens states that "[i]f bright subjects do not have sufficient room for improvement, their learning potential will be underestimated" (1989a, p. 44).[22] According to Babad and Budoff (1974) (as cited by Boeyens, 1989a, p.127) no mean differences in learning potential exists between subjects of higher and lower ability (Boeyens, 1989a, p.27)

Following from the model suggested by Feuerstein (1980, as cited in Murray) in which a mental act consists of three distinct phases, input, elaboration and output, Murray (1988) states that both the pre- and posttest low economic status groups tended to "…deal with tasks in an impulsive unsystematic manner" (p.35) during the input phase. Most of these students were unwilling to ask for help and generally responded with apathy. Once students received mediation on the posttest however, an improvement was seen in terms of less impulsiveness and more clarity regarding test requirements (Murray, 1988).

The most significant finding in the elaboration phase was the often "…inadequate perception of the existence and definition of the problem [on] hand. [Low economic status subjects] lacked spontaneous comparative behaviour … which is an essential prerequisite of any cognitive process" (Murray, 1988, p.36). Lastly, the main finding from the output phase of this study, showed that the experimental group benefitted more from meditation than did the control group. They were able to compare elements in a verbal modality, understand the symbolic nature of words, and also appreciate the differences between connotation and denotation of words. They were also able to express themselves better than the control group when it came to justification of responses.

A finding that was expected by Murray, was that the higher achieving academic group was better able to benefit from mediation than the control group. It seems that academic status played a greater role in predicting success from mediation than did group status (either belonging to the experimental or control group). It would then seem, that on the basis of these findings, intellectual ability (innate ability) contributes to the level of potential within individuals.[23]

4.3.4.4 Conclusion and recommendations

Murray maintains that mediation provided at a later stage in life may not be as beneficial as mediation given at an earlier stage in life, since cultural deprivation along with lower socio-economic status may hinder the learning process to such an extent that any attempt to curtail this "loss" may be ineffective (Murray, 1988). A few suggestions from Murray (1988) concern the need to study the effects that mediators have on the testee, namely, allowing for the mediators to be of the same cultural group of the testee (which may help to lesson the anxiety levels and alienation effects felt by testees when aided by mediators from different cultural backgrounds), as well as to explore the potential of

22 See De Beer (2000) on the issue of ceiling effects and how the LPCAT is able to be utilised for both lower and higher performing students. Boeyens does state that in order to overcome the so-called ceiling effect, a method would be to construct items with predetermined levels of difficulty so as not to overwhelm higher performing students with inordinately difficult items (1989a).

23 One must be careful here not to confuse terminology. Academic ability and innate ability are not the same construct. If no opportunities are afforded one in life, great intellectual ability may never be seen in the academic context. Murray does equate intellectual ability with academic aptitude (pp. 39-40). However often the two are correlated in studies, **they are not one and the same concept**.

investigating the interpersonal dynamics involved in the mediational process itself. The need to study the effects of mediation over a long term should also be looked at within the South African context as well as mediating in the home language of the respondent (Murray, 1988). Also, studies delineating results from mediated efforts in both rural and urban set-ups may prove to be fruitful.

One of the most important conclusions is that "[f]ailure to perform an operation does **not imply an inability to acquire the operation** [own emphasis] (Murray, 1988, Appendix IV). Murray emphasises the differences between low levels of efficiency owing to skills only recently gained or learned and low efficiency due to poor learning potential. Efficiency is characterised by the amount of effort needed to carry out mental acts which will vary based on the familiarity with the tasks. Future research will need to take these differences into account.

4.3.5 Learning potential: A theoretical perspective - a study conducted by J.C.A. Boeyens (1989a)

4.3.5.1 Introduction

The theoretical investigation by Boeyens (1989a)[24] is more of a treatise on learning potential than an actual experimental design study. It serves as the precursor to the study which follows this study emphasising the empirical basis of learning potential. The theoretical departure functions as a broad-based introduction to the later study (see Boeyens, 1989b). Of importance to Boeyens is the insight delivered by Vygotsky, namely that waiting for biological "readiness" is not the view to be taken as propounded by Piaget. In essence, the notion of biological maturation and mental development is, in fact, fundamentally questioned. "…the assumption that ability reaches a plateau in all individuals by young adulthood may be invalid. Learning potential seems to be a more useful concept than traditional ability. The assessment of learning potential [however] poses problems of its own" (Boeyens, 1989a, Abstract).

Boeyens surveys the main debates of the day, namely IQ versus learning potential, intelligence as fixed property within the changing individual, as well as delineating briefly the slow movement from conventional psychometrics to the notion of learning potential. At the time of writing his study, the only research in this field as applied to tertiary education which had been conducted to his knowledge was that of Shochet (1986):

> An educational institution using previous academic performance as a selection criterion usually makes the assumption that all candidates have previously been exposed to comparable educational opportunities. Psychometric tests of cognitive ability largely assess the effects of previous educational exposure. Basing selection decisions purely on cognitive test scores is therefore little different from basing these decisions on previous academic performance (Boeyens, 1989a, p.2).

These sentiments have already been uttered by numerous South African researchers, both prior to and after his study.

Boeyens states further that if present academic performance is assumed to indicate future performance, many disadvantaged students will be considered unsuitable candidates for tertiary education programmes. Practitioners, he adds, are caught in a double-bind situation, in which though more attention is being paid to dynamic models of assessment, practitioners are forced to rely on static measures of assessment. The two paradigms clash but often there is simply no alternative but to make use of static measures.

---

24 Here there is a slight deviation from the stated progression for the Boeyens' 1989a study. The theoretical research forms the foundation for the 1989b empirical research and as such only an introduction and research design will be looked at. The main findings and conclusions and recommendations will form part of the 1989b empirical study.

One of the main concerns in the literature today is the somewhat nonchalant view taken of learning potential tests regarding validation and reliability studies and also that the realm of learning potential research seems to be relegated to retarded children, adolescents and children from disadvantaged communities, where the opportunity for learning (as well as mediated learning) have not been afforded. Another question to be asked is what is the potential value of using learning potential tests for adults who have been educated in a largely western education system? Research of this nature is not really the focus area, but certainly does merit consideration.

“The number of problem solving strategies available to the individual is dependent on the amount of learning that has preceded the actual doing of the test” (Boeyens, 1989a, p.10). This gives rise to two important questions, as Boeyens notes. Firstly, is intelligence a precondition for an ability to learn? and secondly, how is this notion of intelligence measured?

In a situation where different children have experienced a similar socio-economic environment, similar educational upbringing, and share similar experiential world views, traditional IQ tests - which are thought to measure what has previously been taught in traditional western situations and the ability of children to retain the principles of what has been taught may afford a measure of how the child is currently performing on these limited and specific criteria. However, the case is not quite as simple as that as time has shown. Although previous research into learning potential tests in the South African context decries the now defunct segregated system, the effects, although not as immediate and pressing as they were ten years ago, are nevertheless reverberated in our current society. "It is argued that although a new democratic government was elected in 1994, the legacy of the previous segregated education policies still places many students at a disadvantage" (De Villiers, 1999, p.3). This notion is echoed by De Beer (2000) who also states that despite the new dispensation and the changes that have taken place "… since the first democratic election of 1994, the **effect** of the **history** of segregated and unequal living and educational conditions **will affect people for many years to come**" [own emphasis] (p. 64).

De Villiers highlights this retrospective argument further by stating that "[t]he fact of the matter is that the situation in black education has not changed much since the days of apartheid" (1999, p.8). Children raised in a segregated system who are now adults functioning in a democratic society were not afforded the opportunities now present. So a more relevant questions is: what potential resides in these individuals?

As environmental factors may impinge on individuals’ IQ scores, Feuerstein’s and Vygotsky’s conceptualisation of resident potential within all humans and the role of the internal as well as external environment are of importance.

The very nature of the name of Learning Potential Assessment Device signifies that the emphasis no longer resides with "testing" and "measurement" in the conventional sense of the word, but emphasises a supposedly hidden aspect of intelligence which may not otherwise be detectable. This "hidden" facet of intelligence is not directly seen and not found through conventional question-answer techniques. The scenario is less linear in nature and more systemic in its conceptualisation of how potential may be discovered. Boeyens mentions metacognition as another concept which has not always been considered essential to cognitive functioning within static tests. Metacognition is seen to be a reliable index of learning potential (Boeyens, 1989a).

“Metacognition may be defined as ‘knowledge or beliefs about one’s own cognitive processes’. This knowledge may also be used in the regulation of cognitive activities” (The Blackwell Dictionary of Cognitive Psychology, 1994, pp. 225-226). It also includes the conscious monitoring of progress in learning a task (Craig, 1989). Individuals may not always know the full extent of their own cognitive abilities and so need to be co-discovered with another individual. However, Christmann and Groeben (in Valsiner & Voss, 1996, p.55) criticize the somewhat vague distinction between cognition and metacognition. The same principle is used in assessing learning potential by aiding in a task where necessary and leaving the individual to do the next task. An example of metacognition is the ability to transfer a principle learned in one task to another task requiring the same principle but in a different context. Knowledge of one’s metacognitive strategies is a measure of learning potential, “[t]he metacognitive differences between truly retarded children and learning disabled children can be

labelled as differences in the ability to benefit from instruction or intervention. Truly retarded children have very limited learning potential, whereas learning disabled children do not" (Boeyens, 1989a, p.23).

"Learning disabled" must not necessarily be construed as meaning inherently disabled but should also be seen as the result of a deprived learning environment compounded by cultural deprivation. The fact that learning disabled individuals or those who have not had proper access to the education system are nevertheless seen as prime candidates for learning potential assessment interventions, is highlighted by research conducted by Budoff and Corman (1974) as cited in Boeyens (1989a), who found that learning potential is related neither to socio-economic status nor to race, whereas static tests typically correlate with these variables.

Boeyens highlights four approaches to learning potential, namely insights taken from meta-cognitive research (as discussed above), **neo-Vygotskian approaches**,[25] **studies in Russia** (having revealed that the zone of proximal development for learning disabled children is not quite as extensive as the zone for retarded children); that of **Budoff**, whom Boeyens states did not make use of Vygotskian ideas but instead merely sought to look for alternative measures of intelligence instead of making of use of conventional testing, and lastly, that of **Feuerstein,** who discusses the four categories of mental deficiencies and how the LPAD can be used to identify potential for improvement.

Implications for Boeyens' theoretical study include the fact that the learner is able to make use of new knowledge by utilizing metacognitive strategies, the fact that the zone of proximal development can be used to identify retarded as well as learning disabled students, and lastly, that though conventional tests may be able to discriminate between socio-economic class this is not the case with learning potential assessment.

Notwithstanding the earlier comments about the long-term effects of apartheid restrictions, Boeyens' early assumptions (though heavily reliant on the literature available to him at the time) are in general quite accurate as can be seen in the South African research after 1989. He points out an aspect of learning potential research which has been addressed by De Beer (2000), namely that of item difficulty. He states that conventional testing is atheoretical in the sense that static tests do not derive from any specific theory. However in order to determine whether learning has taken place within the learning potential approach, it is necessary to know the difficulty levels of items. "While its may still be possible to distinguish high learning potential from low potential it is impossible to **compare** the amounts of learning potential of two individuals with a sufficient degree of accuracy" [own emphasis] (1989a, p. 37). Furthermore, this issue is exacerbated by the fact that usually no item standardisation is compiled on the sample being studied. De Beer's study attempts to solve this issue using Item Response Theory (IRT) and Computer Adaptive Testing (CAT) within her learning potential test. From the literature, it can be seen that there has indeed been some progress in the field in the past decade.

4.3.5.2 Research design

As Boeyens' (1989a) study is a theoretical one, there was no experimental study as such conducted, but a theoretical learning potential test was designed. Typical of many such tests, Boeyens developed a letter sequence test relying on inductive reasoning skills and subtests such as analogies, classification tests and series completion tests. The letter series is explained and described in detail, noted for its culture-fairness and non-reliance on previous knowledge. A pre-test would be followed by a lesson and completed with a posttest. As Boeyens details the results of the implementation of his theoretical ideas in the next study (the empirical study) this section will not deal with the main findings or conclusions as there were none to be had.

25 As Boeyens states, the Neo-Vygotskian or "typical Russian" method for assessing learning potential does in fact differ from Vygotskian perspectives, i.e. Neo-Vygotskian approaches entail a process of hints delivered to the learner, after which an assessment is made based on the number of hints necessary to complete the task, as opposed to Vygotskian notions of mediation which does not implement this strategy.

4.3.6 Learning potential: an empirical investigation - a study conducted by J.C.A Boeyens (1989b)

4.3.6.1 Introduction

Boeyens' (1989a) first report in the area of learning potential sought to identify aspects of both traditional and dynamic assessment which were pertinent to the study of learning potential, by way of offering both the advantages and disadvantages of the use of learning potential assessments in the field of academics. The context for his study is the area of student selections.

In his second report (1989b) he devises an experiment in which he tests certain assumptions made in his first theoretical report. The lesson devised in the previous report is made use of in this investigation namely, a letter series which has to be completed by the respondent. Of note were the findings of responses to the mediation given as part of the test-teach-test method employed. The pre- and posttests were parallel forms of the same test material.

Three groups of students emerged based on the way they dealt with and used the information gleaned during the teaching phase. One group **successfully integrated** the information into the posttest, utelising the knowledge when and where necessary. A second group **rigidly** applied what they had learned to the detriment of their answers. In other words, they made use of rules that may not necessarily have been prudent to use with each and every letter problem. A third group did **not integrate** the lesson at all. Boeyens concludes that each of the three groups should be assessed differently, and that the learning potential assessment method would perhaps be more suitable for those types of students who can integrate what has been taught to them. The difference score between the pre- and posttest is easily calculated by subtracting the pre- test from the posttest score.

The main question to be answered or hypothesis to be validated can be formulated as follows: "…it [is] predicted that learning potential scores [will] be indicative of the amount of improvement in academic performance that a previously disadvantaged individual currently exposed to an enriched educational environment will evidence during the course of an academic year" (Boeyens, 1989b, p. 18).

Once again, as other research results have shown (De Villiers, 1999; Shochet, 1986; Shochet, 1994; Zolezzi, 1992) less modifiable students or, as in this study, those students who scored below the mean in the pre-test and who consequently did not perform significantly higher in the posttest will benefit more from the traditional mode of assessment than with dynamic assessment. Static measures also predict better for this group. However, learning potential assessment, as opposed to static measures, predicted better for the group who integrated the mediated lesson.

4.3.6.2 Research design

The sample consisted of 183 black matriculation DET students (the then Department of Education overseeing the educational needs of black students). These students were enrolled in colleges which were aimed at improving their results in the matriculation examination. Ninety one students in the experimental group were administered the pre- and posttest with mediated lesson, and ninety two students in the control group received the pre- and posttest without the mediated lesson.

Boeyens made use of two scoring systems, namely a micro-item scoring technique and an item scoring technique and the main reason why this was done was because of the number of questions used in the pre- and posttest assessments. There are thirty letter strings, to which each student has to add the next three letters in the progression in order to complete the string, methods and rules need to be followed in order to complete the strings. A micro-score is given for each correct letter thus amounting to a total score of 90. In this manner, the test has a higher number of items and thus reliability is increased. The item scoring technique administers one score for each completed string. However, only two letters need be correct in order for a mark to be allocated, hence the total amounts to 30.

There were no significant differences between either group on the pre-test score. Likewise no significant difference was evidenced in the mean improvement score for either group. However, a qualitative analysis of the results showed that the experimental group was more methodical in tackling problems as opposed to the control group. The number of micro-items attempted by the experimental group was significantly lower than the number of micro-items attempted by the control group. In other words, although the number of items being answered correctly by both groups was similar (i.e. there was no significant difference), more items were attempted by the control group than by the experimental group. This underlay the different approaches made by both groups; the experimental group were more methodical than the control group.

As stated in the introduction above, three disparate groups emerged, based on the manner in which they answered the letter strings, those making use of the rules in a flexible manner, those who rigidly applied what they had learned in the lesson, to the extent that "…they applied all the strategies taught in the lesson to even the most simple first problem in the posttest" (Boeyens, 1989b, p. 9) and those who did not show any effect of the lesson. In order to more fully understand the usefulness and effectiveness of dynamic assessment for students, it is prudent to classify students further according to **how** they made use of the information in the lesson. In this manner, it is possible to identify more accurately those who can benefit from mediation and those for whom mediation is likely to make no difference or in fact may hinder their progress. (This is a common finding, where a complement of students perform better without the additional information given to them in the form of a mediated lesson, and it is these students who are better predicted for by using static [conventional] measurements as opposed to dynamic assessment.)[26] In fact Snow and Lohman (1984) were able to state that attempts to train already high-functioning students in certain ability areas in fact impeded their performance as overly explicit instruction is "dysfunctional" for this group (p. 371).

4.3.6.3 Main findings

The mean improvement scores for the three identified group evidenced the following: the highest scores were achieved by the group who integrated the lesson information into the posttest, followed by those students who did not show any effect and then, lastly, those who rigidly applied what they had learned. The differences between the experimental and control groups were significant in the mean difference scores. Boeyens states that even though the experimental group was divided into three distinct groups based on the way they performed in the posttest, "… it seemed reasonable to suggest that the lesson did have an appreciable effect on subsequent performance" (p. 10). It is interesting that Boeyens states that there is no significant difference between the **mean** improvement score between the control and experimental groups, and further on he states that there **is** a significant difference between the mean improvement scores for the control and experimental groups (those who did not receive the mediation and those who did). This is accounted for by the fact that in the latter group, Boeyens divides the sample into four groups, one group who did not receive the mediated lesson and three who did. (They were then again divided into the three above-mentioned groups.) It would seem that there is no main effect or difference between the groups and that the difference is due to a moderator effect; the moderator effect being the degree to which the lesson's mediation was used in the posttest (or in other words in the level of modifiability).[27]

Boeyens comes to the conservative conclusion that owing to the differences between the three identified experimental groups, the predictive validity of the learning potential assessment will consequently also be different for each group. Based on this, the assumption then would be that the predictive validity for the group integrating the information will be higher than for the group who rigidly applied the information. At the time of this report, as stated before, one of the very few research results that was available was the work conducted by Shochet (1986). Boeyens frequently notes the similarities and differences in results between the two studies. For instance, Boeyens equates the rigid group with Shochet's non-modifiable group of students.

26 See section 4.5.1.1 where the Pretoria Technicon has also commented on this fact.

27 It is not unreasonable to equate the manner in which the lesson was used with level of modifiability.

The test items as devised by Boeyens for his study were originally theoretical items with theoretical difficulty items. So, in order to validate the difficulty levels of the various test items empirically, Boeyens investigated the scalability of the items. He did so by making use of the SAS set of programmes which yields a " …coefficient that indicates the extent to which empirical scalability approximates perfect scalability…" (Boeyens, 1989b, p. 12).

Validation and reliability of the LP test

In order to perform the validation study on the Learning Potential Test, Boeyens worked with the scores of 202 students that were enrolled at the various colleges. Perfect scalability can be defined then as items which are theoretically very easy to solve followed by items which are more difficult to solve, as it would stand to reason that if everyone were able to answer the more difficult questions then they should likewise be able to answer the easier questions. If this were this case, then the items would be said to have perfect scalability. It may happen in practice that items which are thought to be easier than other items are, in fact, not so when attempted and hence the items are not accurately scaled. The coefficient of scalability, then, is the measure taken to investigate the correlation between the items in practice and what would be yielded in a perfect scale. Kerlinger (1981) brings attention to the relation between the pattern of item responses and total scores, "…if we know [the] … total score, we can predict [the] pattern, if the scale is cumulative, just as knowledge of correct responses to the **harder items are predictive of the responses to the easier items**" [own emphasis] (Kerlinger, 1981, p. 498).

A coefficient of 0.90 (known as the coefficient of reproducibility) is used as an acceptable measure of near approximation of a perfect scale in Boeyens' study. He also makes use of the probability of misclassification of test items as well as the reliability of the scale, taken from Proctor (1970, 1971) (in Boeyens, 1989b). According to Boeyens, a Guttman scale typically deals with the difficulties of two or three items and the SAS programme that was used in his study was able to make use of twelve levels. The thirty items in the learning potential test are thus grouped into ten levels of difficulty. The Guttman scale also works with binary digits and so one digit was allocated to each correct answer (which itself was calculated by awarding a mark to two of the three correctly answered items). Each student then had a string of ten binary digits. The results of the Guttman procedure yielded coefficients (for both the pre- and posttest results) that were above the level of 0.90, (0.969 for the pre-test and 0.966 for the posttest) thus indicting that the items used in the learning potential test were indeed highly scalable and thus reliable; "…and the empirical results were in keeping with the theoretical difficulty values of the items" (Boeyens, 1989b, p. 13). Also the probability of misclassification was very low for both the pre- and posttest results (0.021 and 0.023 respectively).

As far as reliability of the test is concerned, Boeyens calculated Kuder Richardson (formula 20) coefficients on both the pre- and posttest scores for all 202 students. The reliability scores for the pre-test and posttest were 0.96 and 0.97 respectively. The reliability of the difference score was also calculated at 0.76. Boeyens concludes that his Learning Potential Test is indeed reliable and also scalable.

In order to test the hypothesis as stated in the introduction, Boeyens had to obtain a suitable sample for the study, i.e. a sample of students who could be classified as disadvantaged in terms of the inferior education received during their schooling. The sample was obtained from a college whose aim is to aid these students, all of whom were under the old DET system. The mean age of the sample was 25.5 and education already received varied from Std 6 to Std 10. Three measuring instruments were utilised to determine intellectual ability, namely, the pre-test of the Learning Potential Test, the Intermediate Mental Alertness Test (NIPR) and the Number Comprehension Test from the Academic Aptitude Test of the HSRC. Both the latter mentioned tests are (according to Boeyens) used frequently to predict academic performance, even though "…the test's [the Intermediate Mental Alertness Test] manual does not cite any validation studies on black samples….(Boeyens, 1989b, p. 21). Boeyens offers a short description of the reliability of the tests used to determine intellectual ability which will not receive attention here. Two indices were used as a measure of academic competence, namely school results and mathematics results. Owing to the norm referenced nature of the various test results (each school teacher would measure improvement as compared to fellow classmates and the improvement by means of test scores throughout the year would not be an

absolute measure), a mathematics achievement test was devised (HSRC) which was administered twice in the year, once at the start of the academic year and then towards the end; and an index of improvement was calculated as a difference in scores between the two scores. The mathematics test scores were seen to be more reliable than the academic scores over the year as " … mathematics success depends upon a combination of knowledge, present problem solving ability and learning potential" (Boeyens, 1989b, p. 25).

The entire sample of students (N=202) received the above-mentioned instruments as tests at the start of the academic year, whereas only the identified sub-sample (N= 183) received the Learning Potential Test.

Five investigations were conducted;

1. An investigation into the assumed positive relationship between learning potential and academic performance
2. An investigation into the assumption that learning potential scores correlate with mathematical competence more so than do ability scores
3. An investigation into the relationship between learning potential and traditional ability test scores
4. An investigation into the differences between ability, learning potential and school marks
5. An investigation into the comparison between these results and those of Shochet (1986).

The relationship between learning potential and academic performance

No significant findings were shown for the correlations between the mathematics and learning potential scores, although the correlations were higher for the posttest and mathematics improvement scores as opposed to the pre-test and mathematics improvement scores. The hypothesis mentioned above did, however, state that there should be a positive relationship between the two, which there was. Given the three identified experimental groups, Boeyens decided to investigate the degree to which each of the three groups would correlate differently with the learning potential scores. Unfortunately, only 40 students completed the mathematics test, thus making the sub-samples even smaller (to the extent that no worthwhile analyses could be conducted with the rigid group and the group which did not apply what it had learned). However, Boeyens did conduct analyses for the group which did apply what it had learned (N=29). However, these univariate figures were not much different from the entire sample's (N=40) results. When the correlations are investigated, however, a significant correlation was found for the learning potential score and the mathematics improvement score (0.47). Hence "…it can be stated with a fair degree of conviction that a significant relationship between learning potential and ability to benefit from instruction has been demonstrated in those pupils who participated meaningfully in the learning potential testing session…" (Boeyens, 1989b, p. 35).

The relationship between traditional and learning potential tests and mathematical competence

Boeyens ran a series of correlations between the ability scores and mathematics scores, as well as correlations between the learning potential scores and mathematics scores. The underlying assumption which prompted these calculations, was that current ability and learning potential were not the same thing. Current ability scores do not give any indication of future potential scores. Thus mathematical potential should be correlated with learning potential measures more so than with traditional ability tests. Findings showed that neither of the ability tests (Mental Alertness Test nor the Number Comprehension Test) correlated significantly with the mathematics score. The Learning Potential Test, however, did correlate significantly but only in the group which integrated the lesson (N = 29). Owing to the difference in the samples, Boeyens ran correlations for those whose -0.01 correlation was obtained for the Number Comprehension Test and mathematical competence was insignificant and for those whose 0.47 correlation for the learning potential test and mathematical competence (only for those students who successfully integrated the lesson) was significant. If there had been a difference between these two then it would have proved that the learning potential test was better able to predict mathematical competence than the traditional ability test. Boeyens made use of a computer programme developed by Brown (1977) cited in Boeyens (1989b) which tested for this. The results showed that the difference was indeed significant, thus proving the assumption correct for the second investigation (also given the fact that the correlation for the Mental Alertness and mathematical improvement was even lower than the Number Comprehension Test).

The relationship between traditional ability and learning potential

According to the assumptions about learning potential and static measures, learning potential tests are supposed to measure potential, and static tests current level of ability. According to the correlations run by Boeyens, no significant correlations were observed between the static and dynamic measures. This supports the contention that the two types of tests measure different constructs. The two are thus independent. An interesting finding which is not surprising and has been found by others (De Villiers, 1999) is that the practice effect plays a role in posttest results. Boeyens found that for the control group, difference scores correlated with the pre-test scores. In other words, those students completing the pre- and posttest had already experienced the test, and hence the mere fact that they were now sensitised to it would alter the results, even though a parallel form of the test was administered.

Once again the suggestion of making use of both static and dynamic assessment in selection procedures (where predicting performance is of importance) is emphasised. This is true because of the heterogeneous populations that are encountered in admission situations. Disadvantaged status does not necessarily mean that the group shares all characteristics. There are thus relationships between static and dynamic assessment techniques but they are not significant. Learning potential assessment predicts better for those who make use of the mediated lesson content (i.e. those who are modifiable).

The investigation between academic performance, ability and learning potential

In order to assess the degree to which the learning potential test could predict academic performance, Boeyens made use of the averaged March examination results. Correlations were conducted for the Learning Potential Test, the Mental Alertness Test as well as the Number Comprehension Test. All were significantly correlated with the average March results. The next step was to specify which variable contributed more by way of predicting the March outcome and this was done by running a stepwise multiple regression analysis. The results of this showed that the learning potential tests significantly improved the predictability of the March results. Boeyens states that the learning potential test used together with the two ability tests makes for better predictions than when only the ability tests are used.

Boeyens broke down the analysis even further by investigating the correlations of the tests in each of the three identified groups with the March results. The rigid group's numbers (N = 10) were too small for any meaningful analysis. The group which showed no effect of the lesson, did not yield one significant correlation between school marks and learning potential. However, for the group which integrated the lesson, the correlation was slightly higher than for the group as a whole and the static tests correlated somewhat lower for this group than for the whole group. Once again the results showed that the learning potential test added significantly to the prediction of the March results for this group. Hence, the assumption that the learning potential test will add significantly towards improving the prediction results has been affirmed.

4.3.6.4 Conclusion and recommendations

Comparison with Shochet's findings

As mentioned above, Boeyens compared his results with those of Shochet throughout the study, as Shochet's study was one of the very few studies available at that time. Consistently throughout the study it has been shown that the static ability tests make for better predictors of academic success for those students who were classified as low potential students. The findings of Shochet reflect a similar pattern, in which the low modifiable student's results, were better predicted by the static tests. Correlation analyses were run for both the low and high learning potential students on the school marks and the Mental Alertness Test. (Boeyens split them into these groups so as to better understand the results.) With the exception of the results for the Afrikaans examination the low learning potential group scores on the mental alertness test correlated better with the school marks than for those of the high learning potential group. "Although the differences in the correlations are mostly non-significant, they are consistently higher in the low learning potential group" (Boeyens,

1989b, p. 50). Hence the Boeyens study and the study by Shochet evidence similar results. Owing to the fact that these differences were not significant, however, a cautionary note is added by Boeyens in which he states that further research needs to look at the possible reason why this may be so.

Boeyens also warns against the broader applicability of the learning potential test, stating that the predictive validity may only be useful in certain specific subsamples of students (i.e. those who can at least make use of the mediated lesson). In other words, not all students will react in the same way to the lesson. He states that more attention could then be paid to the way the lesson is administered, i.e. a more qualitative approach to the lesson may yield a smaller proportion of students not taking in what they should and in so doing reduce the number of low potential classifications. Lacking also in the study was an investigation into the possible role played by metacognitive aspects, such as motivation and attitude.

4.3.7 An investigation of Feuerstein's theory of mediated learning experience with a disadvantaged community - Master's study conducted by S.J. Henley (1989)

4.3.7.1 Introduction

Henley's study is a mix of both qualitative and quantitative methods. Henley seeks to investigate whether Feuerstein's Mediated Learning Experience (MLE) is the underlying component of Structural Cognitive Modifiability. In an attempt to do so, a sample of 100 randomly assigned school children were placed in either a control or experimental group. The experimental group was further divided into those who had received MLE as children and those who had not received MLE (this information was obtained via questionnaires). The experimental group received mediation on Feuerstein's LPAD whereas the control group did not. All pupils were then tested for transfer measures on tests comparable to the LPAD.

Henley's 1989 study can be construed as one of the earlier studies within the field of dynamic assessment. In her introduction, Henley discusses the dual nature of testing following the arguments that most intelligence researchers follow, namely the disadvantages associated with psychometric testing and the multitudinous advantages associated with dynamic assessment. In particular, Henley discusses Feuerstein and the rationale behind the idea and development of the LPAD. Of note here is the emphasis placed on peaks of performance as opposed to merely accepting final scores on psychometric batteries, "…peaks in performance are an indication of the cognitive potential of the testee, and should not be discounted but should form a central component of the interpretation of results" (Henley, 1989, p. 15).

The sample for Henley's study is drawn from the same sample that formed the bases for both Hoffenberg's (1988) and Gaydon's (1988) studies. Three main hypotheses were postulated, namely:

- H1 - The group identified as having higher MLE (as indicated in the answers to the questionnaire) will have higher academic results as opposed to the group with lower levels of MLE
- H2 - The group identified as having higher MLE will evidence higher levels of achievement on the LPAD tasks as opposed to the group with lower levels of MLE
- H3 - The group identified as having higher MLE will have higher levels of transfer measures as opposed to the group with lower levels of MLE

4.3.7.2 Research design

One hundred pupils were chosen from a population of three hundred pupils, based on academic achievement or achievement in at least one school subject. Fifty four girls and forty six boys were randomly assigned to four groups. Ages ranged from 12 to 19 with a mean age of 14. The design followed a Solomon four-group design (see also studies by De Villiers [1999]; Gaydon [1988]; Hoffenberg [1988]). All pupils were randomly assigned to one of four groups and LPAD components were administered by undergraduate students who were trained by postgraduate students. See Hoffenberg (1988) (Table 2) and Gaydon (1988) for the sample description and design, as this was the same sample and design.

All groups were tested on posttest transfer components which were similar to the pre-test components. In essence group I received a pre-test, mediation and a posttest, group II received only mediation and a posttest, group III received a pre-test and a posttest and group IV received only a posttest. Three LPAD sub-tests were used in the study,[28] the Organiser (A), Verbal Analogies and Set Variations I and II. In order to test for transfer measures, comparable tests were used, namely the Organiser (B)[29] (comparable to the Organiser [A]), Similarities sub-tests of the WISC-R and the WAIS-R (comparable to the VAT) and the Raven's Progressive Matrices Test (comparable to the Set Variations).

Based on the research of Hoffenberg (1988), Henley made use of the LPAD scores and posttest transfer scores already available for this sample and made use of these results in her study. Pupils from the experimental group were divided into high and low achievers based on their academic results (14 girls and 16 boys) and were then interviewed from a questionnaire.[30] The pupils were interviewed "blind", so that the interviewers did not know from which group the pupils came and could thus not prejudice the results. Interviews took place over a three month period and the pupils were interviewed in random order. Only questions which appeared to discriminate between the two groups were retained. Based on these scores, two groups of pupils were identified - those who received high MLE and low MLE during their lives. Those whose scores were borderline were excluded from classification. The groups were then compared on different variables using t-tests. Variables included academic achievement, LPAD scores as well as posttest transfer measures. Henley does not make use of discriminant analysis owing to the small sample size, but rather apportions percentages to both achievement groups in the high and low MLE categories.

4.3.7.3 Main findings

The main aim of the questionnaire was to "…tap early child-adult interactions for the presence of mediated learning experiences…"[31] of which 13 of the original 21 questions were retained, on the basis of their discriminatory power. Henley lists the 13 questions asked and, according to responses, pupils were categorised as either having received high MLE or low MLE. Those scoring eight or above were considered for placement in the high MLE and those scoring five or less were considered for placement in the low MLE group. Those with a score of six and seven were not included in the analyses as explained above. All four groups (high and low MLE as well as high and low achievement groups) were compared on different variables.

As mentioned previously, Henley made use of the data already gathered by Hoffenberg (see Hoffenberg, 1988); who had obtained data for pupils and their subsequent LPAD scores on all sub-tests. Henley converted the results into percentages and then summed the accumulated percentages and thus expressed these as a composite whole. The same formula was used to calculate a composite posttest (transfer) score. Looking at the results and the categorisation that took place based on the questionnaire, as well as the initial categorisation into both low and high achievement groups, Henley states that 70% of pupils were correctly classified into the respective classifications. There was a significant difference between the low and high achievement groups' academic scores, based on previous academic results. There was also a significant difference between the scores in

28 As with many dynamic assessment investigations, tests were conducted in English and Henley states this as a limitation to the study.

29 The same as the Organiser (A) but with different content, as there was no comparable measure according to Henley.

30 A questionnaire which Henley states as not being valid or reliable due to no known measure (at the time of research 1989) existing to test for MLE. Questions included were thus based on face validity and appropriateness with MLE.

31 Henley states that subjective answering of the questionnaire may have yielded answers which put respondents in a better light "…in order to gain social approval" (1989, p.59).

school performance between the high and low MLE groups (as established by the questionnaire). The first hypothesis is thus confirmed.[32]

LPAD scores for both high and low MLE groups were used to calculate t-scores in order to test for the second hypothesis. Although there was a difference between the two groups favouring the high MLE group's LPAD scores, there was no significant difference between the two groups based on the LPAD.[33] This hypothesis was thus not confirmed. However, Henley conducted another t-test to investigate any significant differences on the LPAD scores for both the high and low achievement groups. There was a significant difference. Thus it would seem that the LPAD sub-tests could not for this sample, differentiate between those pupils who had received high or low MLE. One must recall that the questionnaire cannot however be construed as valid nor reliable. "The incremental ability to benefit from mediation in relation to the amount of MLE [is suggested by these results]" (Henley, 1989, p.52).

To test the contention of the third hypothesis that the high MLE group will significantly outperform the low MLE's transfer (posttest scores) a t-test was calculated. Results indicated that there was a significant difference between the two groups. The third hypothesis was thus confirmed. The same finding resulted when comparisons between the high and low achievement groups were compared. Although confounding factors cannot be excluded, MLE does seem to play a role in the transfer effects in the LPAD scores. Henley states that because of these possible confounding effects, various possibilities may account for the high MLE group's better scores, namely:

- The high MLE group was better able to generalise and sustain what they had learned during mediation simply because they had a background more facilitative of MLE than the lower MLE group
- The high MLE was better able to assimilate the mediation than the lower MLE group, and
- That this points to the possibility that previous mediation will better prepare pupils with current mediation.

The fact that both high and low MLE groups emanated from lower income groups and were disadvantaged in terms of education, allows Henley to state that mediation can be effective for disadvantaged groups and that economic status, although disadvantageous, does not necessarily impede the success that comes with mediation. This is in accordance with Feuerstein's MLE theory. Cognitive functions can still develop even in such disadvantaged circumstances.

4.3.7.4 Conclusion and recommendations

Of course, there are drawbacks in using Feuerstein's MLE in the educational context (see Coosner 1999; De Beer, 2000 who address these issues), namely, the issues of cost and time, and the need for highly skilled manpower "…which is clearly a problem with most national education systems working to a tight budget" (Henley, 1989, p.55). She also states that the study is a primary investigation, one needing further research, although it has successfully lent support to the contention that MLE affects cognitive development (since the group identified as the high MLE group achieved academically higher results than the group identified as the low MLE group) and was thus encouraging. The findings underscore the fact that cognitive functioning is modifiable, even as late as adolescence (Henley, 1989).

32 Henley points out the fact that owing to the Solomon Four Group Design, the two MLE groups were not established as being comparable in terms of prior testing on the LPAD and the transfer measures. She states that the high MLE group may have scored higher than the low MLE group owing to other factors as well (Henley, 1989).

33 Judging by these results, Henley states that the questionnaire was limited in its ability to identify MLE that had taken place early on.

4.3.8 Alternative selection measures for university undergraduate admissions - Master's study conducted by S.A. Zolezzi (1992)

4.3.8.1 Introduction

Zolezzi's directed and focused study was conducted in 1991-1992 and thus places it within the political framework at the time - namely, the start of democratisation but before the 1994 watershed year. As with the Shochet's study conducted in 1986 and with so many other studies at the time (those mentioned up to this point) the education system, particularly tertiary educational institutions were looking for alternative measures for student selections in South Africa, in a bid to try and level the playing fields for all groups wishing to pursue further education. The present study proposed that a learning processing paradigm which incorporates learning potential would best facilitate the inquiry into alternative selection measures. "This study aimed to assess the effectiveness of both traditional and learning process selection measures among a group of both advantaged and disadvantaged students" (Zolezzi, 1992, p. iii). Furthermore he states that the dynamic aspect involved in the procedure was the modification of students' cognitive processes and hence modification in their performance, conducted within an information-processing paradigm. Zolezzi's study can be classified as representing the approach of Budoff, in that standard tests are administered in a dynamic manner (De Beer, 2000).

A sample of disadvantaged and advantaged students within the field of commerce were selected to undergo assessment at the start of the academic year and had to complete nine measures or predictors of academic success, namely, traditional measures such as school results, intelligence, home background, motivation and inductive reasoning. Four learning processing measures comprised study processes, learning and study skills, learning processing and learning potential. These test measures will receive brief attention below.

The main findings of the study yielded the inadequacy of using traditional measures as predictors of academic success for the disadvantaged students. However, the matriculation results as well as the test of intellectual functioning were found to be adequate predictors for the advantaged students. The notion of modification of students was supported by a moderator effect for the disadvantaged students yet the best predictor for both groups was the learning process measure. Zolezzi concludes in his introduction that selection procedures and measures should include a module of learning potential measurement, or as he states include the "learning potential paradigm" (p. iii) into academic prediction and to move into the area of modifiability as means of assessment.

Zolezzi states that research prior to his study (1992), namely, the research conducted by Shochet (1986) and Boeyens (1989b), did not investigate the metacognitive aspects of learning within the learning potential approach. However, the De Villiers (1999) study did not only take cognizance of metacognitive aspects (1999, pp. 96-99) but also of non-cognitive aspects. Shochet's 1986 study does however mention and discuss metacognition (p. 144 - 147) but Shochet at the time in which the study had taken place was hesitant to draw too many conclusions from the rather scant data around at the time on metacognition, and in terms of discussing the influence of Feuerstein, Shochet states the following, "[i]t would **appear** therefore that although there is **little research evidence**, a theoretical confluence of Feuerstein's tenets and development in learning strategies and **metacognition suggests** that it **might** be useful to assess students' awareness into *[sic]* their own thinking or into *[sic]* their own strengths and weaknesses as an adjunct in predicting university success" [own emphasis] (1986, p. 146). In very similar words, Zolezzi states "[i]t would appear that a confluence of Feuerstein's tenets and recent developments in learning strategies and metacognition would enhance prediction of successful university students. An assessment of a student's awareness into *[sic]* their own thinking is also seen as a useful adjunct to selection … " (1992, p. 12). Dynamic assessment ideally would be person-specific but practically, a standardised dynamic assessment would be of more use, "…a standardised dynamic assessment instrument with domain-specific skills appropriate to selection of commerce students in particular" (Zolezzi, 1992, p. 12).

Zolezzi tested the following three hypotheses:

- H1- Learning potential is a better predictor of academic success for disadvantaged students as opposed to a traditional static conventional intelligence test
- H2 - Learning potential is also a better predictor of academic competence as opposed to school marks for the same disadvantaged students and, lastly,
- H3 - Learning potential measures as well as learning process measures are together better predictors of academic competence for both advantaged and disadvantaged groups as opposed to either the learning potential measure or the static measure when used alone.

4.3.8.2 Research design

Perhaps the most striking aspect of the sample is the very small size, only 26 students are selected for the study. These students were among those selected for a bridging programme at the University of the Witwatersrand. Age and gender are not included as predictor variables in this study. The sample was adequately divided in to male and female with slightly more males than females. Zolezzi makes a clear differentiation between advantaged and disadvantaged students, namely, that those students who matriculated under the then DET (Department of Education and Training) were to be classified as disadvantaged and those matriculating under any other system were deemed advantaged. Although this is a reasonable course of action to follow, it does not necessarily mean that each and every student passing the matriculation examination under the DET was disadvantaged. However, Zolezzi does state that "[a] primary research question of the present study is whether the categorisation of disadvantaged and advantaged meaningfully distinguishes between the two categories in relation to academic prediction" (1992, p. 15).

Static instruments used in the study

The five static measures that were used were:

- A Biographical Questionnaire (BQ) - which assesses the level of disadvantage and qualities of the student by making use of sub-scales of language and disadvantage. The qualities are ascertained by level of motivation and career suitability, once again measured by sub-scales
- The Mental Alertness test (MA), also employed by De Villiers (1999) - which measures general reasoning abilities and is a group administered test
- The Pattern Relations Test (PRT/T), which measures inductive reasoning and reasoning by analogy (p. 17), abilities which are useful predictors for studies in the commerce faculty. Zolezzi states that this test is similar to but more advanced than the RPM (Raven's Test)
- The Interview Measure (IM) was an interview conducted by an ex-student who had previously been part of the programme, and can be seen to be as complementary to the BQ. It was a semi-structured informal interview conducted face-to-face each student. The IM looked at the same three measures as did the BQ. It was left up to the interviewer's discretion as to the suitability of each student. Zolezzi conducted inter-rater reliability scores for the IM and BQ as they were meant to be complementary, and also because they were used by the same rater. A Pearson correlation of 0.56 was calculated for the inter-rater reliability score. However, a limitation to this study was the fact that the interviewer had to be trained, since a high degree of skill is required to administer the interview as well as to conduct the mediation lessons (Zolezzi, 1992)
- School results (MATRICULATION EXAMINATIONS) - using these results is very much the same system as used in the De Villiers study for the same criterion, namely a point system. Each subject is allocated points based on the level at which he/she passed. This measure accounted for academic achievement.

Dynamic instruments used in the study

The four static measures that were used were:

- The Pattern Relations Enriched Condition (PRT/E) - which determines learning potential. This test was used during an initial assessment period (the PRT/T) after which the PRT/E was used. Zolezzi states that this is in keeping with the Feuersteinian idea of pre- and posttesting, mediation

and measure of modifiability. The difference between the PRT/T and PRT/E was a measure of mediation students received in the latter as part of the procedure

- The Learning Process Measure (LSP) - which consists of students summarising a text in the field of commerce and assessing answers given after questioning. Zolezzi states that this yields information as to how students go about learning information, the mere reproduction of material or transformation of material
- The Study Process Questionnaire of Biggs (SPQ) - which establishes learning strategy (used by De Villiers, 1999) and
- The Learning and Strategies Inventory (LASSI) - which determines study processes and focuses on both "…overt and covert thoughts and behaviours which can be altered though educational interventions" (Zolezzi, 1992, p. 23).

All students received the same tests in the same test sequence, which itself was divided into two sessions, namely a static measures session and a dynamic measures session. The LSP was evaluated by a post-graduate student in the same faculty as well as a second rater and inter-rater reliability scores were calculated to ensure adequate reliability of interpretation by the evaluators. Pearson correlations yielded a score of 0.87 for the inter-rater reliability of the LSP.

The process of administering the PRT/T and PRT/E was dynamic in nature. The PRT/T score was used as a static measure of intellectual functioning followed by four stages of mediation upon which the PRT/E measure was taken. Zolezzi states that the difference between the PRT/T and PET/E can be construed as learning potential and thus becomes another predictor variable. The order in which items were assessed in the enriched condition was to cluster them according to the strategies (rules) needed to solve them. This method, according to Zolezzi, satisfies the Feuersteinian requirement of allowing testees to transfer the rules learned in the traditional session of the PRT/T. Zolezzi further states the criteria according to which mediation should take place, and places a lot of emphasis on Feuerstein's approach and theory of (p. 26). However, a limitation of the study is the fact that Zolezzi was unable to identify the way in which students profited from the mediation process, or the manner in which they responded to it. This would, of course, have yielded pertinent information regarding the needs of different students. The mediation process is also seen as too short to derive any valid data from the learning processes students used (Zolezzi, 1992). Intense mediation over a longer period of time is the ideal.

The design of Zolezzi's study consisted of subject variables (advantaged and disadvantaged students), criterion variables (academic achievement at university as well as mid-year examination results which resulted in five sets of marks being used) and predictor variables (predictors generated from the two measurement sessions, together with the school results and interview results, i.e. the dynamic and static test results). The examination results in July were those marks obtained for four courses, namely, business studies, mathematics, accounting and statistics as well as an average result based on all four subjects. It must be noted that the predictor variables were made up of subscales (as mentioned) and the total number of variables thus amounted to 27. These were then correlated with the five criterion variables. Zolezzi made use of the SAS software programme to compute the various statistics involved in such a correlation analyses. Pearson Product Moment Correlation coefficients were also conducted for all measures. Results were divided into correlations for advantaged and disadvantaged students, as well as for the whole group, and for low and high learning potential students. The results were then presented according to each group's scores.

It must be borne in mind that in separating the subsamples as Zolezzi did, very small sample sizes emerge:

- Whole group - N = 26
- Advantaged students - N = 8
- Disadvantaged students - N = 18
- High modifiable students - N = 9 and
- Low modifiable students - N = 17.

The small sample size did have another effect on the statistical procedures used to analyse the data. "…the statistical technique was limited to a correctional analysis because of the small sample size and large number of predictor variables…[r]egression analysis would be useful…[h]owever, the present sample was too small to allow for this type of analysis…" (p. 63).

4.3.8.3 Main findings

Whole group results

According to the correlation results, the LSP correlated with the June Mathematics (0.55) examination as well as with the June average result (0.40) at $p < 0.05$. The first LASSI subscale (noted by Zolezzi as the "positive attitude towards university scale" [p. 35]) correlated with the June Statistics examination (0.45) as did two subscales of the SPQ, (B4) 0.42 and (B8) 0.43 respectively with the June Business examination. None of the traditional results (MATRICULATION nor IQ) correlated significantly with any of the criterion variables for the group as a whole. It is assumed though, that based on results thus far surveyed, the traditional scores would tend to correlate more so for the advantaged group as opposed to the disadvantaged group. The appended intercorrelations, in Zolezzi's study, between the predictor variables yields a number of significant correlations, the primary being, B6 subscale (measuring an achievement strategy) and L6 (time management) correlated at 0.61 and B4 (deep strategy) with L6 (information processing) correlated at 0.66.

Advantaged students

The main finding for this group is that Business studies correlated significantly with:

- IQ (0.92)
- Two SPQ scales surface motive and surface approach to studying (0.86 and 0.76).

IQ correlated with the June average score (0.75). The LSP correlated highly with the June Mathematics and average scores (0.9 and 0.72 respectively), a subscale of the LASSI, namely attitude and interest correlated significantly with June accounts, Statistics and Average (0.76, 0.76 and 0.80 respectively). A subscale of the SPQ, surface approach also correlated significantly with the June results in Statistics, Accounts and the Average (0.82, 0.95 and 0.88 respectively).

The intercorrelations revealed significant results for the matriculation result and IM (interview measure), surface approach with IQ, and many other important correlations. It is assumed that these intercorrelations, as with the correlations, will yield different results for the disadvantaged group.

Disadvantaged students

The most obvious correlation which showed not only low correlations with the criterion variables but negative correlations, was the correlations between IQ and all the criterion variables, evidencing negative correlations with three out of five. However, the matriculation examination score did correlate at 0.43 with the June Accounting examination results. The Pattern Relations Test (PRT) correlated with the June Accounting results as well. However there are essentially no predictors for this group. The one significant relationship between a criterion, namely Business studies was the B4 subscale of the SPQ (denoting a deep strategy to learning). There were a number of intercorrelations between the various variables, among them an achieving strategy and time management (0.57) as well as an achieving strategy with information processing (0.56). The assumption of static measures not correlating with the criteria has been supported.

Low modifiable students

Zolezzi identified low modifiable students as those who scored below the mean on the LP test (i.e. those who did not benefit much from the mediated lesson given after the PRT/T was administered, after which the PRT/E was administered). The number of disadvantaged students falling into this group was 11 as part of the total number of 17. It must be pointed out here that low modifiability may be due to the fact that scores on the pre-test were already initially high and hence there was no room for improvement left. Moreover, this did not indicate that the students were low on the modifiability measure *per se*, but that modifiability on a particular test was not evidenced. Zolezzi's definition of modifiability thus differs slightly from the usual definition where modifiability is usually seen to represent a general level of modifiability.

There were a number of correlations for the predictor variables and the criteria. Most notable among them were the following:

- 0.48 between a deep approach and Business Studies
- 0. 74 and 0.52 between Matriculation marks and June accounting and June average
- 0.50 and 0.48 between IQ and June statistics and June Mathematics
- 0.48 between the static PRT/T and June Mathematics
- 0.60 between LSP and Mathematics
- 0.60, 0.58 and 0.52 between a LASSI subscale, attitude measure and June accounting, statistics and Mathematics and
- 0.56 between LASSI subscale L7, selecting main ideas and June accounting.

High modifiable students

Zolezzi classified these students according to their scores above the mean for the LP test, i.e. those who benefitted from the mediated lesson given during the PRT/E. Of the nine high modifiable students, seven were disadvantaged, "… demonstrating that this latter group benefitted the most from the period of assisted instruction" (1992, p. 47). However, it must be remembered that the sample was small and that there were more disadvantaged students in the sample in the first place. This was perhaps the most interesting group in terms of correlations with the criterion variables. Firstly, the only significant correlation was between attitude (LASSI subscale L1) and June statistics (0.69). Traditional measures such as IQ and the Matriculation examination results showed a negative correlation with all the criterion variables. This included the static PRT/T measure. "It appears that the more modifiable students do not predict on their manifest level of functioning" (Zolezzi, 1992, p. 51). In Zolezzi's summary of the findings, he states that of all predictor variable correlations, the one appearing most often is attitude, a subscale on the LASSI.

Summary

The highest and lowest correlations are summarised for all four groups:

- Whole group highest, 0.56 between LSP and mathematics
- Whole group lowest, 0.38 between PRT/T and accounting
- Advantaged group highest, 0.95 between surface approach and statistics
- Advantaged group lowest, 0.72 between learning process and June average
- Disadvantaged group, 0.47 between deep strategy and business studies (the only significant correlation for this group)
- Low modifiable group highest, 0.74 between matriculation and accounting
- Low modifiable group lowest, 0.48 for all three, namely, deep approach and business studies, IQ and mathematics as well as PRT/T and mathematics
- High modifiable group, 0.69 between attitude and statistics (the only significant correlation for this group).

Overall, the students performed the best on the Business studies, followed by Mathematics, Statistics and lastly Accounting. Even though the advantaged group outperformed the disadvantaged group on all the criterion measures, the high modifiable students from the disadvantaged group outperformed the low modifiable students. In general, there were also more predictors for the advantaged and low modifiable groups as opposed to the disadvantaged and low modifiable groups.

4.3.8.4 Conclusion and recommendations

According to the results of this study, all three hypotheses were supported. However the first two hypotheses were supported through a moderator effect and not through a main effect. As Zolezzi points out, the more modifiable the students become the less applicable are traditional measures as predictors and vice versa. In other words, manifest predictions decrease with greater modifiability (low correlations for PRT/T) but increase with less modifiability (with PRT/T and MA). "The present study was able to enhance prediction for academic success by supporting H1and H2 through a moderator effect" (Zolezzi, 1992, p. 55).

The third hypothesis is supported, as the results show that the learning process measure predicts better than static measure for all groups. However, it must be considered that since the sample has

an uneven proportion of disadvantaged and advantaged students (18:8), and seeing as the disadvantaged students' academic performance is better predicted by dynamic measures, it would stand to reason that the entire group's academic success would also be better predicted by dynamic measures. If the sample had been more evenly divided, the outcome may not have been as generally supportive of dynamic assessment. Yet, despite this aspect of the research, it nevertheless supports the hypothesis that dynamic measures predict better for disadvantaged groups. Also, in support of this, even though static measures were better predictors for advantaged students, a considerable number of dynamic predictors correlated with the criterion variables. In other words, the static measures are complemented by the use of dynamic measures for advantaged students. Added to the benefits of information complemented by dynamic measures is the fact that the study yielded variables that pertain to the learning process itself, which as Zolezzi states "… is an educational-modifiable approach leading to understanding and remediation" (p. 57). The sample contained considerably more black students as opposed to white students. This is construed as a limitation by Zolezzi, who states that "…there is a need to extend the selection research in other faculties…" (p. 62).

In essence, manifest academic performance predicts relatively well for the advantaged and low modifiable groups, but this is not true for the disadvantaged and high modifiable groups. Zolezzi states that consequent to these results, alternative selection measures need to be looked at for these two latter groups. More specifically, for the group as a whole, the Learning Potential Measure predicted the best as well as a LASSI subscale-attitude. This translates into "[t]he ability to identify main arguments with supporting evidence and transform information" (p. 56) as well as the fact that "[s]tudents who have a positive attitude to university study are motivated [and] achieve better marks" (*ibid*.).

4.3.9 Relationship of static and dynamic measures to scholastic achievement of black pupils - Master's study conducted by L.E. Lipson (1992)

4.3.9.1 Introduction

The Lipson (1992) study is an example of how dynamic assessment can be used in different educational settings. More often than not, selection procedures at tertiary institutions receive a lot of attention. This is, of course, understandable, since the critical situation often encountered at tertiary institutions has to be addressed. However, the context of this particular study is at primary school level. This study sought to compare the conventional General Scholastic Aptitude Test (GSAT) with Feuerstein's LPAD and to ascertain which yielded better predictive results on a number of school subjects. School pupils from an English medium school serving a black community was used as the sample for this study. Group administration of the LPAD to a group of black primary pupils was the first of its kind to be researched in South Africa.

Lipson begins her study with a discussion on the advantages and disadvantages offered by both the psychometric model and dynamic assessment model. Lipson highlights the nature of information obtained from conventional tests which are of a general nature and are not specific, thus making the task of remediation difficult. However, when making use of dynamic assessment, one is able to identify from the domain specific nature of tasks, those areas in which learners may be struggling and thus recommend strategies for remediation. She points out the South African context, wherein the majority of children are not exposed to educational resources necessary to complete schooling successfully, and notes that "[a] depressed test score on such a [psychometric] test by members of disadvantaged communities may simply reflect the lack of learning opportunities, rather than any underlying cognitive deficit or low level of acquired knowledge (Lipson, 1992, pp. 3-4).

Feuerstein's LPAD, Lipson stresses, moves away from identification of differences between learners to differences within learners which is, in essence, a more proactive and positive stance towards identification of learning problems. Dynamic assessment assesses intra-individual changes and not inter-individual differences (Lipson, 1992). Lipson points out the dynamic assessment is itself a learning procedure and is considered as diagnostically and theoretically superior to conventional testing. Lipson cites the studies of Boeyens (1989b) Murray (1988) as well as Shochet (1986), as examples of South African studies conducted in the field of dynamic assessment. She is aware of the

fact that dynamic assessment has not yet progressed to the state of an autonomous method of assessment, and should at this stage only be considered as an instrument which complements the findings of conventional test results. Lipson states her reservations about the LPAD yielding accurate information in terms of predictions on school subjects due to the fact that the LPAD is essentially concerned with the learning process and as such places emphasis on procedural and not declarative knowledge learned in school. As in may instances, conventional tests prove better predictors of academic success than do dynamic measures.

Three predictor variables are discussed briefly by Lipson, the GSAT,[34] length of enrolment at school as well as the LPAD. The GSAT[35] is a test designed to measure academic intelligence but had not at the time been normed on the black population. Those learners' scores (who were identified as "economically disadvantaged") were to be construed only as an indication of academic intelligence. Rather than making use of conventional tests as predictors of success, "… a measure is required which can assess the child's ability to learn as an index of his/her own modifiability (Lipson, 1992, p. 15). The GSAT is divided into four verbal sub-tests (Word Analogies, Verbal Reasoning, Number Problems and Word Pairs) as well as three non-verbal sub-tests (Number Series, Pattern Completion and Figure Analogies). Lipson states that scaled scores for both verbal and non-verbal tests are provided as a total scaled score. Studies conducted by Claassen (1987) and Hugo and Claassen (1991) (as cited in Lipson, 1992) yielded predictive validities of approximately 0.50 and 0.40 for environmentally disadvantaged and non-environmentally disadvantaged students respectively. Also indicated was the correlations between DET students' GSAT scores and school subjects' results.

Three sub-tests from the LPAD were used in this study. Each task encompasses mediation and a test which assesses the mediation. The following sub-tests were used:

- Complex Figure Drawing (CFD) (see Andrews [1996]; Gewer [1998]) - this task involved learners copying a complex figure using different coloured crayons, followed by a memory test where the figure was taken from sight. The reproduction and memory phases formed the pre-test. Mediation was then administered on organisation, location and proportion of the drawing (Lipson, 1992). Following the mediation, subjects were required to reproduce another drawing by copying it and then again to reproduce it from memory, which served as the transfer stage. Quantitative as well as qualitative scoring measures were used in accordance with the LPAD manual
- Numerical Progression - this task involved a pre-test in which learners were assessed on their ability to perceive numerical progression or relationships between numbers. Mediation followed in which learners were taught certain rules regarding progression and various relationships were illustrated. The posttest followed based on how the learners had integrated what they had learned during the mediation phase. Scoring was carried out according to the LPAD manual as above
- Raven's Progressive Matrices/Set Variations 1 - The Raven's test was administered as the pre-test, followed by mediation on Set Variations 1 as part of the mediation phase and then followed by the posttest phase which used the Raven's once again.

Other measures included in the study were information regarding length of stay which was obtained from a questionnaire that the pupils completed. Criterion measures included the accumulated March and July marks. The 27 pupils tested also completed biographical forms and were told why testing was to take place over the duration of four days. All pupils wrote the GSAT on the first day after which they were divided into two groups for the remaining three days. The same testers were used for both groups so as to reduce experimenter bias. The order in which the two groups were tested alternated so as to reduce any effects of time of day and hence fatigue. The following table provides a summary of the events:

34 General Scholastic Aptitude Test.

35 See Nel (1997) section 4.3.14 and De Beer (2000) section 4.3.19 who also make use of the GSAT.

Table 3

Format of testing (Lipson, 1992, p. 21)

Test	Day	Size of group
GSAT	One	27
LPAD - Complex Figure Drawing	Two	14 then 13
Raven's Progressive Matrices Numerical Progressions	Three	13 then 14
Set Variations I Raven's Progressive Matrices	Four	14 then 13

4.3.9.2 Research design

Four main questions were put forward by Lipson, questions which are answered at the end of the study:

1. Of the batteries used, which was the better predictor of academic success?
2. Did students who were mediated perform better in posttests as opposed to those who were not meditated?
3. Was there a positive relationship between length of stay at school and scores on the LPAD and GSAT?
4. Were the scaled scores on the GSAT comparable to those of the general population?

A sample of 27 children was used in the study. An equal gender distribution of children from standard four (grade six) was achieved. This school was a private English school for which entrance is granted based on a selection test. The mean age of the sample was 11 years. Criterion measures consisted of school examination results obtained at three different time intervals.[36] Scholastic achievement almost always correlates with intelligence test scores, perhaps as Lipson states "…because the tests were developed with that criterion in mind" (1992, p.12).

Testers were trained according to the principles of Feuerstein's mediation. The study was a within-subjects design and Lipson stated the independent variables as such:

1. GSAT scores
2. Pre-test scores on the LPAD sub-tests
3. The posttest/transfer scores on each of the LPAD sub-tests
4. Length of time each pupil had attended school.

Dependent variables included:

1. Each pupil's July examination result for all school subjects as well as
2. Composite terms marks for both the first and second term.

The aim was to investigate which predictors perform better on the dependent variables. For these purposes, Pearson Product Moment correlation coefficients, Multiple Stepwise Regression Analysis, a combined Regression analysis, Repeated Measures MANOVA, as well as univariate and multivariate analysis of variance were run. Correlations coefficients were obtained for GSAT and LPAD scores at

36 Note that the criterion variables "…are the school subjects assessed at three intervals: composite marks for the first and second terms collated during March and July, and a July examination mark. The intervention … took place in May… [f]or the July examination results, an average achievement score was computed, using the weighting of one-third each of English, Mathematics and content subjects… Afrikaans was omitted [owing] to possible score contamination by cultural attitude factors" (Lipson, 1992, p. 23).

three time intervals for school subject marks. Regression analyses for GSAT and July marks as well as LPAD and July marks were conducted to investigate the amount of variance explained by each battery. Combined regression was run to investigate the best set of predictor sub-tests from the two combined batteries. MANOVA on the LPAD battery was run in order to determine and pre- and posttest differences. In order to determine whether length of stay at school made any difference to the scores obtained on the GSAT and LPAD, univariate as well as multivariate analyses were conducted.

4.3.9.3 Main findings

Lipson addresses the four questions through the use of the statistical procedures already mentioned. The correlation and regression analyses attempt to answer the first question. The second question is answered by comparing the pre- and posttest scores of the LPAD. The third question is answered by investigating the length of stay at school and performance on the GSAT and LPAD. The last question is answered by an examination of the scaled GSAT scores and an investigation into the distribution of these scores. Lipson delineates her results by grouping them under each of the questions raised at the start of her study. For clarity, the same is done here.

First question - which battery is the better predictor of school subject scores?

In order for this question to be addressed, (a) correlations and (b) stepwise regressions analyses need to be conducted.

- (a) Correlations

Two sets of correlations are considered, namely, correlations between the verbal and non-verbal GSAT scores with the three time interval school results (see footnote on previous page), as well as the three sub-tests of the LPAD with the same school results. A comparison of the two yields comprehensive information. The verbal sub-tests correlate with the school results at each interval, the total verbal score being a better predictor than the individual components, correlating at 0.82 ($p< 0.01$), thus accounting for almost two thirds of the variance in the July examination results. The non-verbal total scores are less able to predict when compared with the verbal total. Of the non-verbal sub-test, Figure Analogies correlates with achievement scores ($r = 0.66$) which is better than the total non-verbal correlation (0.53).

Regarding the LPAD - Numerical Progression results, the pre-test correlations are "generally significant" (p. 26), yet lower than the verbal score obtained for the GSAT. Posttest scores are lower than the pre-test scores.[37] Lipson suggests that this may be due to fatigue. Regarding the LPAD Raven's/Set Variations, the Raven's pre-test and posttest total scores are significantly correlated with some school subjects more so than with others. On average the transfer scores (as measured by the Set Variations) are also significantly correlated with school results. Yet, Lipson states that "[i]n no case is the correlation between the GSAT… Verbal total and a subject exceeded by the correlation between a Raven's sub-test score and that subject" (1992, p. 27).

Regarding the LPAD's CFD, in general none of the pre-test, posttest and transfer scores are significantly correlated with school subjects. Lipson summarises the results of the number of significant correlations between the GSAT and school results as well as for the LPAD and school results. According to the results there are fewer significant correlations for the LPAD than there are for the GSAT. It would seem then that the GSAT correlates better with school results than does the LPAD.

37 Note that the function of the pre-test score really serves as a conventional score - a static psychometric score - and it is not surprising that the pre-test score correlates more so with school subjects than do posttest scores (which is evidenced in many studies in this survey). Once again, this highlights the fact that pre-test versions of dynamic assessments merely function as conventional tests do, and it has been shown that conventional tests yield better correlations with school subjects than do dynamic assessments.

- (b) Stepwise Regression Analyses

Lipson cautions that inferences arising from the results, due to the small sample size, stating that a maximum of two predictors can be considered valid; and where three predictors are used, the data must be duly considered. When comparing the predictability of the GSAT and LPAD results on the school results, the GSAT Verbal score, when used as sole predictor, explains 67% of the variance. "The predictability of the variability in July examination subject scores is considerably poorer when based on the LPAD" (Lipson, 1992, p.33). Of the LPAD sub-tests, the Numerical Progressions[38] pre-test is a significant predictor for four of the school subjects. Lipson's conclusions regarding these results are not too encouraging for the LPAD, although she acknowledges that the GSAT tests declarative knowledge while the LPAD does not.[39] No LPAD sub-test is as strong a predictor as the Verbal score of the GSAT. The GSAT verbal score "…corroborates the widely accepted premise that facility in language skills is the best single predictor of success in school subjects" (Lipson, 1992, p. 41). The main finding is that the LPAD is not considered a good alternative battery when compared with the psychometric alternative. Since no single sub-test of the LPAD can be considered in the prediction model, the complementary idea of the LPAD serving along with the GSAT cannot be considered at this point either. Lipson does acknowledge that the LPAD is really a measure of whether a child would benefit from mediation (which is answered below under question 2) but that this relationship with school performance is not "clearly" defined.

Second question - do pupils exposed to mediation improve in their post-mediation performance?

In this regard the LPAD pre- and posttest scores are looked at in order to determine whether there were any significant differences between them on all three sub-tests used in the LPAD. This was achieved by a repeated measures MANOVA calculation. In order to curtail regression to the mean, those individuals scoring very low on the pre-test and very high on the posttest were not included. This negates the argument of the effects of error versus mediation effects on the scores, as Lipson (1992) states. For all the sub-tests (except for the Numerical Progressions[40] and number of items memorised on the CFD), all pre-test scores were significantly lower than the posttest scores. Lipson states that although the pre-post test scores were significantly different[41] no conclusion can be drawn as to the effects being solely owing to mediation and not any other extraneous variables. This could not be investigated due to the absence of a control group which did not receive mediation (a Solomon Four Group Design). Furthermore, Lipson states that after group administration of the LPAD, it is a wise idea to complement the findings for each individual with individually administered tests in order to ascertain exactly where certain cognitive deficits occur.

---

38 "Numerical Progressions consists of a series of progressions of increasing difficulty … [and]… focuses on the search for an extraction of the specific rule governing the relationship between numbers in a progression" (Feuerstein et al., 1979, p. 165).

39 It is in instances such as these that one questions the validity of using dynamic assessment measures so directly in comparing conventional and dynamic measures in terms of predictability of school results. Early studies emphasised these types of designs. This is not an excuse readily made for the poor results evidenced in this LPAD study. It merely necessitates that a comment be made that dynamic assessment should rather be utilised in an area where testers need to obtain more information about a candidate's potential to succeed in an area of study, or study in general, as opposed to linking conventional and dynamic measures directly in terms of predictions of future conventional performance. It seems possible therefore to predict the results and usefulness of this study. What, in essence, is the use of this study? This is not a criticism levelled at Lipson but a general line of enquiry that needs to be addressed before similar studies are conducted. In order to answer questions in the field of dynamic assessment, one should at least contextualise the study in a dynamic setting as well as ask the right questions (and here the sentiments of Sehlapelo & Terre Blanche [1996] are quite pertinent, when they emphasise what Kirk & Miller [1986] label this as a Type III error which occurs when "…the wrong questions are asked in the first place" [p. 51]).

40 Due to fatigue according to Lipson (1992).

41 The purposes for which the LPAD was intended, namely, to investigate whether mediation has a significant effect on individuals' scores, have been proven. The LPAD might not measure up to the predictive ability of the GSAT, but this begs the question of whether it is truly intended for this sort of study in the first place.

Third question - is there a positive relationship between enrolment at a school providing adequate learning opportunities and performance on the batteries?

Subjects in the sample had either attended the school for 5 months or less, or for 17 months or more. Two groups were thus obtained (N = 15, N = 17). The question here is whether or not either group performed significantly differently from the other based on length of stay at the school. For the verbal scale of the GSAT, a significant difference was noted between the two groups. No significant differences were found for the GSAT non-verbal scores. "None of the F ratios investigating the interaction between pre- and posttest scores by group are significant, indicating that in no case can the difference between pre- and posttest scores be said to be different for … [the two groups]…" (Lipson, 1992, p. 37). Regarding the LPAD, there was a difference for the two groups on the Numerical Progressions pre-test, but no significant differences for the posttest scores was evident. No determination could be made for the interaction effects for the pre-post test scores as this was not considered as part of the calculation owing to fatigue as already mentioned. No significant differences were found for the pre- and posttest nor the transfer measures of the Raven's. However, Lipson notes that the five month group did receive slightly lower scores than did the group who had attended the school for a longer period. Similarly, no significant differences were found for the pre-post test or transfer measures for the CFD test. The main conclusion reached is that length of stay at school does not significantly influence results on pre-post mediation.

On the univariate levels of analyses, the GSAT Verbal scale indicated differences between the two groups. Lipson states that the sub-test responsible for the difference (Word Pairs) can be explained by the fact that since this sub-test relies on testees' knowledge of abstract verbal reasoning, a longer stay at school will result in better familiarisation of English. Familiarity with English is also necessary to perform well on the LPAD's Numerical Progression, hence the slightly lower scores for those attending the school for less than 5 months.

Fourth question - is the distribution of scaled GSAT scores comparable to that of the general population?

Scaled scores are available for the GSAT but not for the LPAD. The mean in the general population is 100 for both the verbal and non-verbal IQ. However, according to the scaled scores, 40% of subjects scored lower than 80 (the verbal mean for this sample was 83) and 30% of subjects had scores lower than 80 (non-verbal mean for this sample was 88). "…[t]he lack of fit to the normal distribution is clear…" (Lipson, 1992, p.39). Lipson states that the subjects in this sample are not representative of the sample population on which the GSAT was normed (the GSAT was not normed on black pupils) and consequently these scores cannot be converted into intelligence scores. These results partially indicate that the level of English proficiency of this group is lower than the general population and that the results of the study cannot be generalised to the population.

4.3.9.4 Conclusion and recommendations

One of the main limitations of the study is the small sample size, which affected the regression analyses and conclusions that could be drawn from the findings. Owing to time constraints, the pupils were tested and administered mediation over too short a period of time, which may have contaminated the results (see Numerical Progressions). Fatigue thus played an unwanted role in the results. The absence of a non-mediated control group is also mentioned as an inhibitory factor when interpreting the results. Lipson acknowledges the process-type of information that the LPAD taps as opposed to the product-orientated knowledge which the GSAT taps and there is a "…lack of match between the skills tapped on the LPAD and those underlying the material on which the pupils are examined" (1992, p. 45). The criterion measure thus has to incorporate a wider spectrum of thinking skills.

Although the results of this study attest to the GSAT's better predictive ability when compared to the LPAD, the notion of a changing education curriculum (see study by Coosner [1999]) necessitates a changing test of predictability as well. The LPAD can be used more effectively when "…tied to a curriculum emphasising process over product-learning" (p. 46). Evaluation should inform not only end status of projects and research endeavours but also the guiding process. This study informs this

process and although mediation was seen to be effective, the results of the study point to greater questions than merely to whether or not the LPAD can predict as well as the GSAT[42] can.

4.3.10 The effectiveness of dynamic assessment as an alternative aptitude testing strategy - Doctoral study conducted by Z.A. Zolezzi (1995)

4.3.10.1 Introduction

In his 1995 study, Zolezzi details the use of a newly devised test battery which he refers to as a "Newtest Battery". The aim of the study is to evaluate the "…effectiveness of a dynamic approach to aptitude testing…" (Zolezzi, 1995, summary) and, specifically, to position the test within the test-teach-test format. Vygotskian and Feuersteinian theory is operationalised in the battery and as such the battery assumes a defensible psychometric position, ensuring through later modifications that the test can be used as an aptitude test within the dynamic assessment field. This approach is deemed "…more equitable and relevant" (1995, summary).

As a researcher driven to understand the learning potential or dynamic assessment school of thought, Zolezzi shares his misgivings about the state of aptitude testing in South Africa and asserts that one of the main reasons why conventional tests are seen to be ill-equipped to test students in the new South Africa is that most tests have not been normed on the very groups to whom they are administered. This is not a new argument. Zolezzi states that only modest and inconsistent degrees of success have been obtained with predictive studies using psychometric procedures at tertiary level. Once again, views opposing conventional testing and espousing the usefulness of the dynamic assessment approach are proffered. And again, unfair educational opportunities are given as causes that have blighted the educational promotion of many students in South Africa, subjected to an inferior education system.[43]

Zolezzi sums up his sentiments towards group aptitude testing in South Africa: "[t]he main problem in South African testing procedures is the inappropriate interpretation of test results for a large group of testees [which] see[s] **little meaning** in being assessed, are **uninvolved** in the test process, and experience the procedure as **unpleasant and threatening**" [own emphasis] (1995, p. 3).[44] Zolezzi speaks of "identity formation" as an essential part of assessment, that is ways in which future competence can be recognised by the testee. He also recognises the fact that matriculation results are no good as predictors of academic success at university for a large proportion of disadvantaged students. Learning potential tests are able to differentiate between faulty learning styles and insufficient acquisition of knowledge, whereas conventional tests merely test for the latter. Zolezzi states that the work of Vygotsky forms an ideal base as point of theoretical departure as does the work of Budoff and Feuerstein (important in this context in South Africa), owing to their group administration approaches as well as the fact that the assessment is aimed at the culturally disadvantaged. Prediction of academic success is better measured by the gains made by students within the zone of proximal development (ZPD) as opposed to initial measures of aptitude.

---

42 See Nel (1997) section 4.3.14; Engelbrecht (1999) section 4.3.16 and De Beer (2000) section 4.3.19 who also make use of the GSAT in parallel with dynamic assessment measures.

43 This cannot be overemphasised as a hindering aspect in South African education. Almost all the studies researched for this study have in some manner criticised the inferior education system in South Africa and one finds a common thread winding through these studies, to such an extent that it does in some instances come across as repetitive. Almost all the studies cited have acted proactively within this domain by administering alternative forms of assessment or at least offering alternative solutions to the problem. It is thus motivating to witness active research in the dynamic assessment field, and not merely being confronted with platitudes about the failed state of education. Dynamic assessment is, by nature, a proactive approach towards assessment and remediation.

44 Contrast this state of affairs with the intended approach of dynamic assessment, where the testee is an involved participant in an assessment procedure which seeks to identify areas of cognitive functioning not working at maximum capacity and, in so doing, remediating and attempting to alter the inefficient functions.

Zolezzi's discussion of intelligence testing involves brief outlines of what he considers to be the four main areas of interest, namely, the psychometric approach, the Piagetian perspective, the Multiple-Intelligence perspective and the Information-Processing approach. Socio-economic differences in status as well as both the old debate of genetic versus environment issues are touched on. The adaptability position and the anti-test movement also receive attention in his discussions on the topic of intelligence. The issue of culture-fair testing is perhaps most pertinent to his study.

Zolezzi states that due to the drawbacks associated with dynamic assessment, such as the lack of reliability and validity of many dynamic assessment tests and procedures, the approach has come under fire and much has yet to be accomplished before this approach can be used as a viable option. The study by De Beer (2000) addresses these issues comprehensively. The fact that the nature of content of dynamic assessment tests often are not similar in content to academic institutions' syllabi also makes for an uneasy acceptance as an alternative procedure. However, the TTT programme[45] is one instance where dynamic assessment has included content based tests and assessment.

In the attempt to satisfy both conventional and dynamic approaches towards assessment, Zolezzi states that mediation should help in equalising the differences between students and at the same time the procedure should be psychometrically defensible.[46] Zolezzi offers a comprehensive tabulated delineation of the various approaches within the dynamic assessment approach. It is lucid and to the point, highlighting the major differences and similarities between different approaches within this heterogeneous field.[47] Drawbacks of dynamic assessment highlighted by Zolezzi include issues such as the non-standardisation of mediation, that the approach is based on a deficit model (which Zolezzi interprets as the approach being predicated on "…a lack of appropriate cognitive skills which are diagnosed during testing" [p. 52]) and that the skills which are diagnosed, bear no relationship to academic success; the lack of standard test items, the issue of ceiling effects during pre- and posttests, unreliability of improvement scores, time and labour involved,[48] the difficulty in grading mediation and generalisability to the normal population (Zolezzi, 1995).[49]

Zolezzi's new test battery is named the "Group Dynamic Test Battery (Newtest)". It is a compilation of two dynamic tests. Zolezzi's theoretical point of departure is made up of the ideas originated by Budoff, regarding standardisation of tests, the contribution of Ferrara regarding domain-specific testing

45 See section 4.5.2.1 for more detailed information on the TTT programme.

46 It is pertinent to highlight the continuum on which dynamic assessment finds itself in South Africa. The TTT programme and its derivation from theory does not allow for the concept of "psychometric" to be used when explaining the approach in the programme. The programme springs from a purely dynamic assessment approach where "psychometric defensibility" is not an issue as the two are diametrically opposing concepts. The need to define the TTT approach psychometrically is at odds with the very nature of the programme. However, almost all other studies in South Africa (within the field) state unequivocally that a dynamic assessment measure simply has to meet the requirements of psychometric guidelines. One can clearly see the continuum of dynamic assessment theory and approach here. Both positions can be said to be dynamic, but which is truly dynamic? It depends very much on the definition that one employs. The TTT programme approach (although no longer in use) may perhaps be more advanced in theory, approach and practice but this may be debatable from certain viewpoints. The De Beer study (2000) and the TTT programme (1985 - 1995) are at opposite ends of the continuum and may not entirely agree in theory and/or approach. Yet, both seem to find niches in the country and one cannot discount either approach.

47 The reader is referred to section 2.9 for more information about the different approaches within the field of dynamic assessment.

48 See De Beer (2000) with regard to ceiling effect, standardisation issues, as well as improvement scores and labour issues.

49 Zolezzi mentions the fact that no studies have yet been documented on the role of dynamic assessment in assessing average students, that is, almost all studies use samples of either gifted people or below-average people. It was the original idea of this study to attempt to devise a test which could be applied to an average performing sample. This could prove to be interesting. However, it seems that dynamic assessment for the present moment is best applied to disadvantaged students in South Africa.

and that of Feuerstein regarding the approach or testing situation. The primary objectives of his newly devised test included: among others,

- Making the test cost- and time-effective
- Allowing for equivalent opportunities to be given to all testees and so allowing them to demonstrate learning potential
- Ensuring that training and testing sessions were separate in order to determine the effects of mediation
- Ensuring that training was standardised
- Ensuring that the test was normed on the group for whom it was intended, namely, university students and that the level of complexity should be at the appropriate level
- Ensuring that random fluctuations did not account for increases in test scores
- Ensuring that, perhaps most importantly, the test be a valid predictor of university success.

Two developed tests are used, namely, the Deductive Reasoning Test and the Pattern Relations Test, the reliabilities being 0.919 and 0.812 respectively (Zolezzi, 1995). Zolezzi states that both these tests purport to measure the prerequisite skills necessary for academic success. The process of mediation was standardised and will be commented on in the next section. Zolezzi states that the learning potential score is equivalent to the second test scores (posttest) minus the first test score (pre-test), that is, the raw scores of both are taken. "Learning potential then becomes the difference between the initial raw score on the first attempt, and the raw sore on the second attempt"[50] (1995, p.62). Modification of the Pattern Relations Test was necessary in some instances. Items were reclustered, similar items were classed together (those requiring similar methods of solutions), and the format was changed somewhat. Mostly, the changes were made for the re-administration of the tests (that is, the posttests).

The main aim of the study as a whole is thus; "…[to compare] the predictive validity of traditional aptitude testing with a dynamic aptitude test procedure among prospective students at a Counselling and Careers Unit (CCU) within a traditional White university" (Zolezzi, 1995, p. 13). Three main hypotheses were tested:

- H1 - Prediction of university success will be significantly enhanced through a dynamic testing situation
- H2 - Advantaged and disadvantages students will have different predictors correlating with university success
- H3 - There is no relationship between current ability and learning potential.

4.3.10.2 Research Design

The design involved testing two different groups of both advantaged and disadvantaged students, using two different methods of assessment, namely conventional testing and dynamic assessment. Both these predictor variables were compared with one another in order to ascertain the correlations with academic success. Criterion variables such as first year academic results were obtained after the first year. The degree of educational disadvantage constituted the subject variable. Non-cognitive factors are not measured in the study, but Zolezzi acknowledges their importance.

The study took place over three years, with a sample size of 50 students.[51] The first group of N=18 were tested in a conventional manner before their commencement at university. Thereafter, the same group was monitored during their first year and end-of-year results were obtained. During this year, the second group N=32 were tested using a dynamic approach and the third year's results were used to monitor these students; their end-of-year results were also assessed. The fact that the sample consisted of mostly voluntary students, skewed the nature (or profile) of the sample since these

50 Note how De Beer (2000) compiles the difference score as opposed to Zolezzi (1995).

51 Once again as with many dynamic assessment studies, the sample size is quite small. In order to norm a new test for a certain population effectively, the normative sample size should be at least 100 strong.

students were more motivated in the first instance. However, Zolezzi states that in order to address this situation, some students were non-voluntary. The students were registered in three faculties at the University of the Witwatersrand, the Arts, Commerce and Sciences faculties, and there were slightly more females than males. The ages ranged from 16 to 20, as the non-volunteer students were students from Std Nine. Two subject variables were distinguished, namely, advantaged students[52] (having received education within the previously white education department) and disadvantaged students (having received education within the previously black education department). Zolezzi's sample is made up of 68% advantaged students and 32% disadvantaged students which is interesting as dynamic assessment studies are aimed more often at disadvantaged students than they are at advantaged.

Traditional predictor variables for the first group (N=18) included school results, Mental Alertness and Reading Comprehension sub-tests of the High Level Battery (B/75), the Standard Level Arithmetic Reasoning Test (A/131), Raven's Matrices, the Pattern Relations Test administered in traditional form as well as the Deductive Reasoning Test administered in traditional form. Zolezzi emphasises the fact that these tests which were administered to the first group, were done without any mediation and were completed within the time limits as specified in each test manual. These tests measure current levels of scholastic ability, intellectual ability, verbal, non-verbal, numerical, deductive and inductive reasoning abilities.

Dynamic assessment predictor variables for the second group (N=32) included the modified forms of the Pattern Relations Test as well as the Deductive Reasoning Test as discussed above. According to Zolezzi, the Deductive Reasoning Test is based on the principles of formal logic and "…examines the relationship between premises and conclusions of a valid argument" (1995, p. 72). The Pattern Relations Test is very similar to Raven's Progressive Matrices and is a test of inductive reasoning. The two tests thus measure potential deductive and inductive reasoning ability.

The criterion variables included year-end examination results for first year studies at the University and incorporated the November B.A., B.Sc. and B. Com. examination results.

Traditional testing took place in 1992 and followed the conventional mode of testing. Seven to eight students were tested in any one testing session. Trained counsellors were employed to conduct the tests. Dynamic testing took place during 1993-1994 and once again trained counsellors were used, but were first trained in the dynamic assessment method, and had to undergo the testing situation themselves in order to appreciate and identify with the testing method more fully. Of particular importance to the present research is Zolezzi's definition of what constitutes dynamic assessment and which can, in part, be explained by the following statement he makes about dynamic assessment, "[a] major part of training involved using the **traditional tests dynamically** [own emphasis] (1995, p. 78). This is a very important aspect in the attempt to define "dynamic assessment". This definition, then, is clearly not the same as a definition characterised by **dynamic tests used in a dynamic way.** The traditional testing sessions lasted just over two hours and the Newtest (dynamic assessment) tests lasted almost four hours. The length of time taken was due to the mediation. Five to six students were placed in groups for the dynamic testing since smaller groups allowed for better mediation sessions. Mediation was standardised[53] and a verbatim transcript was used in each session. In order to control for teaching styles, the entire process was controlled.

After testing was completed for both groups (over the three year period), scores were double-checked by different markers so as to ensure greater accuracy.

---

52 Not all black students are classified as disadvantaged here, as some black students had attended private schools.

53 Once again, pure dynamic assessment aimed at individuals should really be suited to each individual. However, standardisation of mediation is more practical and cost-effective. This illustrates the continuum on which dynamic assessment finds itself. Among all the different types of dynamic assessment available in this continuum, "…differences exist with respect to testing time, test format, degree of standardisation, task and training strategies [and] target processes of training" (Hamers & Resing in Hamers et al., 1993).

4.3.10.3 Main findings

The correlation findings of this study are discussed with reference to the three hypotheses mentioned above in section 4.3.10.1.

Hypothesis I

Both the traditional and dynamic assessment groups' results were correlated (using Pearson Product-Moment Correlations) with the criterion variable (November examinations) with students from all three faculties and then analysed according to the different faculties. Samples under six per faculty were not included for analyses as this was not deemed statistically significant. The only traditional measure which correlated significantly with the criterion score for the traditional group was the matriculation examination score.[54]

The correlation findings for the dynamic assessment group, although not significant, did yield some important statistics. There was an improvement in scores from the pre-test to posttest sessions (that is, from the traditional format of the test to the dynamically assessed format of the same test). "At face value, it appears that for the [f]ull [g]roup of students across both traditional and dynamic tests, Matric *[sic]* marks are the only measures which show a significant relationship with the [criterion variable] irrespective of faculty" (Zolezzi, 1995, p. 84). However, Zolezzi urges that a closer look at the change in results from pre-test format to posttest format be taken; and he states that the dynamic tests enhance the predictability of academic success for different groups.

The mean scores for all three faculty groups are higher for the dynamically tested group than for the traditionally tested group. Whether or not this can be attributed to dynamic mediation is questionable. Zolezzi offers no explanation for this. This is important because, as a group they (the dynamic group) score higher on the criterion variable, and scores also improve between pre- and posttest sessions. Perhaps if they had scored lower or the same as the traditional group on the criterion score, their improvement might not have been as marked. Secondly, the results from the study illustrate that the disadvantaged groups have higher difference scores (or learning potential scores) than the advantaged group. However, Zolezzi does not give the traditional scores for any of the students and thus one cannot tell how high or low the pre-test score was. Difference scores will be lower for the advantaged group purely because they may have scored higher in the pre-test session. The fact remains though, that the disadvantaged groups' improvement between pre- and posttest does enhance predictability of success, as the posttest scores correlate higher with the criterion score than do the pre-test scores. It is also important to note that a few advantaged students actually performed less well in the posttest session than in the pre-test session.[55] "This phenomenon might be worth of further research" (Zolezzi, 1995, p. 99). Means for the disadvantaged group are higher than the advantaged group for both the Deductive Reasoning Task and the Pattern Relations Test.

Zolezzi, as mentioned above, used two different groups for his sample determination. He states that, in light of the above-mentioned findings, "[t]he traditional measures are unable to predict or discern students who might benefit [from]and respond positively to mediation" (1995, p. 86). This may very well be true, but one cannot discount the fact that this statement is based on two different samples of students. The criterion scores for the traditional group are lower to start off with. If the two groups had equal means on the criterion score, then perhaps this statement would hold more sway. Furthermore, it may have been a more prudent idea to test two different groups from within the same sample by matching the students on a number of aspects. A step better would have been to use the same students over a period of two years and to assess them both traditionally and dynamically, using

54 This is to be expected - as it is reiterated in the literature, that the matriculation examination is the single best predictor of academic success at tertiary institutions, mainly for previously advantaged groups.

55 This is a factor that has arisen, for example, at the Pretoria Technicon, where above average performing students do less well in the posttest administration of a computer-based test, primarily because of the renewed way of looking at patterns, ways with which they were not familiar and so performed less well. These students should perhaps be assessed differently.

similar tests and not the exact same tests. This would yield results that are better able to be interpreted and commented upon. As it stands, there are too many extraneous variables that may account for the findings.

Per faculty breakdown, results indicate that the Science and Commerce students improved markedly from pre- to posttest, but that the Arts students for whom results decreased did not. Matriculation examination results remain the single best predictor of academic success for all three faculties. The dynamically assessed tests enhanced prediction for the Commerce and Science students. These results were interpreted for the entire group and have not been differentiated into advantaged and disadvantaged groups;[56] Hypothesis I had thus been confirmed.

Hypothesis II

Both the traditional and dynamic groups were divided into advantaged and disadvantaged groups and correlations were run to investigate any significant findings with regards to the way the two groups performed. Both the traditional tests and dynamic test scores were correlated with the criterion score (November examinations). According to the results, the Matriculation examination results are the only significant indicators for the advantaged groups from both the traditional and dynamic groups. The advantaged group did, however, improve in performance from the pre-test to posttest on the Deductive Reasoning Test. Although this was not significant, it did enhance prediction of the traditional Deductive Reasoning Test.

The correlations for the disadvantaged groups showed that Matriculation examination results were once again the only significant correlations for this group in both the traditional and dynamic assessment groups. Zolezzi does point out that the Raven's Matrices and Arithmetic Reasoning Tests correlate negatively with the criterion, and yet are still retained as tests "…widely used in aptitude testing programs and yet they demonstrate no relationship at all with academic success" (1995, p. 90). Both the Deductive Reasoning Test and Pattern Relations Test enhance prediction for this group and are significant. Zolezzi tabulates the means and standard deviations for both the advantaged and disadvantaged groups, for the dynamic test results. It is shown that the disadvantaged group has higher means and lower standard deviations compared with the advantaged groups. Their performance is also better in the Deductive Reasoning Test than in the Pattern Relations Test, Zolezzi interpreting this as an indication that mediation is not as meaningful for the non-verbal as for the verbal test. It is important to note that here we are not dealing with difference scores (which method itself is problematic as the pre-test scores may well have been higher for the advantaged group to start off with); but the scores used here are posttest mean averages only which, when compared with the disadvantaged groups' posttest average mean scores, are lower.

From the data, the disadvantaged students benefitted more so from the dynamic assessment than did the advantaged group. This is a positive result. Next, correlations were run for both the traditional and dynamic ways of testing for the advantaged group. According to the results, the dynamic measures were slightly better at enhancing prediction for this group, for both the Arts and Science faculties. The Commerce faculty was not shown as there were not enough registered students.

Correlations for the disadvantaged Science faculty group were run for the dynamic measures and the criterion variable. When compared to the advantaged Science faculty group, the correlations are higher for the disadvantaged group. The Deductive Reasoning Test (enriched version) is a better predictor of academic success than the Pattern Relations Test for the science students (both advantaged and disadvantaged). This suggests that deductive skills show a strong relationship with science subjects (Zolezzi, 1995).

56 Zolezzi states that from a qualitative point of view, the testing procedure was considered to be more relaxing and collaborative than traditional testing, as stated by both the assessors and testees in a qualitative questionnaire that was completed by the participants in the study.

Next, Zolezzi separated the groups into high and low learning potential groups, based on scores above or below the average means on the posttest scores of the Deductive Reasoning Test and Pattern Relations Test. As expected, the high learning potential group's correlations are higher than the low learning potential group's correlations. The results show that the dynamic tests are able to discern between high and low learning potential students. Note that the high learning potential group is composed of both advantaged and disadvantaged students. Three important results are thus evident at this point:

- Advantaged and disadvantaged groups have different predictors of academic university success
- The only significant predictor for both groups is the matriculation examination results, however, "…**this measure does not allow** for differences in educational background and because it is a **static** measure, is **unable to discern between high and low** learning potential students" [own emphasis] (Zolezzi, 1995, p.93).
- Other than the Deductive Reasoning Test (enriched version) and the Pattern Relations Test (enriched version), no other test is a significant predictor of academic success.

These results are very encouraging, making visible the fact that even though initial scores on tests might not be very high, subsequent training and enrichment lead to higher results in the posttest assessment. Hypothesis II is thus confirmed.

Hypothesis III

Zolezzi investigated this hypothesis by correlating learning potential measures with the traditional ability measures. The results were negative and hence not significant. These correlations were conducted for the entire group. In other words, there is no relationship between current level of performance (as operationalised in this study) and potential levels of performance. Hypothesis III is thus confirmed.

All three hypotheses were thus confirmed.

4.3.10.4 Conclusion and recommendations

Perhaps the most important conclusion that can be drawn from this study is that dynamic assessment enhances prediction for the disadvantaged group, and that the traditional tests were not able to predict for either of the groups (bar the matriculation examination results). This leads Zolezzi to state that aptitude testing in South Africa needs to be closely looked at. "The results add weight to the continuing disillusionment with traditional testing in the South African context" (1995, p. 97).

Of particular interest in the study's findings is the fact that matriculation examination results correlated significantly for both the advantaged and disadvantaged groups, which does not occur in some other studies.[57] However it must be noted that Zolezzi's sample size was very small.[58] Nevertheless, matriculation examination results are unable to discern high or low learning potential among students. From a qualitative point of view, Zolezzi states that the dynamic measures were more enjoyable to complete.

In essence, the research supports the contentions of both Vygotsky and Feuerstein, who state that given the right environmental support, initial test results can be improved upon. Zolezzi states that according to multiple-intelligence theory (Gardner, 1982, 1983 as cited by Zolezzi), students should be given the opportunity to demonstrate their potential in any learning contexts and with the Newtest Battery, various contexts are available, namely, verbal, visual, mathematical and logical contexts. Another important contribution made by the Newtest Battery is its time efficiency. It is relatively quick

57 See Shochet (1986) for instance.

58 Zolezzi acknowledges this fact and states further that the small sample size precluded regression analyses being conducted, especially for the advantaged group, as it "…would have provided a better and more streamlined outcome with regard to the **relative weighting of the different significant predictors** in the prediction process [own emphasis] (1995, p. 107).

to administer and standardisation makes scoring easy. This battery has also been normed on the population for whom it was intended, namely, university students.

The attempt of the present study was not to create disparity between different groups by manipulation of traditional test material, but was founded on the creation of a learning environment based on sound psychometric principles (Zolezzi, 1995). However, Zolezzi does not discount the status of traditional tests; he merely states that they can be administered in a dynamic manner, and in so doing enhance the predictability of students scores.

If the mediation session had been longer, more intense and individual, profiles on students' responses to mediation would have been a qualitative aid in the study (Zolezzi, 1995). Equalising tests do not have the same effect as equalising the test procedure, as Zolezzi states; and culture-fair and so-called culture-free tests are also not the answer in redressing psychometrics in South Africa. For Zolezzi, learning potential, modification and acculturation are similar concepts and he deduces that learning potential testing is the only way forward for South African psychometrics.

4.3.11 A small-scale investigation of the group administration of Feuerstein's learning potential assessment device - Master's study conducted by S.G. Andrews (1996)

4.3.11.1 Introduction

In a study conducted by Andrews in 1996 in a Kwa-Zulu Natal school, the main aims of the investigation were to:

- "determine whether or not, within a given group of subjects … the mediation offered during an application of Feuerstein's LPAD in a group-administration format would result in modified cognition, demonstrated by improved performance in post-mediation testing and
- to determine whether in the same group of subjects, the group administration format of the LPAD would detect differences in the degree of cognitive modifiability of individuals".

(Andrews, 1996, p.i)

The two hypotheses state that the mediation given to the experimental group would result in modified cognition, which would be reflected in an improved performance on the posttest and secondly, that the group administration of the three instruments from the LPAD would be able to detect differences in the degree of cognitive modifiability of individuals (Andrews, 1996).

Andrews's study set about investigating the use and results of dynamic assessment, that is, he investigated the role played by enrichment or mediated learning. The sample comprised 21 black female students attending a private high school in Kwa Zulu Natal. Due to their proficiency in English (more so than black students in public schools), these students were chosen so as to administer the mediation in a language that was familiar to the students (which in many studies in South Africa has posed a problem) more easily. The division of students into one of the two groups (experimental, N = 10 and control, N = 11) allowed for statistical inferences to be made[59] (Andrews, 1996). The sample was so small because of financial and resource constraints.

The results of this limited study, notwithstanding their somewhat tentative nature, are encouraging in that the level at which assessment of modifiability can be used to determine where an individual makes cognitive errors can be identified in a group administration of mediation. This is a significant finding, since the LPAD has proven more effective when administered on an individual basis.[60]

Andrews's theoretical point of departure is a skilful weave of Sternberg's two metaphors of intelligence (which are themselves a summary of theories of intelligence since the start of the twentieth century),

59 This small sample does lead one to consider if statistical inferences are valid in this study.

60 "The experimental work completed indicates that individual administration of the LPAD clinical battery has substantial advantages over group testing" (Feuerstein et al., 1979, p.319).

namely, the geographic and sociological metaphor of intelligence (Sternberg, 1990 in Andrews, 1996) in combination with the original ideas of Vygotsky (1978) and Feuerstein (1979), as well as Tharp and Gallimore's 1988 model of teaching as mediation (in Andrews, 1996) make for a study grounded in solid dynamic assessment theory. As the ideas of Vygotsky and Feuerstein have already received attention, only a brief description of Andrews's interpretation of Tharp and Gallimore's theory is given:[61]

> Tharp and Gallimore (1988, p.44) integrated neo-Vygotskian theory of development with various considerations central to behviourist and cognitive studies of learning. They provided a theory of teaching which is useful in understanding and operationalising some of the concepts in a model like Feuerstein's. They derived six means of providing assistance within the zone of proximal development: modelling, contingency management, feedback, instructing, questioning, and cognitive structuring (Andrews, 1996, p. 10).

The reason why Andrews makes use of Tharp and Gallimore is what he perceives as a lack of operationalisation of theoretical constructs used by both Vygotsky and Feuerstein, which may be understandable in theory but poses problems to the practitioner who does not know precisely "..what [has] to be done during the mediation process" (*ibid*.).

Andrews makes use of three LPAD instruments, Numerical Progressions, Organiser and the Complex Figure Drawing.[62] The reason why Andrews used these three particular instruments from the LPAD was twofold:

- They are suitable for group administration
- They make for ready analysis across a range of modalities as described by Feuerstein (1979) thus neither favouring nor disadvantaging any student.

A string of statistical analyses was run which enabled the researcher to compare the pre- and posttest scores of both the groups with each other as well as to define the level of difficulty of each item used in the tests. For these comparisons (only run for the Numerical Progressions and Organiser), t-tests were conducted. However, owing to the nature of the Complex Figure Drawing Test, a slightly different approach was necessitated. This test required the students to copy a drawing, and then to re-draw the picture from memory. Hence two sets of comparisons were conducted, a pre- and posttest copy measure and a pre- and posttest re-draw measure for both the groups.

4.3.11.2 Research design

The research design comprises a pre-test/posttest set-up, using matched groups divided into experimental and control groups. Two sessions were held, the first merely to divide the students into one of the groups. The decision to place students into either the control or experimental group was a process which entailed placing the highest performing student (aggregated results obtained in the pre-test phase of the research on all three instruments) into the experimental group, the second highest into the control group and so forth. The results obtained from the administration of three tests were used as a baseline measurement.

Two weeks after the initial testing and placement of students into their respective groups, the experimental group received mediation on the tasks they had already performed and which they were to perform again in the posttest. The control group received no mediation but participated in the posttest. Statistical tests were conducted to ascertain whether or not there were any differences between the two groups' posttest results and whether there were any differences between the pre- and posttest results for both groups.

---

61 See Adams (1994) section 4.4 who also mentions Tharp and Gallimore in a similar context.

62 See also Lipson (1992) who made use of the Numerical Progressions; Gaydon (1988) Henley (1989) and Hoffenberg (1988) who made use of the Organiser; as well as Gewer (1998) and Lipson (1992) who made use of the Complex Figure Drawing Test.

4.3.11.3 Main findings

Numerical Progressions

According to the t-test conducted for both groups, pre- and posttest measures for the Numerical Progressions test, it was found that there was a significant difference between the two groups, that the experimental group's results were significant at the 0.01 level, but the control group's results were not significant. The variation in individual scores also points to the differences between students on various sections of the test. However, as mentioned previously, the group administration did not allow for pinpointing the exact mental operation used or not used by each student. Nevertheless hypothesis 1 is supported based on these results.

Regarding hypothesis 2, (the fact that group administration of this test will be able to "detect differences in the degree of cognitive modifiability of individuals" [Andrews, 1996, p.i]) the degree of variability of results points to the differences of cognitive modifiability of the students. The researcher divided the subtests of the Numerical Progressions into levels of difficulty so as allow comparisons between them for each student. There were in total eleven levels of difficulty identified. The one drawback was the unequal number of tests in each level in the pre- and posttest measures, which seems unbalanced. Because of the greater number of subtests involved in the posttest, the results were percentaged so as to facilitate comparison. The main finding, which supports hypothesis 2, is that "…a comparison of the results achieved by individual subjects in the pre- and posttests of Numerical Progressions, …indicates that many [students] are able to perform at two or three levels higher in the posttest, following mediation, than in the pre-test prior to mediation" (Andrews, 1996, p.41). This proved to be more valid for the experimental group than the control group. However, Andrews does not perform any significance testing on the differences of scores in each level for both groups, hence his statement that "…the **qualitative** results lends some **support** to the second hypothesis in that it is possible to differentiate between the levels of performance of individuals within the experimental group and postulate that these differences are a reflection of the variations in degrees of cognitive modifiability of the individuals concerned" [own emphasis] (p. 43). It may have been more advantageous for Andrews to have performed a more robust quantitative test to investigate his second hypothesis which may or may not have been supported.

Organiser

In the t-test results for the Organiser pre- and posttest results for both groups, there is a significant change for the experimental group, at the 0.05 level, but not so for the control group. Once again, hypothesis 1 is supported for this test. Thus the variable introduced, namely, the mediation administered is the main contributor towards the differences. As with the Numerical Progressions, the Organiser is also divided into different levels of difficulty. The experimental group achieved higher scores in the posttest measure as opposed to the control group, as well as scoring successively higher scores per level on an individual basis than the control group. This is attributable to the mediation. However, Andrews highlights the unfamiliar terminology that is used in this sub-test stating that "…words such as "Maple", "Willow" and "Sofa" are, to South African second language speakers, unfamiliar, and hence difficult to understand in context" (1996, p. 63).[63]

As explained above, the Complex Figure Drawing Test design is slightly more complex than either of the above-mentioned tests owing to the copy and re-call phases included in the design. Unfortunately, the results are not positive, evidencing no improvement for either group on the posttest measure, for both the copy and re-call phase. In fact, the mean score was lower. Andrews asserts that the t-test statistic yields a significant decrease for the control group in this test but does not comment on a similar t-test statistic for the experimental group, saying only that "…the mean decrease from the pre-test to the posttest in the CFDT is **far below the mean increases** of the other two instruments" [own emphasis] (Andrews, 1996, p.58). It is to be expected that the mean results for the re-call phase would be lower than the copy phase since students were not forewarned about the task. Two

63 This is one example where the LPAD, for instance, may need to undergo revision as a test used locally.

explanations are offered as to why the resultant scores were as low as they were, namely, that (a) the posttest example of the CFDT is considerably more difficult than the pre-test example, when compared to the difficulty levels of pre- and posttest examples of the Numerical Progressions and Organiser, and that (b) the CFDT does not mediate on memory but rather on analytical strategies (Andrews, 1996). This may or may not be the reason why the students performed so poorly when compared to the results of both the other tests.

Andrews analyses two experimental group students' drawings of the CFDT and highlights the degree to which one is able to analyse two students' respective degrees of cognitive modifiability which demonstrates support for his second hypothesis. One student's drawing reflects a better attempt during the copy phase as opposed to the re-call phase (pre-test measure), while the other student remains the same in the posttest score between copy and re-call with only one of the students' drawings reflecting an increase from pre- to posttest (on copy and re-call), with the other experimental student actually decreasing (on copy and re-call). This may indeed point to a possible cognitive error for one of the students, which could have been corrected had the administration been on a one-to-one basis.

4.3.11.4 Conclusion and recommendations

The culminating scores of all three tests for both groups' pre- and posttest sessions were conducted using a t-test, which yielded a significant difference at the 0.01. Andrews states that the overall significant increase in both the Numerical Progressions and the Organiser was reduced by the poor performances on the CFDT. This result, however, supports the first hypothesis in totality.

Regarding support for the second hypothesis, Andrews ran another test "…on the combined scores of the subjects [and to] examine the differences between the pre- and posttest scores of the subjects in the experimental group to determine whether those subjects proving modifiable in one area were equally modifiable in the others" (Andrews, 1996, p.58). Should the study ever be repeated, Andrews believes that the CFDT should not be used and instead replaced with "Rey's[64] figure in the posttest and one of the others in the pre-test" (p.64). Furthermore, Andrews mentions a number of issues which may have impacted negatively on his study, but the most influential one remains the lack of explicit explanations of a few concepts used in the Feuersteinian manual. The fact that students did not receive feedback or treatment in identifying possible cognitive errors in thinking may also have been a motivational issue. Andrews concludes with an expectation that dynamic assessment in South Africa at this stage (1996) would not be a financially viable option to the way education is seen to veer.

Despite the financial implications, overall, the results of his study are positive and support the notion of dynamic assessment in South Africa, provided that such assistance is conducted in the home language of the respondent by someone fluent in that language.[65]

4.3.12 Assessment of the predictive validity of the learning ability battery - Master's study conducted by J. Tayler (1996)

4.3.12.1 Introduction

The Learning Ability Battery (LAB), a commercial product used in the assessment of potential employees in firms is, according to Tayler (1996), used widely. It is purported to be of significance in the area of learning ability specifically within the realm of dynamic assessment owing to the nature of the test. The aim of Tayler's study was to scrutinise the predictive validity of this test battery in the South African context. "A major criticism of the instrument is its lack of statistical support, in the form of reliability and validity data" (Tayler, 1996, p.65). The study looks at the usefulness of the LAB in

64 Andre Rey was Feuerstein's mentor, whose own figure prompted the design of Feuerstein's figure which was used in his LPAD (Lidz, 1987).

65 Findings of other studies such as Boeyens (1989b), Henley (1989) and Hoffenberg (1988) echo the same sentiments.

the workplace environment and seeks to validate "…the need for selection instruments which can cut through a lack of education and an impoverished upbringing and highlight untapped potential" (Tayler, 1996, p. iii). A summary of the results of this study shows that reliability and validity scores are valid and that the LAB indeed is suitable as an accurate means of placing prospective employees into various groupings. The results are examined below in more detail.

One of the main aims of the LAB is to test illiterate adults for the workplace. As school qualifications obtained in the era of apartheid still in many instances dominate prospective employee chances of being placed in positions in companies, the need to look elsewhere for a suitable test to administer to previously disadvantaged people needs to be looked at. "It is claimed that the LAB is able to measure the amount of learning potential in the most educationally disadvantaged subject, that is, the illiterate adult" (Tayler, p. 6). The LAB is thus viewed as an alternative to traditional testing, making use of Feuerstein's philosophy as foundation for the development of the LAB:

> The implication of Feuerstein's argument is that a measure of learning ability might provide a fairer and better predictor of workplace success than the currently used selection tests. The advantage of such a measure is that it would not carry the stigma of being either an adapted test, or one which had previously been used as a "screen-out" device (Tayler, 1996, p.50).

Tayler's motivation for the investigation into the predictive validity of the LAB is prompted by the need to move away from the previously held notions of testing, notably, the genetic-environmentalist approaches and she instead advocates the dynamic testing paradigm which she discusses at length. A point of contention in this field in South Africa, as Tayler notes, is the lack of locally produced tests to assess learning potential, and many tests are merely adapted tests from overseas. Tayler cites three locally developed learning potential instruments of which she is aware (ca 1996), namely the Ability, Processing of Information and Learning Battery (APIL),[66] the Potential Learning Index Battery (PIB) and the LAB (which, as already stated, is the focus of her study).

One of the major draw cards for making use of this LAB test battery "…is that is has been developed for the South African market by South African psychometricians, who have taken into account the inherent problems of testing in this country" (Tayler, 1996, p.60). The fact that the non-verbal section has only been translated into Afrikaans, necessitates the use of a translator which brings with it the disadvantages of unstandardised instruction. However, as Tayler points out, the Feuersteinian notion of practice examples goes some way to "easing" the translation process. The LAB battery makes obvious use of static test components such as the visual-motor and verbal assessment tests, and the nature of the instructions and format is quite structured. The test-teach-test approach is also not strictly followed as is usually the case, but it does make use of the pre-test, teaching and familiarisation efforts with a posttest after a long interval period (an objective criterion scores) associated with dynamic testing and as such can be considered dynamic.

4.3.12.2 Research design

Tayler emphasises the need for the learning potential testing arena to take cognizance of validity and reliability issues, seeing as very little has been done to assess these issues in South Africa.[67] Companies willing to make use of dynamic selection procedures will need to be convinced of the statistical significance of dynamic testing, so that not only can a new paradigm shift take place within industry but such a shift will also be a viable and scientific option.

The LAB was first conceived as an idea when a large parastatal company in South Africa wished to measure and group poorly educated and illiterate people as well as group those who had received their school leaving certificate. The wish to "…implement a widespread educational programme of Mother Tongue Literacy and Basic English and Afrikaans for its labourforce, provided the catalyst for

66 See Van Aswegen (1997) section 4.3.13 as well as section 4.4 for more information on the APIL.

67 Taylor is partially correct, since these issues had already received some attention in the field.

this intervention" (Tayler, 1996, p. 52). In Tayler's study, only two of the five group-administered tests were investigated, namely the verbal and non-verbal sub-tests.

Non-verbal sub-tests

The non-verbal sub-tests consist of seven sub-tests, namely, hand-eye co-ordination, visual memory short term, visual memory long term, visual perception, visual insight, quantitative perception and seeing the whole picture. Each sub-test was introduced to the respondents who were allowed to practice the exercises under the supervision of the test administrator. Tayler mentions that the administration of the sub-tests is in keeping with the four main principles or elements of Feuerstein's approach, namely, practice examples, two-way communication between teacher (facilitator) and observer, process orientated administration instead of product orientation administration and lastly the individual being tested is given an indication of his/her learning potential.

Verbal sub-tests

The verbal sub-tests consist of 50 sentences available in 7 ethnic languages which range in difficulty, comparable to the levels of Std One - Std Eight ability. In essence each sentence is missing a correct word, the correct word appearing amongst others in brackets. The choice made by the individual will be assessed as a measure of the type of language programme most suitable for the candidate in which to enter. Due to various factors in the Tayler study, three sub-tests were not included, namely the eye test, the hearing test, and the mathematical test which were substituted by results from the quantitative perception sub-test.

Face validity is commented on by Tayler as being high with test takers acknowledging the test's "…culture-free and unbiased nature" (p. 62). No manual accompanies the LAB and it is scored by computer, which reduces the rate of error when administered or scored by hand. Tayler's sample consisted of 132 "…urbanised Black men and women between the ages of 20 and 65" (pp. 66-67) who were volunteers from five organisations and the study lasted a maximum of seven months and the volunteers ranged from totally illiterate/innumerate to literate/numerate individuals.

The first step in the process was to train five prospective administrators so as to ensure accurate testing results after initial testing. Initial testing was conducted so as to place employees into the correct categories for educational training. These selected administrators were tested on all aspects of the LAB work covered in the training period. In order to ensure that the LAB scores were indeed predictive of the highest level of learning which any one individual was able to attain, a further criterion was established, namely, that after the initial training was received and the selected courses attended a second testing of the sub-test be readministrered. This would suffice as an objective criterion. Tayler cites Huysamen (1987) and Anastasi (1988) as supporting the notion of including a subjective criterion as well which would facilitate the assessment of each individual scored by the LAB. The subjective scoring would take place in the form of rating scales, rated by the test administrators.

4.3.12.3 Main findings

At the end of the procedure, Tayler obtained the computerised LAB test scores (administered by the trained administrators who belonged to one of the five companies mentioned above), the objective criterion scores (the readministration of a sub-test) as well as the subjective test score (rating on a 6 point scale made by the administrator). At this stage Tayler made use of the SAS system and not the LAB software (as SAS contained all the necessary statistical procedures necessary for her analyses) in order to analyse the data at her disposal. In essence then, the method of research proceeded by choosing five organisations, Tayler herself employed by the company who developed the LAB. Five elected representatives were chosen by each of the five companies to receive training as administrators. After a two-day seminar and training course the administrators were tested on their knowledge of the LAB. Each administrator was then sent back to test a number of employees (twelve or fewer members in each group). After initial testing, employees were sent to various courses based on the results of the LAB. After a five to seven month interval during which employees received training, each administrator re-tested the employees, and this score served as the objective criterion score, a measure reflecting whether the LAB had indeed classified employees into the correct

courses. A subjective score was also taken, namely, a rating on a six-point scale of the level of employee progress.

The dynamic nature of the LAB test battery is evident in the test-train-test methodology followed (as mentioned above, this approach is not followed in the conventional manner) the training period being the courses employees attended. Of course in this research, the total length of time taken for the study is considerable. Many studies train respondents within a matter of hours or days. As no validity or reliability data had as yet been gathered, Tayler decided to make use of an exploratory factor analysis to determine whether the LAB "…test items loaded accurately onto the each of the given sub-tests" (p.74). Internal consistency was investigated as was the predictive validity. Pearson correlations were computed to ascertain the degree of accuracy between the criterion variables (the objective and subjective criteria), Cronbach Alpha's were computed to ascertain the degree of internal consistency (and thus measure reliability) and factor analyses were conducted to test for predictive validity. Tayler accepted a Cronbach Alpha equal to or above 0.70 as acceptable (a measure obtained by Tayler from a consolidation of the findings of various authors in the statistics literature).

In order to investigate the reliability of the LAB, the seven sub-tests, the total scores of the sub-tests, both the objective and subjective criterion scores and the scores based on the pre-test (word recognition scores) and lastly the total sub-test scores added to the total readministration scores were used in the analyses. However Cronbach alphas were not computed for the composite scores of the total readministration score nor for the composite sub-test scores, as they were "not legitimate scales" (Tayler, p.85). All correlations were significant at the $p<.0001$ level, except four. All 50 LAB test items were subjected to a factor analysis, the aim of which was to derive at a factor matrix, initially from a correlation matrix. Initially, 29 factors were obtained but were subsequently reduced to 8 due to cut-off points using specified eigenvalues as detailed by Tayler (p. 85). "Promax rotation using a reference structure of semi-partial correlation was found to yield the best fit of variables to factors" (p.86). After analysis of the factor analysis results (in other words the 50 items making up the seven sub-tests of the LAB), the eight factors cumulatively explained 50% of the total variance. In total the various factors loaded relatively well on the items, with only two negative loadings in sum. Scrutiny of the factor loadings prompted Tayler look at the test items again and, in some cases, certain items were scrapped or at least rewritten. Pearson's correlations were run for the total LAB test battery scores and the readministration scores (the objective criterion scores) as well as the pre-test scores (word recognition scores). The highest correlations were found between the objective criterion score and the composite score for the LAB, $r = 0.78$ and for the objective criterion score and the composite word recognition scale, $r = 0.78$. The composite word recognition scores are the scores obtained after the courses had been attended.

Tayler points out that "…a criterion used in the measurement of predictive validity, [should] be internally consistent" (p.90) and according to the Cronbach Alpha's of both the objective and subjective criterion scores, 0.83 and 0.81 respectively, it seems as if the criteria adhere to the tenets of predictive validity. In other words, a good Cronbach Alpha for the criterion scores bodes well for the Pearson correlation between the criterion scores and the LAB scores. The Pearson correlations are thus further supported. "While high alphas do not in themselves signify the credibility of the instrument, **viewed together with the results of a factor analysis**, internal reliability may be confirmed" [own emphasis] (p. 99). Student t-tests were also carried out to investigate the nature of the change from the pre-test to posttest situation. The sample was divided into literate and illiterate people. The calculated t values were significant for both groups, thus indicating better performance after the test as opposed to prior training. There was no posttest for the numeracy group and Pearson correlations had to be used instead as an indication of predictive validity. There was a positive relationship between the objective criterion score and the numeracy scores.

4.3.12.4 Conclusion and recommendations

Tayler's conclusions about the results indicate that the LAB is a valid and reliable test battery to be used in the multicultural dynamic context for which it was designed. The sub-scales show "high internal reliabilities" as well as a "lack of covariance amongst them" (p.95). Validity of the instrument is supported by the correlations between the total LAB sub-test scores and the criterion scores. Of interest is the fact that although the objective and subjective criterion scores are moderate ($r = 0.52$)

the subjective scores do not correlate well with the total LAB scores, thus indicating the administrators difficulty in accurately assessing individuals' learning behaviour. Finally, the students' t-test shows significant differences between the pre- and posttest phases. Of note is that Tayler mentions that although the LAB is widely used, her study validates its use by making available reliability and validity results. The question remains as to why these tests were not carried out before usage of the test battery. However, according to the results, this "semi-dynamic" test appears to perform its intended function. Tayler points out that although the LAB is valid and reliable as a test instrument, it should be complemented by the use of personality questionnaires and interviews.

Notable limitations to the study include the nature of the sample which was composed of volunteers which, by their nature, renders them a unique sample base. Small numbers of women speakers of minority languages made generalisations to these population groups difficult. But perhaps the most important differences between the LAB and the principles and practices of dynamic assessment include the fact that the pre-tests are not scored (which they are with, for example, Feuerstein's LPAD, in order to get a better idea of a candidate's learning potential) as well as the fact that the LAB has to be completed within a certain time frame. This is not true of the LPAD, for instance, which allows the respondent enough time to complete the tests. Or if there is a time limit, it is nevertheless structured around the respondents.

Tayler acknowledges the fact that administrators, as with any administration of any psychometric test, need to be randomly monitored to ensure adequate administration of the LAB battery. This would entail test developers and trained members to oversee certain testing periods. This, of course, involves yet another cost factor. Although the LAB utilises some principles of dynamic assessment, Tayler comments "[i]f the suggestion to produce both a pre- and postfamiliarisation score is utilised, [the] LAB will become a more dynamic measure" (p. 117).

4.3.13 The standardisation of a learning potential battery for the selection of poorly qualified employees - Master's study conducted by M. Van Aswegen (1997)

4.3.13.1 Introduction

The Van Aswegen study is an example of research conducted within an industrial context, a context in which the need to assess people in terms of potential far outweighs their current ability to get the job done. Poorly qualified employees, are often not given the opportunity to develop their intellectual skills owing to the lack of previous education and so cannot enhance their working experience. In this regard, dynamic assessment plays a crucial role in identifying those candidates who will benefit from in-house training opportunities, as it is clearly evident that past school results will not only put them at an immediate disadvantage but will also disqualify them from developing their potential.

Van Aswegen tested the TRAM1[68] battery as devised by Taylor (1994). The goal of the study was to "…evaluate the validity and effectiveness of the Transfer, Automatisation, Memory and Understanding Learning Potential Battery (TRAM-1) to predict future performance of poorly qualified employees on an accelerated development programme" (Van Aswegen, 1997, p. x). Van Aswegen discusses at length the need for alternative assessment measures within industry, and his study looks specifically at the mining industry in South Africa. He argues that unfair labour practices, such as job reservation in the past, restricted most managerial positions to whites. However, this is no longer the case under the new dispensation. In order to ensure that those candidates will function successfully in managerial positions, it is essential that the needed skills are acquired by the right people. Hence the need to choose an instrument which can reliably indicate those employees who are more likely to succeed in managerial positions than those who would tend to perform poorly.

The TRAM-1 battery was developed and normed by Terry Taylor and the battery was investigated in this study in order to find out whether it could be utilised within the mining industry as an effective predictor of performance on an accelerated course, offered by the mining companies involved.

68 See Taylor (1999) section 4.4 for more information on the TRAM-1.

According to Van Aswegen the TRAM-1 is "… a non-verbal paper-and-pencil … learning potential battery" (1997, p.56). It is used to assess individuals who are illiterate and semi-illiterate, or those who have received no real education through to nine years of schooling. The company, whose employees were used in the sample, numbered approximately 60 000, of which 805 are said to have received education below Std Seven (Van Aswegen, 1997). The decrease in unskilled labour and the subsequent increase in skilled labour, has made it imperative that alternative measures are found in order to identify those employees who are most likely to succeed in a skilled position. Van Aswegen states that there are "…currently no instruments available to scientifically and objectively identify employees from historically disadvantaged cultural groups with potential for further development"[69] (1997, p.4). Van Aswegen devotes a chapter to the discussion on employment equity and the use of psychological test measures in South Africa, and points out that more and more trade unions are becoming increasingly aware of the many dubious measures available in terms of norm-referenced tests that were never suitable for cultural groups other than white candidates. Legislative developments within the domain of assessment as well as the influence of cultural differences are discussed in order to offer support for the need to utilise a dynamic assessment approach as opposed to relying almost exclusively on static measures. Once again as with many studies investigated thus far, Van Aswegen contends that the notion of disadvantage in no way implies an homogeneous grouping of people. Not only are language and culture significant factors in the differences but so too are aspects such as geographical location (rural vs. urban), which feature as a so-called third variable in the study.

Examples of unsuitable tests which have been used in previous studies investigating learning potential are (according to Van Aswegen) the Intermediate Mental Alertness Test, which has formed part of test batteries in the studies of Boeyens (1989b) and De Villiers (1999). Van Aswegen also cites the Raven's Progressive Matrices test as being unsuitable for black, coloured and Indian groups. He emphasises the need to change the current labour force into a multi-tasking team if South Africa is to become a global competitor. One way in which to establish this new way of thinking is to place people in positions who are able to fulfil the necessary functions of that position, but who are at the same time also able to cope with the demands of that position. In order to identify such individuals, it is then necessary to devise measures other than the traditional ones to assess them. There is thus a need "…to make use of a **properly validated, scientifically based** assessment instrument that will ensure that the 'high performers' can be identified for accelerated development into supervisory and managerial positions from inside the organisation" (1997, p.17). Van Aswegen discusses the merits of validity studies that need to be conducted in any psychological testing instrument and gives a brief synopsis of the various types of validity.

Intelligence and the TRAM-1 Battery

As a prelude to the discussion of the various intellectual processes measured by the TRAM-1 battery, Van Aswegen details the most recent trends or theories in the broad field of intelligence. One striking feature to which the discussion will return, is that the TRAM-1 battery measures speed as opposed to power, a factor not highlighted in many other learning potential assessments. However, the reason for this becomes clear when Van Aswegen discusses the intellectual theory behind the development of this battery, a theory on which it was built; and he enters the debate about the definition of intelligence, a notion that has still not received unanimous agreement within the scientific community (see section 2.2).

However the assumption that Van Aswegen makes is that in any assessment "… the incorporation of a teaching or learning element into the assessment procedure [is required]" (p. 26). Taylor, the developer of the test, based the TRAM-1 battery on the following three approaches to intellectual assessment, namely,

- The structural psychometric approach (conventional approach)
- The information processing approach
- The dynamic or learning potential approach.

69 However the research of Tayler (1996) discussed above comes to the fore in this context.

This makes for an interesting blend, as the first and third approaches are usually viewed as antithetical, in the sense that they usually do not correlate at all well in terms of the construct they purport to measure. However, it must be recalled that the approaches are often, and should be, used in conjunction with each other in assessment procedures.

Van Aswegen proceeds to give a detailed account of the various theories that fall within the above-mentioned three approaches. The reason why detailed analyses are given is because the TRAM-1 makes use of various intellectual theories. For example, it makes use of the concept of Automatisation as developed by Sternberg (1994, in Van Aswegen, 1997). The notion of speed is attributed in part to the theory of Ackerman (1988, in Van Aswegen, 1997), who critically visualises intelligence in a cylindrical cognitive model, with speed forming the vertical dimension of the cylinder. (For more information, see Van Aswegen, 1997.) Hence the TRAM-1 measure borrows concepts from the information processing approach and structural approach (the idea of fluid intelligence being "… a valid predictor of performance in a cross-cultural environment" [Van Aswegen, 1997, p.31]). Lastly, but more importantly, the learning potential approach is of great importance in the TRAM-1 model. The idea of training and testing is clearly indicated in the use of this battery. Van Aswegen duly discusses the various theories and theorists within the domain of learning potential, discussing contributors such as Vygotsky, Budoff, Feuerstein and Wiedl (see Chapter 2). A concise summary of all three approaches is then taken as a basis for discussion on "…a theory of accommodating the conventional, information processing and learning potential assessment approaches" (p. 47).

The emphasis in the TRAM-1 battery according to Van Aswegen is on the role of speed. Although not the only critical aspect involved, it plays as important a role as do the processes involved in intelligent thinking. Van Aswegen briefly reviews the theory behind the TRAM battery as developed by Taylor. Of note is the unique blending of cognitive theories used as the foundation for the TRAM-1 battery building upon two models of cognitive functioning, that of "…Ackerman's cylindrical elaboration of Snow, Kyllonen and Marshalek's (1984) circular cognitive model as the basis for his theory" (Van Aswegen, 1997, p. 48). Featured strongly in the TRAM-1 is one of the two main dimensions of intelligence, which according to Sternberg (1994, in Van Aswegen, 1997) is Automatisation, which refers to the degree of efficiency with which a person solves problems, after an initial period of learning. Fluid and crystallised intelligence also plays a role in the theory behind the test, which emanates from Hebb's (1949) conceptualisation of intelligence A and intelligence B (as cited in Van Aswegen, 1997), the former reflecting fluid intelligence and the latter crystallised intelligence. Transfer is also a measure focused on in the TRAM-1 battery, the idea being postulated by Ferguson (1954, 1956 in Van Aswegen, 1997). Transfer is the extent to which individuals are able to systematically gain new information through a process of transfer within any given culture, and hence, the culture more or less transfers what it considers to be of importance. Different abilities emerge in difference cultures.

Based on the above amalgamation of theories from all three schools of cognitive thought (structural, information processing and learning potential) the TRAM-1 battery uses five dimensions to measure learning potential, namely,

- Transfer
- Automatisation
- Information processing speed
- Accuracy of learning and
- Memory and understanding.

"This theory forms the basis of a suggested battery incorporating measurement approaches from all three traditions. The envisaged battery used in this study is intended to be appropriate for use on people from all cultures in South Africa" (Van Aswegen, 1997, p. 51).

4.3.13.2 Research design

As the aim of the research study is to determine the predictive validity of the TRAM-1, it follows that one of the more important types of validity is criterion-related validity. Van Aswegen states that two methods can be applied to determine the predictive validity of any test, namely predictive or concurrent validity. Having chosen the former dictated by the time period that passes between the

administration of the predictive tests and the criterion tests at a later stage, a concurrent test would not allow for the lapse of time. The predictor information was collected at the beginning of the year and the criterion data at the end. The design constituted a "…non-experimental, predictive validity design… to evaluate the criterion-related validity of the TRAM-1… for poorly qualified black mineworkers" (Van Aswegen, 1997, p.53).

The sample consists of 101 black male employees working for a particular mining company, and the sample is spread out over a few geographical locations. Van Aswegen cites support for his use of 101 employees as a statistically valid sample number. Those chosen for the study were employees already enrolled in the development programme, a programme instituted in order to help employees obtain their standard eight certificate or N1 qualification. Although there was little written proof of the standards passed in the sample, Van Aswegen states that the average standard passed was Std Five, ranging from Std Two till Standard Nine. It was found that most of the employees could barely read and write. The mean age of the sample was 35.4. The profile of the sample was in a manner of speaking pre-determined, in that those chosen for the course were also those who had proven themselves worthy of obtaining first level supervisory positions within the company. The reason why this is pointed out is because this group is more likely to be motivated than those employees who were not interested in developing their skills. Van Aswegen recognises this fact when he states that "…use was made of a non-probability sampling technique, namely the purposive or judgmental sampling technique" (1997, p. 53). Included in this "purposive" identification of employees were various conventional psychometric tests. The TRAM-1 would (it is hypothesised) be able to assist in management's decision as to who should be placed in the course.

The independent variable is the TRAM-1 test battery and the dependent variables (criterion variables) include four measures which will be looked at below.

Independent variable

The duration of the test is 95 minutes and consists of a Phase A and B, each with their own booklets as well as a memory and understanding test booklet. The dynamic assessment procedure is clearly identified within the test administration, in that the instructors teach a lesson before the administration of the second part of phase A. "Learning occurs throughout the testing process" (p. 56). Aid is administered in a structured manner, hence individual coaching does not take place, but rather a standardised approach to helping those who need assistance. The tasks in the test are novel, as nobody would have seen the items beforehand thus allowing everyone the same opportunity to solve the problems, "(t)he actual task performed by the testee involves translating symbols into other symbols using a special dictionary….[which] are of a pictorial nature" (Van Aswegen, 1997, p.57). Rules govern the type of symbols that are placed together and are not arbitrary groupings.

As discussed above, the TRAM-1 measures the following dimensions, namely,

- Speed of learning - the speed at which learners are able to assimilate and use the information learned is an indication of their basic intellectual capacity
- The accuracy of learning - according to the TRAM-1 explanation of accuracy, making mistakes is not an indication of any intellectual problems but rather an impairment in the "monitoring" process involved
- Transfer of learning - the ability to make use of information learned in one situation and to apply it to a new situation can be construed as transfer of learning. Van Aswegen points out that it is this aspect of intellectual functioning that best distinguishes those who are mentally retarded and those with learning disabilities (Campione & Brown, 1985 in Van Aswegen, 1997)
- Automatisation of learning - this refers to the time taken to learn information and to apply it efficiently; this "learning curve" differs for each individual
- Memory and understanding - in order to perform well on this dimension it is important to note that mere rote learning will not be helpful, since the underlying rules have to be learned and implemented, which implies understanding at a deeper level
- The composite test score - is a cumulated score based on all dimensions, but memory and understanding are given a double weighting. The composite score gives an indication of the initial level of performance, the level of performance due to training as well as the level of performance due to the effects of practice. It provides information on the present level of functioning of the

student as well as the potential level of functioning. This is much akin to the pre- and posttest performance scores, or difference scores used in other dynamic procedures.

Dependent variables

Four measures make up the dependent or criterion variables, namely,

- Technical college results - in order to pass four subjects (Mathematics, English, communication and basic mining) employees have to obtain at least 40% in each subject. The aggregate of the four examination marks is taken as the technical college mark
- In-house results - these are examinations that employees write in the company (Mathematics, English, basic mining and mining ventilation). Once again the aggregate is taken as the in-house mark
- Combined scores - due to the similarity between the in-house subjects and the college subjects, Van Aswegen decided to combine the two and calculate an average of the two, hence an average in-house/technical score
- Composite overall score - is the composite equal weighting scores given to both the in-house and college scores.

The reason cited as to why four measures were used was the more accurate picture obtained when more than one measure is used; "[b]y the use of a number of measures, each contributing a different facet of information, we can limit the effect of irrelevancies and develop a more rounded and truer picture of programme outcomes" (Weiss, 1972, p. 36 in Van Aswegen, 1997, p. 59).

4.3.13.3 Main findings

Van Aswegen made use of the STATISTICA package and univariate descriptive statistics as well as multiple regression analyses. Descriptive statistics such as mean and standard deviation as well as age and education level are described. The multiple regression analyses concentrate on the following;

- Linearity - graphically representing the relationship between the independent and criterion variables
- Reliability - necessary for determining any bias in the tests used. According to Van Aswegen, the unique nature of the TRAM-1 does not allow for the usual methods of reliability tests, hence the need to measure reliability using the following approaches:
  - Speed of learning - correlating the amount of work done in phase A and that work done in phase B
  - Accuracy of learning - correlating accuracy of phase A with that of phase B
  - Automatisation - based on improvement scores of phase A and B and thus correlated with each other (this was based on a different test statistic altogether, using a regression line deviation)
  - Transfer - the same procedure was used for the reliability of the transfer score as used for the automatisation score, a regression line deviation between the score in phase A and phase B. An alternative score that was used was the percentage improvement from phase A to phase B
  - Memory and understanding - here the Kuder-Richardson 20 formula was used, which is a measure of internal consistency.
- Multicolinearity - although all the independent variables used were scales used from theTRAM-1, it is highly likely that they would measure more or less the same constructs, and hence their correlations would be quite high, "…[b]ecause all the independent variables used in this study were derived from the TRAM-1 … battery, **none could really be viewed as being independent** from the others [own emphasis] (p. 63) . Van Aswegen nevertheless calculated Pearson Product Moment Correlations for the TRAM-1 variables and later states that the results are encouraging due to the "…low interrelationships between the independent variables….[which] gives the assurance that no predictors included in this study, measure the same psychological construct" (p.88)
- Correlations between the independent and dependent variables - Spearman Product Moment Correlations were calculated for these variables. In addition, multiple regression analyses were conducted in order to obtain a measure of the predictive power of all variables concerned, predictor (independent variables) and criterion (dependent variables)

- Control of third variables - according to Van Aswegen, so-called third variables may need to be controlled as they may have an effect on the outcome of the results. The variables identified were biographical (age, education and sex) and company.

Results - Descriptive statistics

In the descriptive analyses of the results, it is evident that there were significant differences between the scores of the in-house tests and those at the technical college. Even though the tests measure very nearly the same thing, the differences were evident. Intercorrelations between the college scores and the in-house scores did not reveal any significant findings. However, the results showed that the in-house tests were easier than the college tests, as employees did better on average in the in-house tests.

Van Aswegen notes that the sample selected was a non-probability sample and was, in fact, pre-selected based on those employees who had already proven themselves in the work place. This fact may have skewed the results towards the positive. However, according to the results, the standard deviation of the TRAM-1 composite score as well as the automatisation and transfer scores were high, which means that there were those who scored quite poorly and those who scored quite well. Thus "[the pre-selected sample] did not have a significant influence on the reliability of the data, that is, the possibility of restriction of range[70] has been minimised" (Van Aswegen, p. 69).

Results - multivariate statistics

Four of the five prediction variables yielded high reliability results. Automatisation was the only predictor that yielded a low reliability measure, and according to Van Aswegen should be taken out of the TRAM-1 battery. Each independent variable was graphically correlated with the total criterion score and Van Aswegen concludes that "…all the relationships between the independent variables and the total criterion score were found to be linear"[71] (p. 72). However, Van Aswegen has acknowledged the unreliability of this predictor. Graphically, the relationship between the TRAM-1 composite score and the total criterion score yields the most linear result.

A brief look at the relationships between the independent (TRAM-1 predictor variables) and the dependent variables (in-house and college results) will now follow.

Product moment correlation coefficients reveal that all four criterion variables were significant (at the p = 0.05 level) bar the results for the automatisation dimension. The table presented by Van Aswegen (1997, p. 73) indicates that strong correlations exist between the TRAM-1 and the criterion measures. The strongest correlations were found between the Speed score and basic mining (0.51), the Speed score and in-house Total (0.49), as well as the TRAM-1 Composite score and the Overall score of all the criteria results (0.48).

4.3.13.4 Conclusion and recommendations

Based on these results, Van Aswegen states that the automatisation dimension should be dropped from the TRAM-1 battery (ostensibly for this sample) and that it is possible to use only the composite criterion results, in order to predict the outcome if no other results are used. Significant results were

70 Huysamen (1999) states that "…restriction in the range of scores on the predictor and/or criterion variable due to [decisions taken which lesson the odds of prospective students entering tertiary institutions] has the effect of decreasing the correlation between the predictor and criterion" (p. 133). See De Beer (2000), section 4.3.19 who also makes mention of restriction of range.

71 This is by and large true. However, upon closer examination of the graph depicting the relationship between the total criterion score and the automatisation score, the linear feature is somewhat clouded by what may be outliers, which if taken out (determined by some pre-determined cut-off point) may not really result in a linear relationship.

those yielding 0.25 or above for at a sample size of at least 100 (Anastasi, 1958, in Van Aswegen, 1997).

Considering the fact that speed is part of most of the learning processes involved in the dependent variables, Van Aswegen states that it is no surprise that the relationship between speed and the criterion scores is high. Nine of the fifteen correlations yielded for accuracy and criterion variables were significant, yet "[i]t is suggested that the absence of a relationship between automatisation and the criterion variables could be explained by the fact that automatisation is a measure of **rote learning**, while all the criterion variables reflect learning for **understanding**" [own emphasis] (Van Aswegen, 1997, p.75). Twelve of the fifteen correlations were significant for transfer and the criterion measures. For memory and understanding, all correlations were significant as well as for composite score and criterion measures. Interestingly, as mentioned above, the composite score yielded the highest significant correlations "…[which] …brings up the point that composite criterion scores might be more suitable as criterion measures to determine predictive criterion validity" (Van Aswegen, 1997, p. 77).

Van Aswegen briefly looks at the results of the TRAM-1 dimensions and the correlations with the subjects that the employees had to pass, both in-house and college results. He concludes that the TRAM-1 can be used effectively for determining the performance on these tests, as well as the fact that the results can be used in a diagnostic manner, to identify which cognitive aspects are necessary in order to study and pass the various courses, but emphasises the strong relationship between the TRAM-1 and overall criterion score as being "…the most important criterion scores in the study, as it gives an indication of delegates' *[sic]* overall performance on the total development course…" (Van Aswegen, 1997, p. 81). In the event of no other information being available, the TRAM-1 composite score "…can be used with high confidence, as a valid predictor of successful performance on the **total** development programme" [own emphasis] (p. 82).

Van Aswegen proceeds to calculate the predictive power of the independent variable in forecasting performance. Multiple regression analyses are used for this purpose and the results are summarised below:

- Predictive power of the speed of learning - yielded between 21.4% and 24.2% of the variance explained for total college results, total in-house results as well as the overall criterion score. It is the best predictor of overall performance
- Predictive power of accuracy of learning - yielded a number of significant results, explaining variance for many factors
- Predictive power of automatisation of learning - did not yield any significant results, supporting the idea that it should be withdrawn from the test battery. It is the poorest predictor of performance
- Predictive power of the transfer of learning - also did not yield significant results
- Predictive power of memory and understanding - yielded an especially good result for explaining variance for mathematics. It is the third strongest predictor of overall performance
- Predictive power of the TRAM-1 composite score - can be used "…as a diagnostic tool during the identification of employees who might have difficulty with mathematics…" (p.86). It is the second strongest predictor of overall performance.

The influence of third variables

As noted above, biographical, educational and work-place factors were designated as third variables, that is, those variables that could potentially influence the results. The relationship between age and the independent variables were all negative, and Van Aswegen notes that increasing age corresponds with slower speed in learning (age and speed are significantly negatively related). This may have implications for older employees, since the speed at which they perform may impact on the scores, even though they are adept at handling the information, and hence "…may be unfair to older employees" (Van Aswegen, 1997, p. 90).

No significant relationships were found between level of education and independent variables. In addition, Van Aswegen states that the correlations were not consistent, indicating that the very low levels of education some employees had, made no difference on the test scores. The location of the companies did make a difference however, with employees from rural areas performing less well than

those from urban areas. Van Aswegen states that this is not surprising, since urban employees have greater exposure to intellectual stimulation, albeit in an informal manner.

The main conclusion drawn from this validation study is that the TRAM-1 can be used effectively for the prediction of employees for programmes of further enrichment, and is necessary in the mining industry since the low levels of literacy and the cost involved in selecting the wrong candidates can be inhibitory for companies and is also suitable in a multicultural context. Van Aswegen notes that perhaps future research can include more independent predictor variables such as the Raven's Progressive Matrices; and also that a larger sample be used that could perhaps include females (who are not often found working as miners). As mentioned above, the automatisation dimension should be removed from the battery for this particular sample only, but not removed *per se*, as it is valuable, forming an "…integral part of the TRAM-1" (p. 98).

4.3.14 Die voorspelling van akademiese sukses binne konteks van 'n alternatiewe universiteitstoelatingsbeleid, (The prediction of academic success within the context of an alternative university admissions policy) - Master's study conducted by A. Nel (1997)

4.3.14.1 Introduction

The dissertation written by Nel (1997) has bearing for it examines the use of the dynamic assessment battery, the APIL test as developed by Taylor (1996). The APIL here is used in a larger battery as one of several potential predictors of academic success at tertiary-level education. The main focus in Nel's study is to ascertain the usefulness of the APIL and to determine whether the then current programme of selections and admissions at the Rand Afrikaans University was indeed valuable as a means of predicting success in tertiary education. The context of the study is a project developed by RAU entitled Project 100, the aims of which are to facilitate newcomers' entry into the university. Any student who was admitted based on an alternative admission had to proceed through Project 100. Among other things, this project encourages students by way of lending social and general support, study guidance, reading and thought development as well as language development. As part of the ability assessments, the APIL was used for its unique contribution as a learning potential assessment device - the only learning potential assessment instrument used in the selection battery. The study yielded interesting information about the APIL used in this context. The APIL is of importance because of the "…percentage improvement and development which the testee makes during the process" [own translation] (Taylor & Martiny, 1995, in Nel, 1997, p. 52).

Other test batteries included the GSAT,[72] two language proficiency tests (English and Afrikaans) as well as the traditional M-scores used by many institutions. Once again, this research was prompted by the seemingly disparate educational opportunities given to different demographic groups in South Africa. A more equitable approach was sought and this led to the inclusion of the APIL test battery. The present study focuses on the contribution of the APIL and the relationship of the APIL with other tests as well as the criterion results and how this test works in comparison to the other test batteries. The results of the GSAT, M-Scores and language proficiency tests do not receive too much attention here. Correlations were calculated in order to assess the validity of the various predictors; regression analyses were conducted to assess the predictive validity of the various predictors and discriminant analyses were conducted in order to ascertain whether students had indeed been correctly classified into the right groupings based on their end results.

The usual preliminary chapter details the need to devise a more equitable assessment battery for those students who had received inferior quality school education; and in order to assess these individuals, Nel found it necessary to review the conventional test batteries. Nel rightly argues that the onus on universities to choose the right students for the right courses has never been greater. Language once again plays a major role in any selection battery as it has been evidenced in the literature that language is a very powerful predictor of university success, primarily because in order to grasp fundamental issues in any direction of study, language proficiency is of paramount importance.

72 See also Lipson (1992) section 4.3.9 and De Beer (2000) section 4.3.19 who also make use of the GSAT.

"The value of the language argument is, logically speaking, that there are few things in life which we can learn without the use of language" [own translation] (Nel, 1997, p. 80).

Nel states that learning potential "…reflects the possibilities of an individual to learn new concepts and skills, which are of a cognitive nature, to process and to apply these skills" [own translation] (Nel, 1997, p. 7). Nel also takes cognizance of the importance of non-cognitive factors such as personality, interests, career values, study orientation and study strategy as well as career development. She provides the background of the project, including information about the previous dispensation in South African education, the process of transformation and the problems inherent in this process, different models of education and the unfortunate consequences this has had on the admission of students to university. She also gives attention to the unequal financial subsidies allocated to the previously segregated universities. The idea of transformation of higher education encompasses ideals such as democratisation, liberalisation, socialisation as well as the modernisation of the systems in place. Nel unequivocally states that part of the transformation process entails that the emphasis universities should place on non-authoritarianism, non-paternalistic attitudes as well as a greater degree of transparency.

Nel lists the following universities as playing key roles in the process of transformation and policy changes towards this goal: the University of Natal, the University of Port Elizabeth, and the Rand Afrikaans University[73] (the focal area of Nel's research).

Access and admission to universities in lieu of the transformation process can be underscored as a move away from conventional methods of assessment towards a new method of assessing learning potential (Nel, 1997). Nel mentions and briefly explains the Teach-Test-Teach (TTT) programme run by the University of Natal from 1988-1995 as an example of an alternative admissions policy implementation.[74] She states that the TTT programme represents a total breakaway from the conventional modes of assessment (and hence a move away from total reliance on previous schooling). The programme is highlighted because of the "two-fold" approach of development and selection. Potential students are afforded the opportunity to develop in line with academic standards and are then selected based on the degree to which they can master information that would be expected from them at tertiary levels. Nel states that the process of the TTT programme includes a phase which oversees the development and implementation of the intensive TTT programme, a phase which encompasses the teaching (distance education preparedness programme, which entails three texts that are to be mastered for the up and coming selections test), a testing phase which is the period of selections and finally the second teaching phase which credits students with attending a basic course in the Faculty of Social Sciences and Law.

The information gathered from the TTT programme allows the University to make an informed decision as to who should be allowed to proceed with first year studies and who should be routed to a bridging programme (an expansive degree year). This is the only intervention programme aimed at alternative selection models which Nel mentions are dynamic in nature. Four universities are mentioned by Nel as implementing alternative structures but that are not dynamic such as, for instance, the University of Stellenbosch which at the time of her research considered various ways of implementing alternative procedures. These included a possible bridging year for candidates who were unprepared for tertiary studies, the facility of improving students' results by way of academic programmes as well as possible collaboration with other post-secondary institutions in order to better prepare students for admission into the university. Nel mentions that learning potential tests would be given attention as alternative forms of assessment but does not divulge what types of learning potential assessment tools could be used.

The University of the Western Cape is cited by Nel as another example of a university which aids students by way of a mentoring programme aimed at lending support for new students. However, no

73 These three universities' initiatives in the field of dynamic assessment are detailed in sections 4.5.2.1, 4.5.2.7 and 4.5.2.8 respectively.

74 See section 4.5.2.1 for a brief discussion on the TTT programme.

indication of dynamic assessment is given. Lastly, she cites Rhodes University as an institution embracing the concept of affirmative action. Rhodes strive for no discrimination in a manner which will not adversely affect academic standards. The discretion of the dean is used as well as potential students' non-cognitive factors, such as leadership, motivation and home language. No dynamic assessment procedure is used at Rhodes either. The University of Port Elizabeth is the last university cited by Nel as having alternative admission policies towards disadvantaged prospective students. However, this university merely complements access by means of academic support and development.[75]

Dimensions of prediction

Nel briefly introduces the concept of prediction and criterion variables, and defines the one in terms of the other as the measure of relation between the predictor (variable X) and the criterion variable (variable Y). For the regression analyses Nel made use of the Lotus 1-2-3 software programme. Academic achievement was used for the criterion scores, and according to Nel, criterion variables should themselves be valid and reliable and should also reveal high discrimination value. However, defining exactly what is meant by academic achievement in her study, Nel focuses on the first semester's final marks as indication of academic success. The issue with this sort of criterion is that candidates can be classified into either a "pass" or "fail" group, which neglects the wealth of information about the students in between the extremes. An alternative would be to classify students into groups such as "pass", "fail", and "probability of gaining entrance to the examinations". Nel further states that merely comparing the students' average first semester scores may not be directly comparable owing to the different types of courses undertaken. Here, the process of transforming the scores into z scores[76] comes into play. These normalised standardised scores are at least comparable.

Nel cites the work of the TTT programme at the University of Natal, as well as that of Skuy, Hoffenberg, Visser and Fridjhon (1990) and of Skuy and Shmukler (1987)[77] as evidencing some local South African research into dynamic assessment. Nel's study pays only limited attention to dynamic assessment. Nel does acknowledge the work of Feuerstein and Vygotsky as precursors to dynamic assessment. However, as learning potential assessment is deemed important, it is included in the RAU alternative admissions battery (1997).

4.3.14.2 Research design

The main hypothesis as mentioned already is that it is assumed the battery as used by RAU is a valid and reliable predictor of academic success. Sub-hypotheses have been delineated by Nel and are all essentially the same. The only difference between them is the specific test used as predictor in each hypothesis. Only those which have a bearing on dynamic assessment will receive attention.

1. The first null hypothesis of importance to this study states that there is no significant relationship between students' results on the Learning Potential test and their academic success (students who were admitted via the alternative admissions policy)
2. The next hypothesis of importance to this study states there is no significant relationship between students' scores on the Learning Potential test, the GSAT, the English reading comprehension test as well as the M-scores **in combination** and their academic achievement (students who were admitted via the alternative admissions policy)
3. Regarding discriminant analysis, the next null hypothesis states that, through the use of the APIL, GSAT, Afrikaans Reading Comprehension test and English Reading Comprehension test, no satisfactory classification can be made either as to whether students will pass or fail, about those who will gain access to the examinations

75 See section 4.5.2.7 for the latest information on this university's efforts which would not have been available to Nel at the time her study was conducted.

76 See Shochet (1986) section 4.3.1.3 who also made use of transformation scores.

77 These two texts are not cited in this study. See APPENDIX 1 for reference details of these works.

4. On a variation of the above hypothesis, Nel cites another similar hypothesis which states that through the use of the APIL, GSAT, the Afrikaans Reading Comprehension test and the English Reading Comprehension test, no satisfactory classification can be made as to whether students will be successful or unsuccessful.[78]

The research design is a non-experimental correlational design (Nel, 1997). The measuring instruments are those test measures which measure learning potential (APIL), general intelligence (GSAT) and reading comprehension (English and Afrikaans Reading Comprehension). In order to determine whether matriculation examination results can be used as a satisfactory predictor, M-scores were used as well (as predictor variables).

Nel discusses the merits of each measuring instrument and pays attention to the respective reliabilities and validities and overall use of each test. Looking specifically at the APIL test battery, Nel states that, according to Taylor, the battery is intended to identify those candidates who would most likely benefit from further university training. Three sub-tests[79] were used in the selections process, namely,

- The concept formation test which is designed to determine the testees' ability to form concepts and also to determine reasoning ability. Each item consists of a number of sub-items which the testee has to study and must then decide on which item does not fit in with the rest of the figures. This supposedly gives an indication of "g" or general intelligence, first hypothesised by Spearman
- The learning curve test in which the testee is given a work book as well as a dictionary. The dictionary translates the symbols into meaningful units. This test is carried out over four sessions. The point of the sessions is to familiarise the student with the task so that he/she will be able to recall which symbols refer to what
- The memory and understanding test is a multiple-choice format, during which the testee must answer questions in relation to the dictionary but at this stage the testee no longer has the dictionary in front of him/her.

Nel indicates the reliability and validity results of the test battery. According to Taylor (in Nel, 1997), the Kuder-Richardson formula (KR-20) yields 0.87 and 0.81 respectively for the Concept Formation Test and the Memory and Reasoning test. Two variables which indicate improvement in the learning curve test are percentage improvement and regression line deviation, and the correlation between them is equal to 0.92. Item intercorrelations are also high. The validity of the APIL rests on the sample used by Taylor in the validation study which yields a correlation coefficient of 0.67 with a criterion (an examination written by testees) as well as 0.79 (with another criterion). The test is said to be culture-fair.

Regarding matriculation results, it is assumed by Nel that the M-scores will not correlate with academic achievement and will thus prove ineffectual as a predictor of academic success. The sample used in the study consisted of N = 274. According to the alternative admissions project 192 students were recommended (but only 188 were accepted by the various faculties) and 82 were not recommended by the project. By and large, states Nel, those students entering the alternative admissions project were those from previously disadvantaged schools. Due to logistical problems (Nel, 1997) 171 of the original 188 students are included in the sample, all of which completed the APIL test battery. Considerably fewer students completed the English and Afrikaans Reading Comprehension Tests. Regression analyses as well as discriminant analyses were carried out using SPSS.

78 It is interesting that Nel should state her null hypotheses in the negative, as the common practice is to state null hypotheses in such a way that results will support them. In this case, however, one expects the null hypotheses to be refuted. She states her alternative hypotheses in the positive and one expects these to be supported.

79 See section 4.4 for further information on the APIL where it is also administered in its shortened form.

4.3.14.3 Main findings

According to the correlations between the APIL and academic success, the first null hypothesis can be rejected and the alternative accepted.[80] The significant correlation between the two variables indicates that the APIL can be used as a predictor of academic success ($r = 0.279$ and $p = 0.0001$). A further hypothesis introduced by Nel states that the combination of M-scores, APIL scores as well as GSAT scores yields a significant relationship between those admitted via the alternative admissions programme and academic success. According to the multiple regression analyses of the above mentioned three predictors, "…all three independent variables deliver a significant contribution to the regression…" [own translation] (Nel, 1997, p. 108).

Nel also carried out two hierarchical regression analyses in order to determine which variables (and to what extent) explained prediction variance. The initial analysis first included the M-score as variable (due to the fact that most students seeking admission will have an M-score) followed by the GSAT and APIL test scores. Nel also reversed the order and placed the APIL second and GSAT third in the second analysis to investigate whether there were any differences. Interestingly, there was no difference between the two analyses - each revealed the same degree of variation explained by the presence of the variables. In both analyses, altogether 11% of the variance was explained by the combination of the three variables. The M-score together with the GSAT explained 10% of the variance; and with the addition of the APIL, this rose to 11%. When the APIL was calculated before the GSAT, it increased the explained variance from 6% to 9%.

Next, discriminant analyses were run in order to determine whether the test battery could distinguish between those students who were successful or unsuccessful and in a more focused attempt to ascertain whether students could be correctly classified into the "pass" and "fail" as well as "no entrance to examination" categories. Regarding the three classificatory groups, the question arises as to how the students who completed the alternative admissions battery faired. In other words, the question is, could the test battery correctly classify students into one of the three above-mentioned groups? The predictors included the APIL test, the GSAT as well as the M-scores. The reading comprehension tests were discarded owing to the small sample number who completed them. A 50% correct classification resulted from the discriminant analysis. Nel details the figures necessary for correct classifications due to chance which amount to 38%.[81] Hence, "[i]n total, the three tests combined correctly classify the students significantly more so than correct classifications left to chance" [own translation] (Nel, 1997, p. 114). However, closer inspection of the data reveals that the GSAT yields a correct classification figure of 50.82% (above the chance level of 38%), and the APIL and M-scores respectively yield 49% and 50%. Yet the two classification groups ("no entrance to examination" and "fail") are problematic. The three tests cannot adequately discriminate between these students and hence the null hypothesis has to be accepted, in other words, the APIL, GSAT and M-scores cannot adequately discriminate between students who pass, fail and gain no entrance to the examination. The only group that can be correctly classified using these three tests is the "pass" group (Nel, 1997).

Owing to the possible homogeneity of the two groups ("fail" and "no entrance to examination"), Nel decided (based on the above poor classificatory attempts) to merge the two groups and to re-try the analysis with only two groups, namely, "successful" and "unsuccessful". This analysis yielded a 64% correct classification. Chance of being correctly classified was 49.9%. Thus the correct classification appears higher than the chance factor. (However, the M-scores alone were not a good predictor of group membership, 53% correctly classified, only marginally higher than the percentage correct classification due to chance - 49.9%). Yet the null hypothesis could thus be rejected in the instance of this two-fold classificatory system - the APIL, GSAT and M-scores combined could adequately discriminate students into the two correct classification groups - "successful" and "unsuccessful".

80 It is perhaps pertinent to note that H0 should be rejected and Ha be accepted.

81 This does sound quite high for a figure that is due to chance alone.

In sum, all null hypotheses pertaining to the APIL test battery were rejected bar the one including the three classificatory groupings. One must, of course, bear in mind that with the present survey of results, this dissertation explores the usefulness of the APIL albeit in combination with other tests. In her conclusions, Nel lists (in order of predictability), that the Reading Comprehension Tests, correlated best with academic success, followed by the GSAT, APIL and M-scores. It is not surprising that the M-scores correlated weakly (yet significantly in this study) with academic success, as it has been shown in much literature, that with disadvantaged students (coming from disadvantaged educational backgrounds), M-scores will not predict well, as it is based on prior learning. However, one cannot neglect the finding that the M-score accounted for greater variance explained in the regression analyses than either the GSAT and APIL.

Nel's findings suggested to her that the GSAT and APIL predicted equally well on academic success, noting that the two tests may measure similar properties[82] and also supporting the contention that the APIL is a culture-fair[83] test. Ironically these two findings may appear contradictory, seeing as the GSAT is a conventional static test (in other words measuring prior learning). For both the GSAT and APIL results, the sample used in Nel's study performed worse than those samples used in the validation studies conducted for the GSAT and APIL. The conclusion reached is one of cautious optimism signifying that the tests be used but not in isolation. This applies to the APIL as well. Results of the disadvantaged students on the APIL were lower than those obtained with the norm group, yet this correlates with academic success.

4.3.14.4 Conclusion and recommendations

In essence, the battery as used by RAU is useful and the APIL[84] as used in the battery is also deemed useful. Variances that cannot be explained by the battery are attributed to non-cognitive factors (Nel, 1997). Factors which may have negatively impinged on the research results include the manner of testing, not all students were tested on the same day and anxiety may have played a role in adversely affecting the results. Of concern here is the use of the APIL. Nel mentions that the APIL is intended for use on smaller samples, as it is necessary to invest a lot of time to the correct interpretation of instructions (which are quite difficult). Although individual attention is required as well, Nel's study did not seem to give as much attention to this. Nel does acknowledge that it is possible that those who did not completely understand the questions may have been negatively affected by the lack of individual attention (as prescribed in the APIL manual). This could have resulted in scores lower than the true potential of some students.[85] The marking of the tests is time consuming, and attention to detail is paramount, an area which may have resulted in unnecessary errors on the part of the marker.

Nel also considers the possibility of transfer between the two tests, (GSAT and APIL) during which time after a cessation between testing sessions, it may have been possible that the students could have transferred some "training" in cognitive tests from the one to the other. She supports this contention by highlighting the correlation between the two tests. Perhaps the most glaring critique given the APIL is the time taken and the costs involved in administering the test. According to Nel, the GSAT is computerised, faster and cheaper than the APIL. Taking into account the correlation between the two, it would seem that testers would opt for the cheaper and faster test. However, if the GSAT is thought to be incomplete, the APIL should be employed to give the student two chances to prove himself (Nel, 1997). If an institution would prefer a more culture-fair assessment test, then the APIL should be considered of primary importance over and above the GSAT. Nel also suggests that

82 Engelbrecht (1999) also concluded that the APIL and GSAT measure the same construct.

83 Note Engelbrecht's conclusions (1999) section 4.3.16.4 which do not deem the APIL a culture-fair test.

84 According to De Villiers (1999), the sub-tests of the APIL and academic performance yielded correlations varying between .043 and .45 ($p < .01$) (Kotze, van der Merwe and Nel, 1996 in De Villiers, 1999). This study had 5000 first year university students as a sample and made use of dynamic assessment principles. However, Engelbrecht's (1999) results were not encouraging with regard to the APIL battery.

85 Once again, one recalls that Feuerstein (1979) states that improperly trained administrators can have such a negative effect that testees' scores are deemed low because lack of potential, when in fact this is not the case.

the APIL test format could be changed to a multiple question format as this may save time. Of course this criticism should not be taken lightly as this would mean that the test construction would need to be revised with all its complications and costs, as the format of any test has an influence on results.

The research conducted by Nel (1996 - 1997) may have precluded any use of the computerised adaptive test (LPCAT) as developed by De Beer (2000). Perhaps for future research, the APIL can be computerised or indeed the LPCAT can be used as well as the APIL in correlation studies and these two compared for correlations and the like. The main contribution of Nel's findings to this present study is the fact that the GSAT and APIL, even though they make meaningful contributions towards prediction of academic success, do not yield much more when combined than when used separately (Nel, 1997).

4.3.15 Uncovering potential: dynamic assessment of non-verbal reasoning ability in educationally disadvantaged children - Master's study conducted by A. Gewer (1998)

4.3.15.1 Introduction

During 1998 Gewer conducted research into the non-verbal reasoning ability of disadvantaged children by making use of dynamic assessment. Gewer interchangeably uses the word "capacity" and "ability" when discussing learning potential. As with many studies emanating from the University of the Witwatersrand, Gewer's made use of the LPAD developed by Feuerstein (1979), the RCPM (Raven's) as well as the Rey-Osterreith Complex Figure Test (ROCFT) and, in the analyses of results, made use of qualitative measures as well, namely, a scoring method developed by Lurie and Kozulin (1996) which added to the quality of interpretation that could not be inferred from the quantitative results.

Gewer (1998) introduces the reader to the concept of dynamic assessment and relates the current situation in South Africa regarding assessment of disadvantaged children. Gewer states that mention has been made of the medical model[86] when diagnosing learning disabilities in children. This model attributes the source of the problem to the child and not the environment. This makes sense when one considers the ideas of Vygotsky which specify that adult or peer mediation aids in the development of children's cognitive development. The medical model, it is assumed, does not allow for this interaction to be considered important. Duffy and Wong (1996) state that the medical model arose from an attempt to diagnose and label certain disorders in patients, linking these disorders to some biological aetiology. "The medical model left at least two important legacies in traditional psychology. One is the reliance on diagnostic labels…[t]he other legacy is the assumption of authority and power by the professional over the patient" (p. 95). In support of this notion, Gewer reiterates the necessity propounded by Kriegler and Farman (1996) (as quoted in Gewer [1998]), that the "…provision of equitable educational services will not be effective in the South African context, if the conceptualisation of special needs continues to focus on intrinsic **deficits**" [own emphasis] (p. 2).

Gewer points out what many dynamic assessment researchers at times fail to emphasise and this is that the goal is not only to develop potential within each child but also to enhance his/her ability to function effectively. The two-fold process involved in dynamic assessment encompasses assessment as well as remediation, the latter often not receiving the attention it should. Gewer reviews various models of dynamic assessment, such as those of Feuerstein, Budoff, Campione and Brown, Guthke and Wiegenfield as well as Carlson and Wiedl,[87] and briefly attends to the nature of assessment in each of these models. Focusing on Feuerstein's LPAD, Gewer outlines the battery of tests included in the LPAD, the rationale behind the test itself, the concept and idea behind Mediated Learning Experience. It is important at this stage to note what Gewer emphasises, when discussing the LPAD, as Haywood and Wiegenfield (as quoted in Gewer, 1998) state, that dynamic instruments have in the

86 Dissatisfaction with the medical model or prescriptive model as Lidz (1987) refers to it, "…led to the development of diagnostic-prescriptive teaching" (p. 17) which in itself does not really adhere to dynamic assessment definitions as such. But this was a step away from exclusive reliance on conventional tests.

87 See Chapter 2 for a brief discussion of these theorists work.

past been shown to deliver adequate reliability and validity results when used in a static mode. Whether this notion justifies the lack of verifiability and reliability of dynamic assessment instruments as a whole when used in a dynamic manner is, however, questionable.

The goal of Gewer's study was to find an instrument which can adequately address learning potential for disadvantaged students in South Africa. Three main aims addressed are:

- The usefulness of dynamic assessment in a South African township
- The extent to which mediation will aid in the measurement of cognitive modifiability of the these educationally disadvantaged students
- To test the usefulness of qualitative information gained from a scoring system known as the Lurie and Kozulin Qualitative Scoring Method.

Instruments able to measure ability in black children are often instruments normed on white children. The situation may be even worse when considering test instruments for those children who also possess a learning problem. The LPAD is used to ascertain the severity of the learning problem, at the same time promoting the development of each child, as opposed to merely serving as a classificatory instrument. Gewer states that dynamic assessment may be the answer to the present education system which emphasises Outcomes Based Education,[88] "[d]ynamic assessment provides a mechanism for facilitating the realisation of such outcomes, by invoking the mediation process as a basis for discovering learning potential" (1998, pp. 12-13).

Gewer states that studies have been conducted in South Africa using the LPAD but not specifically for the "…general black school-going population" (p. 13).[89]

Hypotheses

The four main hypotheses are:

- H1 - Children undergoing testing on the RCPM will significantly improve on their below average scores once they have undergone mediation as implemented during the use of certain tasks in the LPAD, as opposed to those children who do not receive mediation
- H2 - Children undergoing testing on the ROCFT will significantly improve on their below average scores once they have undergone mediation as implemented during the use of certain tasks in the LPAD, as opposed to those children who do not receive mediation
- H3 - The quality of responses from the above mentioned two tests, from those children receiving mediation on the LPAD tasks will be significantly better than those who do not receive mediation
- H4 - Those receiving mediation on the ROCFT test will be better able to transfer their learned skills onto a similar figure as opposed to those who do not receive mediation.

4.3.15.2 Research design

A total sample size of 72 children referred to an adolescent psychiatric unit were placed into either an experimental (N = 48) or control group (N = 24). The sample for the control group was matched to the experimental group. The mean age was 10.96 but ranged from 9 - 15. The children referred to the clinic were those supposedly experiencing learning difficulties as identified by their respective teachers. Gewer details a number of pertinent biographical as well as some family background information details as well as the status of the child (i.e. the problem associated with the child at present). Once again, one of the few reasons cited for the poor performance of these children in school (undoubtedly exacerbated by some children's family background and standard of living) is the former education system in South Africa.

88 Coosner (1999) section 4.3.17 pays attention to the role played by Outcomes Based Education (OBE).

89 Gewer's study is not a study based on the "general black school-going population" either. Other research in South Africa has indeed included black school children as this present study shows.

Gewer pays attention to the specific measurement instruments used, namely the Raven's Coloured Progressive Matrices,[90] the Rey-Osterrieth Complex Figure Test[91] as well as the LPAD. An important issue which Gewer brings to the fore is the usefulness and indeed applicability of reliability in the field of dynamic assessment. He quotes Lidz (1997) as stating the inappropriateness of the need for reliability[92] measures. "While intra-test reliability for the pre-test and posttests is important, high test-retest reliability would invalidate the DA [dynamic assessment] process as DA seeks **to bring about** [entrench] **change rather than stability**" [own emphasis] (Gewer, 1998, p. 22).

The Lurie and Kozulin qualitative analysis of matrices is explained and the scoring system is based on four underlying criteria on how a problem is solved:

- Defining the problem
- Basic principles underlining the problem
- Implementation of the principles of the problem and
- How precise the response in fact is.

As has been mentioned, the reason for inclusion of this information is to assess the effects of mediation in a qualitative manner.

Dynamic assessment using the RCPM and the LPAD set variations 1 as mediation

The experimental group was asked to complete the RCPM, after which, in groups of five, they received mediation on the LPAD Set Variations 1. Trained post-graduate students completed the mediation sessions. The children were then asked to complete another session of the RCPM. Gewer asserts that this serves as the post-test measure as well as indicating transfer. The control group received no mediation in between the pre- and posttest. A slight modification in research design becomes evident at this point, as Gewer states that owing to logistics, some students could not complete the pre-test session nor the mediation on the same day. In order to control for this variable, two broad groups were created, those who completed the pre-test and mediation on the same day and those who did not.

Mediation was not conducted in the home language of the children and as such Gewer required the assistance of nursing staff for some translations. This aspect of the research is interesting, as it may not be construed as entirely scientific; firstly, because the main aim of the study is to assess the usefulness of the LPAD in a "township setting" (where mediation in the home language would have been better) and secondly, the nursing staff may not have been the best mediators, since they were not trained for the mediation they administered.

Dynamic assessment using the ROCFT and an equivalent figure for purposes of mediation

The two-fold process of measurement included copying a figure shown to the children, and then drawing the figure from memory. Mediation followed, suggesting ways in which to complete the task. Once again the process was repeated, followed by an instruction to draw from sight a second equivalent figure as well as to draw it from memory. Assessment of the second figure would give an indication of how successfully the strategies taught were transferred. The control group underwent the same procedure but the testees were not exposed to the mediation. At this point, only near transfer of skills was assessed and not far transfer.

---

90 Note that Gewer makes mention of the fact that the "Raven's Matrices" according to research cited in his thesis cannot be used as an "…accurate measure of intellectual ability within a cross-cultural context" (p. 19). He does however makes use of the Raven's Coloured Progressive Matrices. The tacit assumption then is that there is a difference between the two forms.

91 Note also the studies by Andrews (1996) and Lipson (1992) who used the Complex Figure Drawing Test from Feuerstein's LPAD.

92 This same sentiment is echoed in the Teach-Test-Teach programme (TTT), run in the past by the University of Natal (see section 4.5.2.1).

4.3.15.3 Main findings

Two-way analyses of co-variance were conducted to ascertain the effectiveness of the mediation for the RCPM. The reason for this was that Gewer had to rule out any significant variables (pre-test score variability) which may have played a role in the data collection between the two broadly defined groups (those receiving mediation on the same day as the pre-test and those who did not).

For the qualitative scoring method employed, Gewer made use of three independent raters, who would rate the figure drawings of the children. In order to determine the level of agreement between the raters, the Kappa statistic was used which according to Gewer "…is a measure of agreement between independent raters to certain categories…it measures the proportion of times the raters agree, to the maximum times they **could** agree" [own emphasis] (Gewer, 1998, pp. 27-27).

One-way analysis of variance was conducted for the ROCFT results. Both the copying and memory tasks were analysed. The qualitative analysis was once again carried out in a slightly modified form, and the Kappa statistic was once again calculated, based on the level of agreement between raters.

RCPM results

The different days on which certain parts of the research was/was not conducted was not significant. However, no significant differences were evident for the groups in terms of mediation, even though there was a "…noticeable difference in the gradients, in favour of the experimental group" (Gewer, 1998, p. 29). The shift between pre- and posttest scores for the experimental group indicated that this group's performance increased from below average to almost average whilst that of the control group remained more or less stable below average. For a small set of matrices in the RCPM (B8 - B12), the experimental group showed a significant improvement as a result of the mediation.

The qualitative analysis of the RCPM did not reveal satisfactory results, as the inter-rater reliability was not significant. Gewer states that the raters were not accustomed to the matrices nor the principles behind the scoring technique.[93] Nevertheless, Kappa analyses were carried out and was shown to yield lower levels of agreement for the experimental group, thus suggesting that there was a greater shift in the responses for this group as opposed to the control group. In light of the mediation received this would be a positive result. The control group did not vary as much in their responses (yielding a 58% level of agreement). The overall level of agreement on the RCPM was 15.4 (which according to Landis and Koch [1977] as quoted in Gewer [1998] indicates "…poor agreement beyond chance" and 37.82% for the control group "…fair to good agreement beyond chance" [*ibid.*]).

ROCFT results

No scoring system was used for the equivalent figure, only qualitative analyses follows. Significant results were found in favour of the experimental group, evidencing improved performance of the task after mediation. The control group showed no such significant improvement. Once again inter-rater reliability was not successful owing to the reasons cited above. However, results showed that there was variability on the tasks performed by the experimental group as opposed to the control group. This means that after mediation the experimental group showed changes in their copying and memory tasks.

Although the overall results for the RCPM were not significant, important differences were seen in the experimental group as opposed to the control group. This was emphasised in the qualitative response analyses. A sub-section of the RCPM (B8-B12) did evidence significant differences in favour of the experimental group. Gewer states that the RCPM when used statically may not be useful in a cross-cultural context, but most certainly can aid in the learning potential field with regard to educationally

93 This begs the question of whether these raters were trained well or not, and if not, why did they receive such poor training if their roles are so crucial to the outcome of the study? These methodological issues seem to play a negative role in many such studies. See Chapter 5 for further elucidation on this issue.

disadvantaged children. The mediation on both the RCPM as well as the ROCFT lasted 3 to 4 hours. Moreover, the qualitative analyses conducted would enable teachers to identify specific areas that may need attention more closely. The types of issues can be addressed when the RCPM is used in a dynamic manner as opposed to a static manner. The identification of cognitive areas which may need working on is a key characteristic of dynamic assessment. Even though inter-rater reliability was poor, Gewer states that this problem has plagued dynamic assessment since its inception. Over reliance on "…statistical conceptualisation detracts from the focus on …. the quality of change" (p. 42).

4.3.15.4 Conclusion and recommendations

Looking at the results and in an attempt to explain the variations in scores between the RCPM and the ROCFT, Gewer states that performance in the ROCFT is easier to interpret since the test lends itself to finer evaluation than does the RCPM. What occurred with the RCPM was that although there was a decrease in the number of incorrect responses, there was not the equivalent gain in correct responses. The ROCFT, as Gewer states, "…lends itself more to being a process-based assessment instrument by the very nature of its application" (Gewer, 1998, p. 44). This is useful information for those researchers who would attempt remediation with students. As with the RCPM, the experimental group was better able to complete the figures from memory as well as sight as opposed to the control group. A low Kappa is indicative of low levels of agreement between pre- and posttest results, which supports the effect of the mediation, as for the control group, high levels of Kappa suggest less changes between the pre- and posttest scores, meaning that no change was evident as they were not exposed to mediation.

Aspects such as planning, strategising and anxiety played a role in how the two groups performed on the tasks. The control group evidenced higher levels of anxiety (when viewing their small drawings, for instance) than did the experimental group. Impulsiveness, ability to pay attention or plan as well as motivation were all aspects which were assessed by the dynamic manner in which the measurements were taken. It is aspects such as these that allow teachers and potential mediators the opportunity to help the child after he/she has undergone testing, and would not be forthcoming if he/she were assessed by means of conventional testing.

The main crux of Gewer's study was to show the usefulness of dynamic assessment as a measure of learning potential which would be able to service disadvantaged children in a black community. In so doing, he concluded that "[t]raditional, static assessment procedures would not have allowed such an investigation" (p. 47).

Gewer states that time constraints did not allow for further use of LPAD sub-tests which may have yielded more information about numerical and verbal abilities; and that the poor inter-rater reliabilities hampered the study somewhat. Further to the study, the qualitative measuring/scoring system could well be employed to investigate the Raven's Standard Progressive Matrices as well. These form part of the LPAD. Also of note, a fact to which Gewer does not refer more than once, is the language used in the study. Home language mediation would have been preferable and may have had an impact on the scores achieved after mediation.

4.3.16 Leerpotensiaal as voorspeller van akademiese sukses van universiteitstudente, (Learning potential as predictor of academic success of university students) - Doctoral study conducted by M. Engelbrecht (1999)

4.3.16.1 Introduction

The aim of this study was to determine the effectiveness and usefulness of the APIL Battery[94] as a potential predictor of academic success of first year university students and also to investigate how the results compared to those of the GSAT and SAT. Engelbrecht sampled students from two different universities in South Africa, namely Potchefstroom University for Christian Higher Education (PUCHE)

94 See the study conducted by Nel (1997) section 4.3.14 which also makes use of the APIL in a similar setting.

and Rand Afrikaans University (RAU). In her study, Engelbrecht provides information about the current status of tertiary education in South Africa regarding predictability of tests and the present dissatisfaction with psychometric tests as tools with which to predict academic performance. This sentiment is of course not new and is echoed by almost all researchers in the field today. Engelbrecht foregrounds the efforts of tertiary institutions and their on-going efforts to try to find a valid test battery or battery of tests which will yield accurate information in a fair and non-discriminatory fashion. She does, however, point out the folly in thinking that an open-door policy will necessarily yield high pass rates; "…an 'open-door' policy does not guarantee the delivery of successful students…" [own translation] (1999, p. 2). The aim, she states, in accordance with the Green paper for Transformation of Higher Education, is to do away with Matriculation exemption and entry criteria and to use the envisioned single national qualification criterion according to which students will or will not receive entry to tertiary studies.

The previous approach to assessment (when based on the psychometric method) was one of inhibiting students from entry whereas a more proactive stance will better secure students' futures at tertiary institutions (assuming that the tests and assessments can be shown to be valid predictors of success). Engelbrecht states that the idea is eventually to replace the matriculation results as criterion with evaluation of learning potential.[95] Engelbrecht states further that there has been " … relatively little research conducted in South Africa with the eye towards predictions of academic success" [own translation] (1999, p. 5).[96] Her study seeks to add to the current research already available. If the research yields positive results and the APIL is seen to function accurately, then, Engelbrecht states, it can be used in place of the existing admissions tests used at both universities.

The four main aims of the study are:

- To determine whether the APIL is a good predictor of academic success of first year university students
- To determine whether the APIL is a better predictor of academic success than the GSAT and/or Senior Aptitude Test (SAT)[97]
- To determine whether the APIL is a culture-fair test
- To develop norms and cut-off points for different university study directions.

Engelbrecht briefly discusses the influences of Vygotsky (1978), Feuerstein (1979), Budoff (1968) as well as Campione and Brown (1987) (as cited in Engelbrecht, 1999) and their efforts in the field of dynamic assessment. She takes a critical look at the on-going debacle between the psychometric and dynamic approaches towards assessment. Criticism is levelled at the former and evidence of the latter's usefulness is given. She quotes Ionescu and Jourdan-Ionescu (1983) as stating that psychometric tests ignore underlying mechanisms of cognitive functioning.[98] She also reiterates the sentiments made by other researchers that no remedial aspect is implied in psychometric testing - a major point of departure for dynamic assessment as well as the beneficial use of dynamic assessments' approach towards diagnoses and description of cognitive functions.

She draws attention to the difference between "trainability"[99] and "learning potential" (citing Taylor 1994; 1996), noting that they are complementary concepts and not equivalent to each other;

95 However, it has been stated in numerous studies that the learning potential of students should really serve in a complementary role and not be used exclusively.

96 This is questionable as it is evident that research has been conducted and it is debatable as to what constitutes "little" research. However, she does note the research of Boeyens (1989b), Shochet (1992) (not cited in this present study), Skuy, Zolezzi, Mentis, Fridjhan and Cockcroft (1996) and Kotze (1996) (not cited in this present study).

97 "The SAT was constructed with the aim of giving a measurement of a number of aptitudes for pupils in standards 8, 9 and 10, and of adults that can be used for counselling and selection purposes" (Smit, 1996, p.233).

98 One is tempted to think of Feuerstein's cognitive map (1979) in this instance as providing a comprehensive assessment of cognitive functions.

99 See section 4.4 .

trainability referring to more practical type tasks which are often expected of workers in a physical context more so than a mental task. Briefly, Engelbrecht discusses the history of learning potential approach, highlighting four main phases in its development, that is,

1. The structural/conventional approach - an important precursor to the learning potential approach even though it is not directly linked to learning potential. Pioneers in this field such as Thorndike, Thurstone, Cattell, and Jensen[100] are mentioned as well as the so-called "g" and "s" factors which typified the period and approach of this method of investigating intelligence.

2. The information processing approach - Engelbrecht states that the literature at times yields multi-directional starting points for this approach and that certainty does not always govern its origins of this approach. She speculates that parallel approaches developed when researchers did not necessarily know of others' work. Basically, this approach studies the manner in which information is stored, processed and received. Eminent researchers within this field such as Hick (1952), Broadbent (1958), Treisman (1960) and Sperling (1963) are mentioned, with a brief description of their various experiments which were conducted. The bottleneck approach, capacity approach, Sternberg (1969), as well as the theories of Posner and Snyder (1975), Newell and Simon (1972), Hasher and Zacks (1979) are reviewed.

3. The cognitive approach - an approach recognised for three main approaches within its field, namely;

   - The cognitive correlates approach - supported in part by researchers such as Eysenck (1986), Jensen (1982) as well as Hunt (1985)
   - The cognitive components approach - supported in part by researchers such as Sternberg[101] (1988) and Carrol (1976)
   - The cognitive training approach - perhaps closest to the learning potential approach stemming from the cognitive approach. It is narrowly connected to the above two approaches, and emphasises the need for individuals to possess some skills in order to process new information from the environment accurately; and in this way emphasis is placed on the teaching and learning of new information.

   "According to [the above-mentioned] writers the study of intelligence should not consist of an analysis of intelligence tests, but the information processing approach **should explain intelligence as a component of the cognitive approach**" [own emphasis and translation] (1999, p. 45).

4. The learning potential approach - the emphasis of Engelbrecht's study. The differentiating feature for Engelbrecht, for this approach is that the aim is to modify existing cognitive structures and to promote improvement and in so doing train learners. Here she discusses the efforts of Vygotsky, Feuerstein, Budoff, and Campione and Brown among others.[102]

The next subject of Engelbrecht's literature study is the often researched topic of tertiary admissions policy, but this time the focus is on other countries other than South Africa. She discusses the policies and efforts of these institutions in the light of dynamic assessment. Transformation of various tertiary institutions across the globe in a sense shows that the South African initiative in this regard follows world trends. After highlighting certain South African institutions' efforts, Engelbrecht focuses on

---

100 An awareness of the more prominent theorists within the realm of intelligence testing is assumed, as important and significant contributions made by these past researchers always impinge on present-day thought and theory. As progress is being made all the time within this field, it is nevertheless vital that cognizance be taken of early twentieth century theories and contributions.

101 Sternberg's Triarchic theory of intelligence is an example of the cognitive components approach.

102 See Chapter 2 for a brief discussion on the history of dynamic assessment as it is merely repeated in Engelbrecht's discussion.

overseas' efforts in admissions and their attempts to remain as open to all students as possible. She mentions the policies of Australia, Britain, Ghana, Canada, the Netherlands, New Zealand, the United States and Zimbabwe. She states in sum that "…most South African universities have an alternative admissions policy for prospective students, especially those from disadvantaged groups" [own translation] (1999, p. 98).[103]

Engelbrecht (1999) also gives attention to the debates surrounding predictors of academic success and the role played by biased test instruments. It is this effort to confront bias and yet maintain a valid and accurate measure of predictability that thwarts the field of academic prediction within the dynamic assessment approach. She cites a few sources of bias such as inapplicable content, inapplicable standardisation samples, language bias, different social environments, measurement of different constructs, differential predictive validity and cognitive style. Culture-fair tests do seek to minimise the effects of the above-mentioned or even to rid tests of these characteristics.

Engelbrecht acknowledges the dual feelings of matriculation examination results as accurate predictors of academic success and states that although researchers vary as to their opinions of matriculation results as predictors, one fact is the usefulness of these results for those students lucky enough to have been schooled in a previously advantaged environment, and that matriculation results play almost no role when trying to predict for disadvantaged students. This fact has been emphasised by numerous South African researchers already discussed in this study. The General Scholastic Aptitude Test (GSAT) has, according to Engelbrecht, been shown to be a valid and reliable predictor of academic success provided that the test is administered in the language of the testee.

Engelbrecht states that the Senior Aptitude Test (SAT), is used, in part, to determine academic potential but has not been researched lately with regard to its applicability for different race groups.[104] Engelbrecht also states that the PUCHE has in the past made use of Feuerstein's LPAD as a supplementary admissions test but this has since been scrapped.[105]

4.3.16.2 Research design

"A one-shot cross-sectional design of first year students at two universities was used" (1999, p.vi), consisting of all first year university students at both the PUCHE and RAU who were tested in 1996 (N = 3250). Two thousand three hundred RAU students were tested on the APIL, SAT and GSAT and 950 PUCHE students were tested on the APIL and SAT. Approximately 1000 students represented black students. (Xhosa, Tswana and Sotho among other languages were spoken.) The sample can be construed as an "availability population", and both inferential and descriptive statistics were employed to investigate the results. A further variable used in the study was first year academic success (obtained from various results throughout the year), which was given to Engelbrecht and was included in her database of results. The students represented a cross-section of various disciplines, such as B.A, B.Com, B.Sc. specialising in Biology, B.Sc. specialising in Physics, Mathematics or Economics, B.Ing, (Engineering) H.E.D, Law, Optometry, Nursing and Social Work. The sample was divided into the following three groups:

- Those who passed completely (all subjects passed)
- Those who passed partially (those who passed the minimum number of subjects to enable them to proceed to the second year)
- Those who failed (those who failed their first year and in so doing were unable to complete their studies in the required time).

103 This finding concurs with the findings of this present study. The informal interviews conducted with all tertiary institutions in South Africa (see section 4.5), although not yielding very much in the way of dynamic assessment research, did yield information about the many alternative admission policies available for all students, especially those from disadvantaged backgrounds.

104 In fact, Smit (1996) states that the latest revision of the SAT was published in 1978.

105 See section 4.5.2.5 for the latest information.

Criterion variables included the average result obtained at the end of the first semester and/or the second semester as well as both semesters taken together in the different subjects. The outcome of the third year students (for those who did complete their studies in the required time) was also used as a criterion variable.

Only 10 sub-tests of the SAT were administered, all six sub-tests of the GSAT, as well as the shortened version of the APIL, consisting of three sub-tests. The APIL was used in the capacity of a learning potential test, in order to determine whether the students could master cognitive information in a formal educational set-up, and results give an indication of the potential to develop cognitively (Engelbrecht, 1999). Of the three sub-tests that were administered in the APIL, only one sub-test included a training session. A learning curve was worked out for each student on the learning curve sub-test, in which speed of learning is ascertained. The sub-test consisted of four sessions, with a learning phase in between each. The percentage increase in improvement between the first and last sessions, is revealed as the amount of learning that had taken place. The APIL was originally standardised on a sample of 122 technical workers and one of the aims of the Engelbrecht study was to norm the APIL on tertiary students.

Multiple regression analyses as well as discriminant analyses were conducted on the data collected, in order to determine predictors of the identified criterion variables and to determine group membership based on variables that could adequately distinguish between the above-mentioned three groups.

4.3.16.3 Main findings

Engelbrecht analyses all the disciplines' results from both universities separately and offers descriptive statistics and correlation analyses (between the independent and dependent variables) for the data set.
The correlation results for the University of Potchefstroom yielded the following interesting findings:

- None of the variables correlated with academic success for the B.A, B.Com, B.Com (Law), B.Sc. (Home economics) and B.Sc. (Biology) groups
- Three sub-tests of the SAT correlated with academic achievement for the B.Sc. (Mathematics, Physics and Economics) group
- Two sub-tests from the SAT correlated with academic achievement for the B.Ing. student group as well as the four sections of the second sub-test of the APIL
- 3 sub-tests of the SAT correlated with academic achievement for the H.E.D group.

Correlation results for the Rand Afrikaans University yielded the following interesting results:

- None of the variables correlated with academic success for the B.A, Law, B.Com, B.Com (Law) and B.Sc. (Mathematics/Physics/Economics) and B.Ing
- Three sub-tests of the SAT as well as one sub-test of the APIL (the first sub-test, which is the memory sub-test[106]), as well as three sub-tests of the GSAT correlated with academic success for the Optometry group
- One sub-test of the SAT, two sub-tests of the APIL (the concept formation and memory sub-tests) as well as the total IQ result of the GSAT correlated with academic achievement for the B.Sc. (Biology) group
- Four sub-tests of the SAT, three sub-tests of the GSAT, as well as the non-verbal IQ and total IQ of the GSAT, and one sub-test (one of the tests of the learning curve sub-test) of the APIL correlated with academic achievement for the nursing students
- Nine SAT sub-tests, two sub-tests from the APIL (sub-tests one and two), all the sub-tests from the GSAT, the verbal and non-verbal IQ as well as the total IQ scores from the GSAT correlated with academic achievement for the social work group.

Of note is that both universities' correlations regarding B.A, B.Com and B.Com (Law) groups did not correlate significantly with academic achievement.

106 Which does not include a training session.

Engelbrecht also conducted multiple regression analyses which yielded the best combinations of variables which would best predict academic achievement. Based on the results of the regression analyses, the following was identified for the University of Potchefstroom:

- The various sub-tests from both the SAT and APIL do not contribute much to the prediction of achievement. The highest contribution to the explained variance comes to the fore with the B.Sc. (Mathematics/Physics/Economics) group for whom one sub-test of the SAT explains a certain amount of variance. For the black students, the first sub-test (concept formation) also contributes a significant amount to the prediction model. In combination, the SAT and APIL contribute more than when used on their own. For certain groups only, the regression model obtained (using both the SAT and APIL) can be used but not for other study directions.

Results for the Rand Afrikaans University indicate that:

- As with the Potchefstroom sample, the SAT, APIL and GSAT did not contribute much in the way of predictability and their contributions are subsequently of a low order. The highest contribution of explained variance emanates from the GSAT total IQ score, for the B.Sc. (Biology) group, the contribution of the first sub-test of the APIL for the B.Ing. group as well as one sub-test from the SAT for the B.Sc. (Biology) group.

Some findings in the Engelbrecht study are at odds with what others[107] have found in terms of "best predictors", but this, she cautions, is usually the result of different combinations of predictor variables being used within the regression model. In sum, regarding the findings for both university groups, Engelbrecht states that, based on the results of the regression analyses, no specific tendencies come to the fore when predicting academic achievement and hence there should be no reliance on cognitive predictors of academic success of first year students.

The APIL, SAT and GSAT were used as variables in discriminant analyses. The APIL was included to measure learning potential. Three criteria were used in order to classify students into one of three groups, namely, "passed", "passed partially" and "failed". Other variations included amalgamating the two "passed" groups, so as to end up with two classifications. The results of the discriminant analyses used in the Engelbrecht study were also utilised for differential prediction and selection of referrals. It is vitally important that it is known if prospective first year students can be grouped into different groups based on cognitive performance, as well as whether or not the measuring instruments are able to allocate individuals to the correct classification groups correctly. The following notable deductions of the discriminant analyses for the University of Potchefstroom group are highlighted:

- The SAT contributed more so than did the APIL to the classification functions for a variety of study groups
- The APIL did not contribute very much to the classification function of any group
- The percentage correct classifications that were made across all groups, however, ranged from 62% (B.Com group) - 91% (B.Sc. Mathematics/Physics/Economics group). A higher percentage of correct classifications was made when the classification variable consisted of two groups ("pass" and "fail") as opposed to three groups ("pass", "pass partially" and "fail")[108]
- No single predictor can be highlighted as significant regarding the weight of the variable in the different classifications functions in which it is represented.

The following notable deductions of the discriminant analyses for the Rand Afrikaans University group are highlighted:

- The SAT delivered a greater contribution to the classification functions than did either the GSAT and APIL
- The APIL did not contribute very much to the classification function of any group

107 In this instance, the APIL test developer, Taylor (1995, 1996) as cited by Engelbrecht (1999).

108 Note that Nel (1997) section 4.3.14 also had to reduce the number of classification groups from three to two which were also very similar in nature to the groups used by Engelbrecht here.

- The percentage correct classifications that were made across all groups, however, ranged from 62% (B.Com group) - 85% (B.Sc. Biology group)
- As with the Potchefstroom group, no single predictor can be highlighted as significant regarding the weight of the variable in the different classifications functions in which it is represented.

Even by implementing discriminant analyses[109] with the different cognitive measurement instruments as discriminant variables, not one good discriminator could be selected in order to differentiate between successful and unsuccessful students.

4.3.16.4 Conclusion and recommendations

Engelbrecht states that based on the all the findings for all study groups at both universities, the "…combination of …the various sub-tests of the APIL, SAT and GSAT explain between 17% and 56% in the variation of academic success for the Potchefstroom group and between 13% and 58% for the Rand Afrikaans University group" [own translation] (1999, p. 285). Furthermore,

- Only a limited number of regression models can be devised for certain groups at both universities
- Only a few groups can benefit from discriminant classification models (some for whom it will not be valid) in which students will be correctly classified into pre-determined groups
- The APIL does not contribute significantly in explaining variance in academic achievement and thus, based on the results, should not be used to determine academic success of first year students[110]
- Even when the GSAT, SAT and APIL are combined, they do not yield that much explained variance in terms of prediction of academic success, and that not one of the predictors is better than either of the other two
- The APIL is no better a predictor of academic success than either the GSAT and SAT
- For the above reasons, states Engelbrecht, the APIL[111] cannot be used as a culture-fair[112] learning potential instrument
- As a result of the poor results found for the APIL, Engelbrecht's second main aim of the study was not investigated, namely, determining norm and cut-off points for the APIL for the various study fields.

Engelbrecht states that non-cognitive factors should be looked at closely when embarking on prediction batteries, and that the APIL should perhaps undergo further refinement in terms of item analysis in order to determine the difficulty level of determined items as well as to determine a discriminant index for the battery. In sum, she claims that the prediction of academic achievement was not aided by the inclusion of a learning potential test but this is surely debatable as other studies have proved.

109 Engelbrecht (1999) states that nowhere in the literature to date (up to 1999) was there any literature found to make use of the APIL in discriminant analyses in the context of academic achievement. However, the Lopes, Roodt and Mauer (2001) (section 4.4) study which uses the APIL in discriminant analyses should be noted, although this was conducted in a different context, namely the work place. Also see the results of the Watson and Ncapayi (1998) study (section 4.5.2.7).

110 Perhaps it should be stated that, though this may be the case at these two specific universities, one cannot unequivocally make this generalisation.

111 The APIL as used in this study, made use of the shortened version of the battery. Only one of the three sub-tests was dynamically administered. Perhaps the total APIL battery could have yielded other more positive results. However, the fact remains that the shortened version, too, is valid and reliable as other studies, reviewed in this present dissertation, have shown.

112 See Nel (1997) section 4.3.14.4 whose results show that the APIL is a culture-fair test.

4.3.17 Dynamic assessment - a practical strategy for school educators - Master's study conducted by C.D. Coosner (1999)

4.3.17.1 Introduction

The research conducted by Coosner (1999) is perhaps one of the most qualitative research results detailed in this study. As has been mentioned in Chapter 3 (section 3.2), the branch of dynamic assessment dealing almost exclusively with Feuersteinian theory (that is, studies whose main aims are to detail qualitative findings on mediation and re-mediation) do not form the focus of this study. There are a number of such known studies which are not mentioned. However, it is felt necessary to include at least one such study.

In her study, Coosner's focus is the applicability of Feuersteinian mediation within the school classroom context and the effects that mediation may have on teachers as well as learners, in particular Structural Cognitive Modifiability (SCM). The new Outcomes Based Education[113] (OBE) system would be more than fitting as a vehicle in which to administer Feuersteinian tenets. Her argument rests with the fact that OBE is especially conducive to the use of dynamic assessment, in South Africa, in particular. OBE is purported to be learner-centred, holistic and process-orientated. Two important terms used by Coosner in her study are MLE (Mediated Learning Experience) and SCM (as above). The aim of this study was to detail the experiences of teachers who were trained in MLE and who were taught the fundamentals of SCM. As these aspects of education are focused on the mediator (the teacher) it was imperative to investigate the teachers in this study more so than concentrating on the students (as is more often the case with dynamic assessment research).

Coosner introduces the socio-economic state of South African education and highlights the areas that need to be looked at from a monetary and socio-political point of view. The shift from knowledge-based curricula to outcomes education promotes the need for more process orientated learning, hence the need to train teachers in mediational skills. "Both the educator's mediational role and an understanding of cognitive processes impact on Outcomes Based Education" (1999, p. 4).

OBE is grounded in the belief that all individuals can learn, which underscores the tenets of dynamic assessment and learning potential. It would thus seem that the both OBE and dynamic assessment are uniquely suited to one another. Coosner casts an analogy with OBE/dynamic assessment and the traditional education system/static tests. She points out the criticisms levelled at dynamic assessment (such as the costs involved as well as the time taken to administer this type of assessment), but notes that these can in part be overcome. Coosner cites the Department of Education's (1997) envisioned role for educators in the OBE structure. Whilst the list promotes the OBE philosophy one can clearly see the almost overlapping ideas this approach has with dynamic assessment in terms of mediation. The two are closely aligned. Coosner's study has as its aim a "…practical, implementable and user-friendly [way in which to promote dynamic assessment and as such]…empiricism is not an issue for the practical strategy which is proposed" (1999, p. 46). The main emphasis in Coosner's study is the mediation administered in the classroom. Her dynamic approach is thus not the usual test-train-test/pre- posttest approach.

4.3.17.2 Research design

Coosner states that both qualitative (unstructured, informal interviews) and quantitative (an anonymous questionnaire) analyses were conducted in the study. The sample consisted of voluntary educators from three primary schools in the Cape Peninsula and so constituted a convenience sample. Due to attrition, the number of teachers present in the study totalled 15. Brief training of about an hour in total was provided before the strategy was implemented. Of the many approaches to

113 In a statement given by Professor Kader Asmal (1999), it is stressed that active learning takes place through outcomes based education and he also highlights the role of teachers in this initiative. Teachers need to be developed more so on a professional basis now than ever before.

dynamic assessment which Coosner discusses (test-train-test, graduated prompts, LPAD, and testing-the-limits), she highlights Lidz's curriculum-based dynamic assessment approach as being nearest to her research endeavour. In this approach three versions of a task are designed for learners that may be experiencing problems, and are placed in a pre-test - mediation - posttest evaluation set-up. Lessons are then planned according to the assessments of the problem areas encountered. Coosner states that this approach relies quite heavily on Feuerstein's mediational criteria. The pilot study thus consisted of a lecture in which teachers were taught mediational skills and strategies, a workshop on the Feuersteinian approach to cognition as well as a follow-up evaluation and assessment one month later. The training process encompassed and input phase, elaboration phase and an output phase.

Input phase

This phase emphasised the need for MLE in the classroom. Training of educators lasted an hour and the content used by Coosner consisted of literature compiled from Jensen and Feuerstein (1987), Lidz (1991), Haywood and Tzuriel (1992) as well as Kozulin and Presseisen (1995) (as cited in Coosner, 1999). The educators completed an evaluation at the end of the training session. Certain criteria of MLE were explored and taught in this phase. Criteria[114] such as "intentionally and reciprocity"; "meaning"; "transcendence"; "competence"; "self-regulation and control of behaviour"; "sharing behaviour"; "individuation/separateness"; "goal planning"; "novelty and challenge"; "self-change"; "optimistic alternative" as well as "feeling of belonging" were issues considered. These aspects mentioned above were ideas that were intended to be operationalised within the classroom setting.

Elaboration phase

This phase took place two days after the input phase discussed above. According to Coosner, this phase lasted 40 minutes and was more interactional than the input phase. The aim of this phase was to concretise the notions of MLE as experienced in the classroom. Educators were asked to keep a journal of their experiences and were given a form to complete. Their experiences were tabulated and forms were to be handed in one month after the second phase took place. The educators were exposed to the manifestation of incorrect cognitive functioning of learners in the classroom context and were trained on how to "spot" faulty thinking in terms of the Feuersteinian cognitive map.

Output phase

This phase took place one month after the elaboration phase. This session lasted 40 minutes and comprised educators' experiences in the class after having received brief training on MLE. These discussions as well as the questionnaire that had to be completed formed part of this phase's activities.

4.3.17.3 Main findings

Qualitative results

Coosner notes that a recurring theme in the questionnaires that the educators filled in was one of "…consciously being self-reflective in … practice" (1999, p. 62) and that this had significantly enhanced their interaction with the learners. It would seem then that the MLE training had achieved its intended goal. Other noticeable reactions to the initial training were teachers' responses to their learners, being more empathetic, in-tune with them, discussing their progress and even tracking them in the classroom. Reciprocal effects such as increased learner self-esteem were also a "spin-off" from the training. These findings were noted at all three schools where the programme took place.

114 See Feuerstein (1979) who discusses these types of aspects regarding the examiner-examinee relationship.

Quantitative results

According to the questionnaires completed by the teachers, the majority felt that they were unsure about identifying the specific phase in which a learner might be experiencing cognitive difficulties. This led Coosner to aim for a more thorough input training phase, as this is crucial in identifying and remediating learners. However, according to discussions with the teachers, most felt that they had integrated the MLE into their teaching quite sufficiently, so much so that their teaching was now more learner-centered than before. However, only 67% felt that their teaching was more process-centred and according to Coosner "…were having difficulty making the paradigm shift to transformational OBE…" (1999, p. 63). Perhaps a more important finding or at least more elucidating one in the context of dynamic assessment in South Africa, is the fact that Coosner notes the dissatisfaction of teachers toward conventional assessment and the lack of integration of dynamic assessment into the mainstream of assessment.

Although the informal discussions with teachers yielded positive results regarding learners' attitudes, in terms of motivation and effort, the questionnaire results showed that on average only 60% of the teachers noted a difference in the learners' attitudes and motivation. All the teachers mentioned that they were more self-reflective after having gone through the programme, and most felt that the strategy was practical and implementable. 80% were motivated to continue using the strategy yet only 53% noted their willingness to continue with further training. Coosner states, in this regard, that this may have been due to the nature of the programme, it needing further refinement, more focused tracing and longer training sessions.

A practical dynamic assessment strategy for school teachers (educators)

Coosner reflects on the findings both the qualitative and quantitative results in order to come up with a better programme that could be used within schools. The pilot study above was used as a means towards refining the programme. Coosner discusses the programme in greater detail, identifying six sub-phases that would span the practical strategy. The first sub-phase seeks to introduce the teacher to the realm of dynamic assessment, process-orientated education and move towards OBE from the previous static mode of education. This sub-phase will typically last forty minutes. The second sub-phase serves to introduce Feuerstein's concepts of MLE as a means towards implementing dynamic assessment, and the criteria involved in mediation. This would last sixty minutes, followed by a twenty minute break.

The third sub-phase introduces the teachers to Feuerstein's cognitive map and is more academic and theoretical in nature. The duration of this sub-phase is forty minutes. The fourth sub-phase provides the training necessary to identify problems that learners may experience during any of the three phases (input, elaboration and output phases). This lasts forty minutes. The fifth sub-phase allows the teachers to use their own experiences as examples of how meditation could take place in certain instances where it is needed, and teaches them to cope more effectively with remediation and strategising and planning in these tasks. This lasts for forty minutes. The last sub-phase emphasises the need to keep the teachers motivated and "on track". It necessitates that they remain consistent in their efforts. This is an ongoing process.

4.3.17.4 Conclusion and recommendations

Coosner states that "[i]t is important that throughout the process, the golden threads of intentionality and reciprocity, meaning and transcendence are constantly mediated" (1999, p. 69). Furthermore, she states that the prototype used in the study has refined areas of weakness, so that the strategy can be better utilised. Perhaps one of the more prominent findings in this study is the new attitudes of the teachers towards their learners. This new appraisal speaks favourably of the programme. Also indicated by the teachers was the fact the this dynamic assessment strategy was not biased against any group, immigrants or otherwise. Of course, this is one of the tenets on which Feuerstein's theory was built. However, even though Coosner states that this method allows teachers the opportunity to assess each learner on an individual basis, it is perhaps not that easy to implement in practice all of the time principally because it presupposes well trained facilitators. Another important conclusion is that even though a learner might not exhibit difficulty in mastering the material, learning potential is not

necessarily absent. It is this aspect which needs to be identified in schools that are still at a disadvantage. Early identification of "at risk" learners is also another point which can be identified through these means.

Coosner states that a major limitation to her study was the lack of a strict test-teach-test format. One can understand her reservation but, on the other hand, one can appreciate the more qualitative stance of the study. The time allocated to teacher training was too brief; the number of teachers (N - 15, after drop-outs) was too large for a pilot study; and important biographical information from the teachers was missing. This study failed to show that, when dynamic assessment is taught as a teaching strategy, more good comes from it than not. A qualitative look at mediation, however, proved to be quite successful.

4.3.18 Disadvantaged students' academic performance: analysing the zone of proximal development - Doctoral study conducted by A.B. De Villiers (1999)

4.3.18.1 Introduction

The main aim of De Villiers' lengthy 1999 study is to determine alternative predictors of academic performance other than the traditional matriculation examination results used in South Africa. It seeks to differentiate between those students who can most benefit from the mediated experience and those who cannot. "The study relates the students' response to meditation to their academic performance and analyses the role that non-cognitive factors such as motivation, approaches to learning and learning strategies play in cognitive performance" (De Villiers, 1999, abstract). The study was conducted through a number of phases with a sample of 400 first year students at a South African Technicon studying in one of four faculties. The first study comprised the use of two dynamic assessment batteries and the effect of the mediated lesson; the second study looked at correlations between the static and dynamic tests; while the third study sought to find predictors of academic performance (De Villiers, 1999). A conclusion found in this study echoes the conclusions made by other researchers in the South African context (Boeyens, 1989b; Shochet, 1986; Shochet, 1994; Zolezzi, 1992) that for less modifiable students, the best predictors of academic success are the conventional manner of testing, namely matriculation examination results. For more modifiable students, the results indicate that alternative predictors are more helpful.

This broad study looked closely at the role of mediation, the role of moderator variables such as demographics, the predictive validity of conventional versus dynamic tests as well as the differences between more and less successful groups of students. The unique problem with which selection committees are faced with at South African institutions of higher learning is the need to assess accurately and select those candidates with a "predetermined probability" of success while, at the same time, taking into account culture-bias and previous backlogs in learning. "Prior learning has always been considered a good predictor of future learning, but this assumption depends on the quality of the prior learning experiences" (De Villiers, p.13). In other words, the need to choose successful candidates from a pool of previously disadvantaged students is challenging. Once again, the manifest ability and true potential ability are the focal point (Boeyens, 1989b; Shochet, 1986).

The notion of the zone of proximal development in this study is understood to relate to the sensitivity with which candidates are able to benefit from mediation.

Four sets of academic performance scores were used in this study, namely, the year marks of students (calculated from test and assignment results) also referred to as the grade point average, an end of the year examination mark, the final year mark as well as a the number of credits obtained in the year of study. Data analyses were carried out using SPSS sofware. De Villiers delineated three hypotheses, namely,

- A significant difference will be encountered between students who receive the mediated lesson in comparison to those who do not
- Academic prediction will be enhanced by the use of dynamic assessment as opposed to the use of static tests only

- The differences between three sub-groups of students will be significant in terms of modifiability, schooling, demographic variables and length of time spent at the tertiary institution. More modifiable students will perform better than less modifiable students, with the latter having better prediction results garnered from the more static and conventional test scores such as matriculation examination results.

De Villiers mentions as part of his third hypothesis that no differences will be marked between the different socio-economic groups, nor between male and female nor between English first language and English second language students.

Seven measuring instruments were used in the study with two being classifiable as dynamic tests. The Learning Potential Test (LPT) and the Conceptual Reasoning Test[115] (CRT) make use of non-verbal reasoning tasks and according to De Villiers makes use of "novel material" not seen before in students' previous education experiences. The LPT makes use of a pre- and posttest phase with a lesson in between, whereas the CRT "…only uses a lesson with hints in the first part of the lesson" (p. 25). These two dynamic tests are group administered and are formal in the sense that the structure of a formal pre-test, standardised training and formal posttest is conducted. Once again, owing to time and cost effectiveness, conducting the study on a one-to-one basis would not have been feasible. The zone of proximal development was created in a collaborated effort between the researcher and respondents by means of verbal and non-verbal interaction. The pre-test LPT score was used by De Villiers as an indication of static measurement[116] as no mediation had then taken place. The posttest score of the LPT and the only score from the CRT were used as dynamic test scores. Previous academic achievement results were used, rated by the Swedish Rating System (SR) used by De Villiers as a means of standardising matriculation examination results from both advantaged and disadvantaged groups.

Due to disillusionment with previous selection procedures at some tertiary institutions in South Africa, De Villiers points out four such institutions which have made use of dynamic assessment procedures to select potentially successful students, among them, the University of Natal, the University of the Witwatersrand, the University of Cape Town and the Rand Afrikaans University. De Villiers (1999) presents the rather circular and self-fulfilling prophetic argument that since conventional tests seek to assess prior learning, the results for disadvantaged students would prove dismal and thus the success ratio would be similarly low. However, dynamic assessment seeks to avoid this by not aiming to assess prior learning, but only what **can** be learnt.

Although the De Villiers study also concentrates on the role of non-cognitive factors which influence predictions of academic performance (such as interests, personality, attitudes, motivation, approaches to learning, learning strategies, as well as study methods and habits) the main focus here is to relate the findings of the two dynamic assessment tests. However, it must be kept in mind that "…non-cognitive factors, such as motivation, learning strategies…seem to play an important role in the mediation phase of dynamic assessment procedures" (1999, p. 102).

4.3.18.2 Research design

De Villiers employed seven measuring instruments in the study, divided into cognitive measures and non-cognitive measures. The two dynamic assessment measures emanate from the cognitive test battery, which in total consisted of the Conceptual Reasoning Task (CRT), the Learning Potential Test (LPT), The Reading Comprehension Test (RCT), The Mental Alertness Test (MA) as well as the Electrical Aptitude Test (EAT). The non-cognitive measure consists of the Study Processes Questionnaire (SPQ) and the Motivated Strategies for Learning Questionnaire (MSLQ). An overview of each of the various above-mentioned tests is presented, noting reliability and validity issues as well as giving a general description of the instruments. The present dissertation does not go into the

115 See section 4.4 for further results using the CRT.

116 The pre-test of any dynamic assessment can be used and is used as a static measure, as no learning has yet taken place.

details of the tests used but comments briefly on aspects that are pertinent to the chosen topic. Details can be read in De Villiers (1999, p.132-167).

The SPQ measures the type of approach students have towards their studies (non-cognitive factors), namely, a deep approach, surface approach and achieving approach with a composite approach known as deep-achieving. De Villiers conducted his own reliability and validity study for this test and compared his results with those of other researchers. The test can be said to be both reliable and valid for the purposes of his study.

The MSLQ measures non-cognitive factors and "…is a self-report instrument with a seven point Likert scale designed … to assess students' motivational orientations and their use of different learning strategies" (De Villiers, 1999, p.139). Two sections of this test make up the motivation measure as well as the learning strategy measure. Once again, using reliability data from the original devisors of the test as well as a sample study De Villiers indicates that the instrument is indeed reliable with good alpha coefficients. De Villiers also manages to replicate the validity findings that other authors have found for the MSLQ with the exception of one subscale which did not evidence as clear a structural pattern as in previous validity studies. Where there are differences between De Villiers' findings and those of predecessors, he states that "[i]n the cases where the subscales are used, the interpretation of findings would be done with due consideration of the findings of the validity studies done with the present sample of students" (p. 151).

The CRT which forms part of the two sets of tests considered as dynamic assessment (measuring cognitive factors) is a "…non-verbal test of inductive reasoning ability" (*ibid*.). The reason why this test can be construed as dynamic is the inclusion of a lesson with hints for students, so as to ensure a more equitable approach for all students. The test is made up of four problems, namely, series problems, two-way classification problems, transformation problems and operation problems. As with the previous two measures, De Villiers conducted a reliability test with his own sample and compared the results to previous studies, with which comparisons can be made. Validity studies with his sample of students compared favourably with a study conducted previously by Boeyens (1990) (as cited in De Villiers, 1999). De Villiers concludes that the CRT "…has adequate predictive validity" (p. 156).

"The LPT [also a dynamic measure which measures cognitive factors] was developed as a way of operationalising potential … modifiability" (p. 157). It consists of an initial assessment, an intervention and a posttest. De Villiers' own sample compared well with a previous sample's reliability data (a sample used by Boeyens, 1989b as cited in De Villiers, 1999), making both the pre- and posttests reliable. However, he points out that due to a difference in scoring, the reliability of the difference score could not be replicated in his study. De Villiers also correlates the LPT pre- and posttest scores with the EAR and CRT and finds moderate correlations. Owing to the incomparability of the difference scores between his study and the Boeyens (1989b) study, De Villiers made use of the posttest score only, as a measure of change.

The RCT is a cognitive measurement test. It assesses the ability to understand written English. It is not a dynamic test and is "…closely linked to an assessment of formal academic work" (p. 161). Of a number of samples that were used for the assessment of reliability, two samples were found to yield unsatisfactory results. However, the sample used in the study by De Villiers reveals adequate reliability results. As a possible explanation for the poorer reliability results, he offers the fact that the students' first language was not English. For the assessment of validity, De Villiers conducted two sample studies, with students from different faculties. Two intercorrelation matrices were compiled, showing the results of the RCT and MA as well as LPT results. De Villiers states that the "…RCT compares favourably to *[sic]* the other tests in predicting students' academic performance" (p. 163) but cautions that interpretation of results of this test should be viewed with "certain consideration" (*ibid*.).

The MA test "…evaluates the ability to profit from formal academic work, and provides a measure of general intellectual functioning" (p. 164) and thus is not a dynamic test. De Villiers uses reliability data gathered by a previous researcher but gives no indication of any studies that he conducted to ascertain reliability as "it was not possible to calculate a reliability index for the sample in this study" (p. 165). Nevertheless, the results indicate suitable reliability. Validity data are however available for his

sample and he concludes that the MA compares well with other tests (provided in an intercorrelation matrix with the RCT and CRT as well as academic performance).

Lastly, the EAT "…was designed to measure high school students' aptitude for electrical and electronic engineering" (p.166). De Villiers relies solely on the reliability statistics yielded by an Australian study but states that the test can be used to make "…reasonable judgements about individual student aptitude" (*ibid*.). The same is true for the validity studies. De Villiers states that these reliability and validity studies were not conducted for his samples. After a thorough analyses of basic reliability and validity results from his own sample studies, those from overseas and other local studies, De Villiers proceeds to use these tests proving their usefulness within the South African context.

4.3.18.3 Main findings

De Villiers made use of the SPSS and Statistica computer analysis software for all calculations in his study. As mentioned previously, this study comprised three phases, namely an investigation into the effectiveness of the mediation during the administration of the LPT, an investigation into the correlations between the static and dynamic tests and lastly, a third investigation into the predictors of academic performance. The results are now looked at below.

Effectiveness of the mediated lesson - LPT study one

The LPT had both a pre- and posttest with a lesson in between, whereas the CRT only has a lesson preceding the test. Two studies were conducted for the LPT and two for the CRT. For the first study using the LPT, a Solomon Four Group Design was employed as most suitable for the LPT. Four groups were employed, two receiving lessons and two not receiving lessons. The four sets were divided into experimental and control groups. According to the ANOVA results, no significant differences were found for the pre-test groups who received the lesson and those who did not receive the lesson although there was a general improvement in mean scores for those who did receive the lesson. A two-way analysis of variance was conducted to see if there were any significant differences between the posttest groups. This allowed for simultaneous investigation of variables such as effect of the lesson, the sensitisation effect and interaction effect between the lesson and pretesting (De Villiers, 1999).

According to the two way analysis, there was a significant difference between the group who received a pre-test and the group who received none. Hence the pre-test served to sensitise the group and De Villiers concludes that "…the mere fact of being pre-tested affected the performance of the group that was pre-tested" (p. 171). De Villiers offers an explanation as to why there was no difference between the two groups regarding difference scores, namely, that as the control group received a placebo by way of being given hints (as opposed to the experimental group which received the mediated lesson) they may have gained as much knowledge as the experimental group. There was also no interaction effects between the pre-test and treatment. The conclusion drawn from the first LPT study is quite general and vague, with sensitisation effects playing a role in the experiment as well as no significant differences for the groups involved.

Effects of the mediated lesson - LPT study two

Once again the Solomon Four Group Design was employed with random allocation of students to one of four groups. A new sample was used. Once again, there were no significant differences between the pre-test groups for this sample. However, there was a significant difference between the control and experimental groups after a two way ANOVA was conducted, showing that the lesson did indeed play a role. Pre-testing did not have an effect (as with the first sample) and there were no interaction effects. The results for the second study are more supportive of the use of dynamic assessment than the first study. The difference scores were significant and, because the design controls for the practice effect, De Villiers concludes that the mediation was responsible for the significant difference. Firstly, De Villiers attributes the more positive results in the second study to the fact that this study randomly assigned individuals to groups, as opposed to the first study in which students who studied in a specific field were put together. Secondly, the fact that the first sample's control group received a

placebo in place of a mediated lesson (which affected the results) and no placebo was given to the control group on the second study may also have affected results.

CRT test - study one

In order to assess the effectiveness of the lesson given during the CRT, a posttest control groups only design was employed. An experimental group consisted of students who were admitted into the Electrical Engineering faculty at the beginning of the year and the control group consisted of students who were admitted into the same faculty in the middle of the year. The experimental group received the mediated lesson and the control group merely completed the test part of the CRT. According to a t-test, the results showed that there was a significant difference between the two groups with the experimental group scoring higher than the control group. The mediated lesson thus had a significant positive effect.

CRT test - study two

First year Information Technology students were randomly allocated to one of two groups, the experimental group receiving the mediated lesson and the control group receiving no mediation. The results showed that even though the experimental group evidenced higher scores, these were not significant. In order to fully understand the reason behind this, De Villiers decided to split the groups into those students who had attended DET schools and those who had attended non-DET schools. As many studies have thus far indicated, dynamic assessment usually yields positive results for those students who come from culturally deprived backgrounds and as such it is to be expected that the DET experimental students would show a significant improvement in results. This is exactly the case, as compared to the non-DET students' results. De Villiers states that the two groups (DET and non-DET) "formed two distinct groups" (p. 178). Thus according to both CRT studies, it can be concluded that the dynamic assessment procedure involved yielded positive results. The extent to which this dynamic test can predict academic performance when compared to static tests, was De Villiers' next challenge.

Comparison between static and dynamic test results as predictors

The static measures as mentioned above (MA, RCT and EAT) and the CRT and LPT as dynamic assessment tests were compared in their capacity as predictors in a sample of Business students as well as Electrical Engineering students at the end of their first year. Matriculation examination reseults (scores allocation based on the Swedish Rating system - [SR]) were used as the past performance scores. The results of the two studies are now discussed.

Comparison study - Business Students

Pearson Product Moment Correlations were computed to find significant correlations between the variables. Results indicated that the CRT correlated more so with the end of the first year results than any of the other tests (i.e. the MA, SR [Swedish Rating System] and RCT). The Swedish Rating system scores correlated significantly with academic performance but not to the same extent. The traditional static MA test correlated significantly with the SR results (as would be expected since the MA measures for those who would be more successful at tertiary institutions). The correlation between the CRT and SR results were not significant. De Villiers ran a stepwise multiple regression analysis to determine the relative weighting each test contributed to the overall prediction model. The SR and CRT tests contributed the most in the model. De Villiers cites Cohen's effect size (1977) as delineating the value a result has to attain in practical terms before it can be considered large enough for prediction. However, it turns out that the variance explained by both the tests is indeed quite small, with the CRT and SR explaining only 14% and 15% of the variance respectively.

Comparison study - Electrical Engineering students

For this sample, the CRT, RCT and EAT and LPT tests were used. The SR scores were used once again as past results. The main difference between the CRT and LPT administrations, was that the LPT test included a pre-test whereas the CRT included no pre-test. The findings for this sample were

not supportive of the dynamic assessment hypothesis. Pearson Product Moment Correlations were once again run, but yielded only significant correlations between the SR and Technicon results. The rest shows "weak correlations" (p. 184). Another stepwise regression analysis was conducted, which yielded a contribution from the SR scores but when the LPT was included as a variable the percentage variance explained increased, but was construed only as "useful" by De Villiers (*ibid*.).

The two studies above indicate that the SR scores were more useful in predicting academic performance than either the dynamic and static measures. Correlation patterns also differed for both disciplines. One important aspect of the two studies is that only cognitive tests were used and excluded the non-cognitive tests, which as De Villiers points out may be the reason why such a small degree of variance was explained by cognitive measures only.

De Villiers conducted numerous calculations, striving to find a suitable case for the use of dynamic assessment and further investigated the relationship between past academic performance and manifest potential, learning potential and various cognitive and non-cognitive factors. However, the preliminary results mentioned above (CRT and LPT case studies) yield the general findings. Furthermore, the non-cognitive measures consisted of subscales that were also individually analysed and these subscales were also looked at in detail in order to explain the importance of non-cognitive factors and the role played by these factors in selection procedures. De Villiers' further studies highlight the evidence in favour for the use of dynamic assessment tools. These results are touched on only briefly. The sample for both Electrical Engineering and Information Technology students were once again first year students in these respective faculties. Cognitive measures were taken at the beginning of the year as part of the selection procedures and the non-cognitive measures were taken after one to three months.

Non-cognitive factors - Electrical Engineering students

As with the cognitive measures studies above, the non-cognitive test measures were added and correlations were computed for all variables. For the group as a whole only the SR scores correlated in any significant manner with the first year academic results. However, when regression analyses were conducted, a non-cognitive factor as well as the LPT scores explained a significant amount of variance in the model. Hence, although the SR score was the only significant variable which correlated with academic success, the contribution made by the LPT and non-cognitive factors cannot be ignored (De Villiers, 1999).

Non-cognitive factors - Information Technology students

The same variables were used for these students. However, the correlation results showed that the SR scores as well as some of the non-cognitive scores correlated significantly with academic performance. Although no cognitive factors correlated with academic performance for this group, after regression analyses were conducted, the RCT (a cognitive measure) was seen to explain a certain amount of variance in the prediction model, along with non-cognitive measures and the SR scores. Even though the two groups differed with regard to the role played by non-cognitive factors, both groups' predictions were enhanced by the measure of non-cognitive measures.

Measure of previous academic scores with high and low scoring students

As has been shown by other researchers in the dynamic assessment field (Entwistle, Percy & Nisbet, 1977 in De Villiers, 1999; Miller, 1992 and Shochet, 1986), matriculation examination results correlate more positively with SR scores for those higher scoring students than for lower scoring students, for whom matriculation examination results tend to correlate less and less well with academic performance the lower these results are.[117] In order to analyse the role played by dynamic tests, De

117 Zolezzi lends support to this in terms of a dynamic approach to assessment, when he says that"**[t]**he concept of modifiability has important implications in the South African context where traditional predictors of academic success have demonstrated little or no relationship to tertiary academic success" (1992, p. 11).

Villiers divided the sample of students into those who received high matriculation examination results and those who received low results. Not surprisingly, for those lower performing students in Electrical Engineering, the LPT scores correlated significantly with academic performance as opposed to any of the other variables, especially the SR scores which, in fact, correlated weakly and even negatively with academic performance. When correlations were run for the higher performing students, both the dynamic assessment measures (LPT and CRT) showed not only negative and weak correlations but significantly negative correlations (CRT) with academic performance. The conclusion that De Villiers arrives at based on the results, is that different selection criteria should be used when assessing students for Information Technology and to a lesser extent for Electrical Engineering (given the lesser difference between the higher and lower scoring Electrical Engineering students).

Individual Matriculation subjects

De Villiers took a closer look at individual matriculation subjects to determine their respective predictive values as predictive of academic performance. Of the number of regression analyses conducted, it was found that individual matriculation examination results were inconsistent in their prediction of academic performance for students in the different faculties. However, there were individual examples of matriculation subjects playing an important role in predicting academic success some of the time for some students. Three conclusions that can be drawn from the De Villiers study, as he points out are the following:

(1) According to the findings, it can be said that in general the mediated lesson assists in increasing scores in tests
(2) SR scores are still the main predictors of academic success
(3) However, dynamic assessment scores are in general better predictors than static scores especially when the dynamic assessment scores are used in the prediction of academic success for lower SR scorers (i.e. those who received lower matriculation scores). Hence, CRT and LPT scores were better predictors for lower SR scorers, whereas the SR scores were more predictive of academic success for those with higher matriculation results.

De Villiers conducted secondary analyses and studied the "…prediction patterns for different groups of students [which] include the students' response to the mediation provided in the LPT…[s]tudents can also be differentiated along the lines of secondary school attendance, gender, language and socio-economic status" (De Villiers, 1999, p. 205). According to De Villiers' findings, former DET students should not necessarily be assessed using the same criteria as non-DET students. Dynamic assessment tools should rather be used for these students, and the non-cognitive learning factors may be more appropriate in differentiating between more and less successful Technicon students. Dynamic assessment measures did not seem to reflect differences between students' socio-economic status as much as the static tests did (as they should not).

Secondly, female and Afrikaans students were commented on owing to the fact that female students scored higher in the anxiety subscales and thus were prevented from scoring higher marks in final year subjects. For Afrikaans students, it was found that according to their matriculation results, better first year performance results should have been obtained. De Villiers states that English language proficiency may be the reason why they may have underperformed. An interesting finding that should be mentioned is the results from a discriminant analysis that was conducted to "find an equation that best combines the different variables, SR score, cognitive factors and non-cognitive factors, to predict passing or failing" (De Villiers, 1999, p. 237). The discriminant analysis revealed that the SR score as well a non-cognitive subscale (metacognition) were the two variables which correctly classified students into either the passing or failing category. These two variables correctly predicted 73% of student classifications. "This is an indication that it would be incorrect to use the SR score exclusively in selecting students for Information Technology and Electrical Engineering and that a non-cognitive variable such as metacognition strategies plays a role in enhancing prediction of academic performance" (De Villiers, 1999, p. 238). In essence, differentiating between less and more modifiable students could be done when looking at the results of the SR score, the posttest of the LPT, the use of metacognitive and cognitive strategies.

Results of De Villiers' hypotheses

Three hypotheses were set out by De Villiers and were discussed at the end of the research thesis. These three hypotheses and their conclusions are briefly reviewed:

- H1 - There would be a significant difference in the scores between those students who received the mediated lesson and those who received no mediated lesson as part of the dynamic testing procedure
- H2 - There would be a significant enhancement in prediction of academic performance of students using dynamic testing as opposed to making use of static measures, including the SR (Swedish Rating Score) system
- H3 - There would be significant differences between different patterns of correlations among various students based on modifiability, schooling (DET or non-DET) as well as socio-economic status, gender and language.

The first hypothesis was supported mainly in the second study conducted with the LPT test where a mediated lesson was used in the assessment with the experimental group. Using the CRT also revealed significant differences between groups, and the CRT yielded significant results in the second study after more thorough analyses were conducted. The fact that the sample formed two separate groups (DET and non-DET) may have initially influenced the results.

The second hypothesis was "partially supported by the data" (p. 214). Although matriculation examination marks served as the better predictor of academic performance for the group as a whole, DET students were better able to be predicted when the results of the CRT and LPT tests were used. This was not the case for non-DET students, whose predictions were better served by conventional tests, such as matriculation examination marks as well as the SR scores. Not only were dynamic assessment scores better predictors, but the use of non-cognitive factors enhanced prediction for all groups concerned.

The third hypothesis "was generally supported by the results" (p. 242). More modifiable students benefited more from the dynamic assessment procedures than less modifiable students. Neither socio-economic status, nor language, nor gender revealed any differences in prediction. Schooling backgrounds did however make a difference in scores, and had a moderator effect on prediction scores. Static assessment procedures, namely the SR score as well as the non-cognitive measure, were able correctly to predict students who would pass or fail their academic year and accounted for almost three-quarters of correct classifications.

De Villiers does point out a few pertinent issues when he discusses the results of his study. A matter of contention within the dynamic assessment domain is the usefulness of using pre- and posttest designs. The practice effect or pure memorising of test procedure needs to be discussed. The use of control groups as well as implementation of a Solomon design should negate any such practice effect, as this is controlled for in the use of a control group. In other words, one is able to measure the degree to which learning has taken place and true gain in terms of memory and reasoning. The use of a parallel test also negates this effect. Regarding the response to mediation, De Villiers states that "[a] result that stood out and which differs from the results of Shochet's study was the high and significant correlation between the CRT and posttest of the LPT (which are examples of dynamic tests) and academic performance for the more modifiable group of students. Shochet (1986) did not report any predictor of academic performance for the more modifiable students"[118] (p. 260).

4.3.18.4 Conclusion and recommendations

Regarding the use of a mediated lesson, De Villiers states "…the majority of the evidence seems to indicate that the mediated lesson was effective" (p. 243) and the more disadvantaged students were better able to benefit from the mediated lesson than the advantaged students. De Villiers also states

118 Zolezzi (1995) section 4.3.10 concurs.

that the use of the sample, who were tertiary students are, as a group, a select one and a reason why static measures were the single best predictor for the group as a whole may be attributable to this.

Despite the reassuring research results, De Villiers convincingly discusses the usefulness of dynamic assessment. Dynamic assessment should be viewed as a complementary assessment procedure to be used in collaboration with static tests when assessing potential students. Dynamic assessment and the more conventional tests more often than not measure similar constructs such as general intelligence, but it is the way in which dynamic assessment is administered that makes all the difference. Static tests to some extent measure what each test deems to measure as intelligence. However, dynamic assessment allows the testee to make better use of his/her potential during the testing situation. "The possibilities of enhancing the cognitive and affective processes during a period of mediation play an important role in changing the width of the zone of proximal development and in its turn affects the way individuals utilise their intelligence" (De Villiers, 1999, p. 263). Another positive finding of this study was that it offered support for the use of dynamic assessment for disadvantaged students. Since lower SES students initially scored lower on the LPT pre-test (which is essentially a static test, as mediation has not been administered yet), but scored on a similar level to the higher SES group on the posttest score, attests to the usefulness of dynamic assessment. These students were able to benefit from the mediated lesson.

The LPT is a non-verbal assessment instrument and thus the assumption is made that language will not play a role in students' pre- and posttest scores. This assumption is supported by De Villiers' data. When English first language and English second language students were assessed, it was found that English second language speakers performed less well on academic tests at the end of the year (even though their SR scores were on average higher than English first language speakers), a contributing factor may be that since the medium of instruction was English, this may have affected their academic results. Their scores on the pre- and posttest LPT scores were also not very encouraging, producing scores significantly lower than that of the first language speakers. Nevertheless, results on the CRT (dynamic test) indicate that both first and second language speakers performed equally well on the assessment which provides support for the ability of dynamic assessment to "…circumvent the language barriers that exist for the ESL [English second language speakers]" (De Villiers, 1999, p.270). De Villiers' main conclusion from his research is that:

> [o]verall the results obtained in this study seem to suggest that Vygotsky's theory of learning potential can be successfully operationalised and applied in an assessment procedure. Furthermore the differentiation of different groups of students on the basis of their response to mediation seems useful as a tool in improving the validity and accuracy of predicting academic performance of disadvantaged students (p. 275).

Implementing a quota system in which different students are assessed differently is seen by De Villiers as one way of overcoming previous test bias. But the attendant problems with such an approach are numerous. The contentious issue of the quota system is elaborated upon by Miller (1992) and Huysamen (1996). Huysamen discusses models that are used in selecting students for higher education. Three types of quota system models are explored, such as multiple fixed cut-off requirements (in which prospective students have to attain a certain level of achievement on two or more criteria in order to be allowed entrance), compensatory models (in which high scores on one test may compensate for poor scores on another test) and multiple-regression models (in which any variable which is deemed a predictor may be used in a regression model to predict those who may benefit from further education). "Throughout the world it has been found that groups formed in terms of some or other biographic or demographic variable perform differently on tests of cognitive ability (Huysamen, 1996, p. 201). Miller (1992) states that "…a quota system does not solve the problem of selection but merely displaces and contains it within the limits of the quota" (p. 100).

De Villiers' study has demonstrated that the use of matriculation examination symbols as academic predictors of success for disadvantaged students (i.e. previously DET students) is rather biased and not a successful predictor, which is reiterated by Miller (1992) and Huysamen (1996). "Unless an alternative to the matriculation-based points system is developed and instituted without undue delay, universities will face a crisis, not only of confidence from without, but also of conscience from within"

(Miller, 1992, p. 98). These are prophetic words indeed which still have not been heeded to the extent that they should have been.

It is possible that different predictor variables can be used for different demographic groups where the different groups receive different weightings for the same variable. Dynamic assessment procedures are able to identify those students who will most likely benefit from tertiary education, as the test can identify those who are more and less modifiable. In other words, it is at least possible to select promising students who are disadvantaged, as opposed to merely selecting students because they are disadvantaged. Using different criteria in selecting students may seem to be less practical in terms of cost and timing, but looking at the situation from a cost-benefit point of view, tertiary institutions may well reap the rewards later on and hence recoup any money spent initially, by the mere fact that these students have a greater chance of passing their studies. Although the study focused on students from the Peninsula Technicon, the findings are generally true, even if the assessment procedures have to be contextualised to other settings such as university students.

Limitations of the study

As with most other studies in the dynamic assessment arena, a few limitations are evident, which De Villiers summarises. They are in brief:

- Small sample sizes
- The short-term nature of the assessment, in comparison to longer periods devoted to more students and with more interaction within these settings
- Metacognitive processes were assessed based on self-reported questionnaires versus identification of these processes during testing procedures
- Transfer issues were not studied in detail
- More examples of dynamic assessment batteries could have been utilised to evidence greater validity of findings
- A broader scope of intervention strategies such as lessons may have yielded a greater of comparisons to be conducted.

Even though individual testing would be ideal, the costs involved are still at this stage too great to warrant their use. De Villiers' use of the Solomon Four Group Design aided in control of the practice effect to some degree. This study showed that the use of dynamic assessment as a means of selection of students at a tertiary institution was quite beneficial. It has been pointed out that even though post-apartheid South Africa may be a country that has changed in many respects, the education system like many other institutions will take quite a while longer to change to any measurable extent. Thus, although the DET as a name may no longer exist, the actual change will take significantly longer to attain. Hence Miller states that "[students will] remain victims of the DET education system, a situation that in the short-term will not change when the DET changes its name" (1992, p. 99) and this is underscored by Huysamen (1996) who states that "…for at least the immediate future the poor teaching is likely to continue at some schools and [yet] candidates for academic support programmes have to be selected…" (p. 200).

4.3.19 The construction and evaluation of a dynamic, computerised adaptive test for the measurement of learning potential - Doctoral study conducted by M. De Beer (2000)

4.3.19.1 Introduction

De Beer's (2000) study is a pioneering study in South Africa, owing to its unique attempt to assess candidates in a multicultural set-up, using the latest computer software, as well as its fusion of Item Response Theory with Computer Adaptive Testing in a dynamic manner, utilising non-verbal figural items in the test. No prior scholastic knowledge is necessary, and even though it is administered on computers, no prior experience with computers is necessary. It also tests for a range of abilities.

The development of the Learning Potential Computerised Adaptive Test (LPCAT) is, as far as can be ascertained, the first of its kind in South Africa and will offer much to the field of dynamic assessment

in this country in the future. This doctoral study is a culmination of many years' work making use of large sample sizes for validation and other studies in three provinces in South Africa.

De Beer states that the need to devise a test that will adequately assess individuals' learning potential in as economical a manner as possible, adhering to the criteria put forth by the governing body in South African (South African Professional Board for Psychology) is urgently needed and although the LPCAT is a dynamic assessment technique it should render results in a quantitative fashion. Foremost among the reasons why a test such as this is sought is the need to assess multicultural groups with one test battery.
By combining learning potential measurement, the dynamic testing approach and computerised adaptive testing based on IRT, a psychometric instrument that

- Makes use of the improved IRT statistical procedures for test development
- Contains training as part of the assessment procedure to take diversity in educational backgrounds into account [and]
- Focuses on the measurement of learning potential was developed.

(De Beer, 2000, p.3)

The usual criticisms levelled at dynamic assessment, such as high cost factor and time intensive evaluation can be set aside when one considers the multiple benefits of the Computerised Adaptive Testing (CAT) approach of De Beer. Two problems often associated with dynamic assessment are the cost efficiency of administering the tests and the questionable accuracy of the results. The LPCAT addresses these issues directly. De Beer states that although Item Response Theory (IRT) is not a new theory, its application in psychometrics has been slow, possibly due to the later development of technology (such as programmes needed to run the necessary statistical procedures). The advent of CAT[119] is also uniquely suited for the administration of dynamic assessment techniques.

Highlighted in the study is the very old notion of potential, first brought to prominence by Alfred Binet,[120] (De Beer, 2000) although not necessarily conceptualised by him initially. De Beer acknowledges Binet's early attempts at trying to assess children's learning potential, his work later being used for the basis of IQ testing, an approach not initially intended by Binet. She states that his original intention for the use of his test has only today "come full circle", after a period of misuse of his test (whether inadvertent or not).

The LPCAT seeks to assess general reasoning or the notion of "g". De Beer summarizes the background of intelligence testing but focuses on the contribution made by Binet as the precursor to learning potential, in essence. Binet's ideas are seen to be ahead of his time, in that IRT and CAT are now readily available to fulfil his intention of the measurement of learning potential. De Beer also makes use of Vygotsky's zone of proximal development as fundamental to the theory underlying research in the field and notes the similarities between Binet's and Vygotsky's early attempts to aid low-performing children, devising measures in which potential can be identified, and points out that the way in which Vygotsky conceived of the measurement of intelligence was in fact quite similar to the way in which Binet's original concept of change could be seen in IQ test performance. "…a new approach to dynamic assessment is proposed that will allow the extension of Vygotsky's ZPD to the broader ability spectrum….[t]he extension of the ZPD concept to the broader ability level as well as the ZPD scores is included in the interpretation of the developmental (ability) level" (De Beer, 2000, p. 105).

De Beer emphasises the fact that both the initial level of performance and the difference in pre- and posttest scores need to be taken in account. Relying on difference scores only will not yield the true picture of a person's ability or developmental level (see also Boeyens, 1989a). De Beer states that

119 Embretson and Reise's (2000) study outlines recent detailed information on IRT and CAT as used in psychological assessment.

120 De Beer is one of few local researchers who acknowledges the importance of Binet's early contribution in the field of learning potential. It must also be noted that Thorndike defines intelligence as "the ability to learn" (Guthke in Hamers et al., 1993a).

both the **actual** level of performance **as well as the ZPD** are malleable and not fixed, and can be changed through training and instruction. The ZPD "…is represented by the difference between the potential level of development after training (posttest) and the initial (pre-test) level of development" (De Beer, 2000, p. 76).

She devotes a chapter to the history of intelligence and its measurement as well as a chapter on the development of dynamic assessment and the various learning potential schools or approaches. Here, she draws largely from the classification framework of Grigorenko and Sternberg[121] (1998) among others, as it is pertinent to the study, but will not receive attention here. However, in the delineation of dynamic approaches, De Beer systematises the various approaches and, where needed, indicates the level of similarity or difference between the mentioned approach and the focus of her study or approach. The following summary highlights these similarities and differences as presented by De Beer (2000):

- The enrichment approach (Feuerstein) - training is provided as part of the assessment procedure, the way in which it is provided reflects a clinical approach to assessment. The objective is to achieve "…enduring change in the cognitive functioning of the individual" (De Beer, 2000, p. 80)
- Budoff - where his approach makes use of standardised tests administered in a dynamic manner, the LPCAT makes use of a dynamic test administered in a dynamic manner, but the figural items used in many standard tests are used in the LPCAT. Both approaches make use of non-verbal tasks. Further similarities between the two are that the LPCAT serves to measure general abilities and these results are used as alternatives for conventional test results
- Campione and Brown (graduated prompts) - both measures make use of computerised administration in order to standardise the procedures, but once again the LPCAT is a dynamic test, whereas the Campione and Brown approach makes use of conventional tests. Secondly, the difference score is calculated using the pre- and posttest scores and is not determined in the manner used by Campione and Brown (where the difference score is equated to the number of hints or prompts required). Furthermore, hints to the candidates are administered when and where the need arises, whereas the LPCAT's administration is uniform and all candidates receive the same quantity and quality of instruction
- Carlson and Wiedl (testing-the-limits) - the LPCAT approach does not have much in common with this approach, with De Beer stating that owing to random assignment of members to groups no valid measure of pre-test performance is available (as no definite pre-test was administered) and also that because the Carlson and Wiedl approach allows no individual comparison to be made (the tests are group administered) the only real similarity is that the two are dynamic assessments in terms of training which is given in both
- Guthke (learning test) - perhaps the most similar aspect in this approach to the LPCAT is the test-train-retest approach with a distinct emphasis on the psychometric properties of the test as well as the use of IRT in the development of the tests. However, the LPCAT test focuses on the present level of performance in order to direct the next question to be asked of the testee, whereas with the learning test concept, other factors are used to determine which type (or level) of question should be asked, such as the amount of help needed and the nature of the help (prompts given)
- IRT approach - although IRT is a method used to analyse and choose items that are not biased, it has been accorded a status worthy of inclusion as an "approach" in the field of dynamic assessment by De Beer. IRT basically serves to equate the difference scores obtained (since the pre- and posttest scores are on the same scale), as opposed to classical test theory which "…leads to measurement problems regarding difference scores" (De Beer, 2000, p. 95). The powerful combination of IRT and CAT, according to De Beer, seeks to overcome many obstacles previously encountered in dynamic assessment measurement when worked with from a classical test theory point of view. She cites Embretson (1987) and Sijtsma (1993) as support for the move toward this approach. This study is based on "…standard IRT parameter estimation and CAT procedures" (p. 95) and includes the "latest trends" in differential item functioning (DIF) making use of two separate item banks, one for the pre-test and one for the posttest scores.

121 See Chapter Two section 2.3.1 for more information on the comprehensive article by Grigorenko and Sternberg (1998).

Reliability and validity issues, as mentioned, plague the field of dynamic assessment, but these issues are solved when using IRT and CAT as part of the testing procedure. It is interesting to note that many advocates of dynamic assessment try to steer away from classical psychometric means of measurement. A number of researchers do, however, acknowledge the usefulness of maintaining both. De Beer states that her instrument "…belongs to the psychometric approach to dynamic assessment, with the focus on **accurate measurement**" [own emphasis] (p. 96).

De Beer pays attention to the fact that group test scores (IQ) have changed over the years, and contributing factors such as socio-economic level have had a major influence in this trend; and she cites a number of studies conducted in South Africa to illustrate the narrowing of averages between different cultural groups in the country. Language proficiency[122] is cited as an important factor in tests with a strong verbal component. The past situation in South African psychological testing leaves a lot to be desired as most of the tests used in this century were international tests which were standardised for local conditions (even if this only really meant that items were translated). According to De Beer, early attempts at tests which aimed at considering learning potential in South Africa (or adaptability) were the General Adaptability Battery (GAB) developed by Biesheuvel in the 1940s and 1950s. In his battery, Biesheuvel takes cognizance of the influence of cultural and educational variables in test scores. In accordance with De Beer, the present research details that the next "wave of interest" occurred in the late 1980s at the University of Natal (TTT[123] programme). The thesis written by Boeyens (1989)[124] followed two Human Sciences Research Council reports (using the same data set) that were also published in 1989 in South Africa. The research by Shochet in 1986 was, of course, also a precursor to the renewed interest in dynamic assessment.

This signals quite a large gap in research in dynamic assessment in South Africa, a country which sorely needs this type of research. This gap is further highlighted by the international research conducted into dynamic assessment starting in the 1950s and gaining momentum in the 1960s. De Beer states that local psychometrics followed international trends with little pioneering work in the field of locally developed tests. The work of Taylor, which utilises two test batteries, the APIL and TRAM in a dynamic manner, is used in industry as well as at tertiary institutions. However, the present research concurs with what De Beer states (2000) namely that "[u]nfortunately, although these latter instruments are used in industry, research publications on their results could not be found" (p. 5). The thesis by Van Aswegen (1997) does, however, detail the findings of the TRAM-1 test battery as developed by Taylor as discussed earlier in this chapter so too the research by Nel (1997), Watson and Ncapayi (1998), Engelbrecht (1999) as well as Lopes, Roodt and Mauer (2001) who make use of the APIL and the research into the locally developed Learning Ability Battery (LAB) as studied by Tayler (1996) discussed.

CAT and IRT

De Beer discusses the shortcomings of local research and pays attention to the language barrier issue, educational development (or lack thereof), current political situation, as well as test-specific issues such as verbal and non-verbal (figural) aspects of psychometric tests used in South Africa. One of the most important topics in the field of multicultural assessment is the degree of bias present in a test. De Beer acknowledges this and devotes a lot of time to the issue of bias analysis also known psychometrically as DIF (returned to later). Through the use of CAT and IRT theory as opposed to classical test theory, many disadvantages of learning potential assessment can be overcome. The broad-based aims of the LPCAT as set out by De Beer include the need to:

---

122 In a recent report of language proficiency at a local South African university it is revealed that "…poor language skills are contributing to the under achievement of students at university…[and] at least a third of all [these] students require remedial language teaching" (http://www.news24.co.za/News24/South_Africa/0,1113,2-7_1053215,00.html). It was evidenced that poor language skills, or skills inability "… [were] suppressing their achievements" (*ibid*.).

123 De Beer (2000) incorrectly refers to the Teach-Test-Teach approach as the "**test**-teach-**test**" approach, which although minor in print, is quite contentious if the TTT programme's aims are studied in detail.

124 The thesis written by Boeyens (1989) is an amalgamation of his two reports (1989a) and (1989b).

- Assess the impact of initial training and its beneficial role
- Assess the extent to which learning potential tests narrow the difference in results between the various cultural groups
- Assess the equability of the distribution of scores on learning potential tests
- Assess the predictive validity of learning potential tests as opposed to conventional tests. In other words, the question of whether learning potential tests are able to indicate future educational performance is posed.

As this is a validation study, an in-depth look at and an understanding of psychometric theory, as well as the statistics used is necessary. Specific aims of the study include the following:

- Evaluation of DIF between language, gender and cultural groups
- Evaluation of the reliability and validity of the battery according to APA test development standards
- Assessing predictive validity by measuring other relevant results
- Assessing construct validity by comparing the results with standard psychometric test batteries
- Assessing the usefulness of the battery in a multicultural set-up.

(De Beer, 2000)

The study illustrates the plethora of variables which have an impact on low scores usually obtained with previously disadvantaged cultural groupings and attempts to disentangle the role played by each variable. In so doing, De Beer analyses results from three South African surveys (the October 1995 Household Survey, the 1996 Census results, as well as the 1999 Reality Check Survey) which support the notion of unfair distribution of wealth and educational opportunities. These are environmental influences which play a large role in standardised intelligence tests. The standardisation sample used in the study consisted of N = 2 454, with an almost equal gender split, but with an unusual racial split - 49% African, 27% white and 24% coloured.

It is necessary to point out the various advantages in using IRT with CAT in the dynamic LPCAT:

- IRT entails the use of two separate item banks which means that the items are not the same and the practice effect, so often a problem in research designs, can be ignored. (However, the research conducted by De Villiers [1999], Gaydon [1988], Henley [1988] and Hoffenberg [1988] overcame this problem quite effectively with the use of a Solomon Four Group Design.)
- The CAT enables the testee to complete items that are commensurate with their present level of functioning and in so doing is able to allow assessment of an individual quite speedily. Since the items are adapted, neither ceiling nor floor effects are likely (which is often the case with people who initially score high and are unable to score much higher on the posttest)
- Each item is measured for its reliability and offers a more refined assessment of individual performance as opposed to reliance on the number of items that are correct in classical test theory
- Lastly, the latent trait theory of intelligence is important in De Beer's study. Anastasi states that "[a] fundamental feature of this approach is that item performance is related to the estimated amount of the respondent's 'latent trait', symbolised by the Greek letter theta…" (Anastasi, 1988, p.221). Furthermore, Anastasi states that theta is a statistical construct and not a psychological construct. Changes in scores between the pre- and posttest represent a change in performance due to training and not due to the difficulty of the test items.

De Beer devotes a chapter to the reasons behind the use of IRT and CAT, as well as the historic development of the theories and duly explains the advantages and disadvantages surrounding their use in dynamic assessment. The principles underlying IRT and CAT are also discussed at length. The applicability to the study is obvious but will not receive detailed attention here. However, it is prudent at this point briefly to view the discussion on IRT and CAT.

IRT and CAT as discussed in the study

IRT and CAT are discussed in detail by De Beer and once again the original intentions of Binet are highlighted, as these intentions are very close to the manner in which IRT functions in the LPCAT. The disadvantages of classical test theory are also discussed with the need to focus on a new theory that is able to deliver more accurate information regarding people's learning potential. One of the more serious disadvantages of classical test theory is the varying nature of the measure of item

difficulty, which differs when used with different groups. Also true scores cannot be directly measured but must be inferred from calculations using the error score and observed score.

Reliability is thus dependent on the ability level of the respondent as well as the level of difficulty of the test. This leads, among other things, to results that are incomparable with other testees if comparisons are not made within the same specific test. This is so because test results are dependent on item difficulty as well as testee ability. It is also very difficult to determine how a testee will fare on a test item. Of note is that the tests administered in the LPCAT are not speed[125] tests, which differs from the battery of Taylor (TRAM-1), in which speed is a crucial aspect of the test procedure (cited in Van Aswegen, 1997).

IRT Model used

Of the three IRT models available, De Beer chose to use the three-parameter model, as the required sample size was available. The three-parameter model allows three parameters to vary; these parameters being the difficulty level of the item (b-value), the item-discrimination (a-value) and the pseudo-chance parameter (c-value) - the extent to which items are answered correctly by guessing. In the three parameter model, all three values are allowed to vary (with the one- and two-parameter models, one or two variables are at any one time constant). De Beer states that in order to make use of the three parameter model large sample sizes are needed (N > 1000), "…with sufficient numbers of low ability examinees to allow accurate estimation of the pseudo-chance…index" (De Beer, 2000, p. 118).

Another reason why this model was employed was because the format of the test items was multiple choice and the chances for guessing items correctly increase in this format, and the value of c is evident. In order to encompass this variance of c (the third parameter) the third model was chosen. De Beer also used a large enough proportion of low ability respondents:

> [o]ne of the advantages of IRT is that is resolves the problem of sample dependency by providing ability parameters which are invariant over samples of subjects, and not dependent on the particular items that are administered. This allows for more accurate comparison of different test scores of the same individual as well as comparison of test scores between individuals…. (De Beer, 2000, p.113).

De Beer made use of the MicroCAT system (for item analysis) in order to run the three-parameter item analyses program. Certain assumptions were made regarding the use of the three-parameter model, such as unidimensionality, invariance of item parameter estimates as well as the invariance of ability parameter estimates. These constructs were empirically measured by De Beer and will receive attention later. The item characteristic curve (ICC), item parameters and the test information function are discussed at length.

The use of CAT - Computer Adaptive Testing

The main advantage of CAT is that the test is adapted for each individual which makes for more accurate measurement as well as enhancing testee response. The testee does not struggle unnecessarily with difficult items but also does not become bored with items that are too easy. The amount of latent trait or theta is continuously adjusted either up or down depending on the answer given to each item. The estimated theta value is found in this manner, and this technique thus uses far fewer items than classical test theory, making it more efficient and less time consuming, a major inhibitory factor in dynamic assessment, as already noted.

125 See also Taylor (1999) who states that speed in the API is very important, section 4.4.

4.3.19.2 Research design

Construction of the LPCAT

De Beer outlines the process of test construction and in particular the construction of the LPCAT. The need for such a test battery has been mentioned. Dynamic measures or tests are often not replicable. However, the LPCAT serves to correct this issue among other aspects already mentioned (such as time saving). Features which are of importance in the LPCAT, other than those already mentioned, include the test-train-retest method of administration, incorporation of standardised training and the use of multicultural groups for the item analysis, standardisation and validation of the test battery. Three types of non-verbal items were used in the test, a pool of 270 items, divided into 90 items for each type. The types consisted of figure series, figure analogies as well as pattern completion. The range of items, (which were initially administered using paper and pencil), covered the whole spectrum of abilities, and items were scrutinised by professionals in the field of cognition. Owing to the large number of items necessitated, two administrations were given, after which item analyses were conducted. Of note is that during both administrations, 66 anchor items were used so as to have some items common to both groups in order to calculate item parameters.

Forty-one schools were chosen to participate from three provinces. The schools were identified on a random basis and chosen for the willingness of school psychologists and teachers to co-operate in the research. Sixty pupils from each school were chosen, 30 each from Grade 9 and 11 and gender was split exactly. In the administration of the paper and pencil tests (both formats were alternated to ensure equal distribution of items), all cultural groups were included except the Indian group. De Beer points out that this group does not make up a large percentage of the South African population, but that they usually score relatively closely to scores obtained by whites on tests (as seen with previous research in psychometrics), and also that the provinces chosen for the study did not have a large contingency of Indians. Although not exact representational proportions of the cultural groups in South Africa, therefore, the demographic spread allowed for item analysis using the three parameter IRT.

De Beer analysed the items using both classical test theory and item response theory. The software used to conduct the item analysis from the **classical** perspective was the **ITEMAN**[126] program of MicroCAT and the two item pools (both forms) were analysed separately and hence two sets of values were obtained for the anchor items. Alpha coefficients were also calculated for both forms of the test and yielded values above 0.925, indicating high internal consistency. This supports the unidimensionality of the LPCAT test battery.

IRT analysis was conducted on all 270 items (they were pooled together using the anchor items, for both forms, and as the software allowed for the coding of items which were "reached" and not "reached", the programme could complete the analysis. The programme used for this purpose was the **ASCAL**[127] programme of MicroCAT). In other words, those items that were not answered by each group were simply coded as such; this dichotomous scoring process allowed for the procedure to continue. Five items were discarded in the process and the items chosen were based on classical test theory as well as IRT analyses, with emphasis on IRT.

As mentioned previously (under the IRT model used), the general assumptions of IRT are unidimensionality, item parameter invariance and ability parameter invariance. For **one-dimensionality**, factor analysis was conducted to determine whether the items measured the same construct for the whole group and subgroups. The construct purported to be measured in the LPCAT is non-verbal, figural reasoning. Principal component factor analysis was conducted for both forms

126 This programme "…scores items that are not reached as incorrect, and this affects the values obtained" (De Beer, 2000, p. 149).

127 "The ASCAL programme…was used to calculate the IRT item parameters…[which] estimates IRT item parameters according to the two- and three-parameter IRT models (De Beer, 2000, p.152).

and yielded "…support for a one-dimensional structure for both the total group and the various subgroups" (p. 154). Scree plots (for all subgroups as well as the whole group) illustrated the cut-off points used, a point where the factors no longer account for too much explained variance, and hence the number of factors are contained. Of note is that factor 1 accounted for more or less the same variance for the African and Coloured groups, but accounted for less variance for the White group. The one-dimensionality of the LPCAT is supported and is also significant because the training that is administered is simplified.

For **item parameter invariance**, De Beer quotes Lord (1980), one of the founding figures involved in IRT, as stating that item parameter invariance across groups is of great importance. In order to assess whether the item parameters are indeed invariant across independent groups, De Beer chose to divide groups according to gender (male vs. female) and language (English/Afrikaans vs. African languages). Scatter diagrams were calculated for all three parameters for both these groups. The linearity of the results preclude much invariance. The correlations were all highly significant indicating that the item parameters were indeed invariant. The b-parameter plots were more linear, and can be considered as favourable, since the b-parameter indicates item difficulty. The item parameter invariance can thus be confirmed for the LPCAT items.

For **ability parameter invariance**, De Beer states that the ability parameters of a person are not affected by the items used to assess ability. The ability parameters should also not differ from one test to the next. An empirical way of investigating this aspect was to calculate the ability parameters of a group of testees using three different sets of item types (namely figure series, figure analogies and pattern completion). This was carried out to determine the ability parameters of the total group and to see whether the results concurred. All three scattergrams support the notion of ability parameter invariance and De Beer states that these findings are similar to those made in other studies. The support garnered for the three parameters allows De Beer to state unequivocally that the three parameter IRT model is the most suitable one to use for her study.

DIF - Differential Item Functioning

Differential item functioning is the detection of the measure of bias in a test. In other words, it detects those items which may unfairly discriminate against a group or groups. Differential item functioning is similar to the classical test theory measure of test bias. De Beer states that DIF is still an evolving technique not without its problems. DIF is not only able to differentiate over particular ability levels but across different ability levels. DIF investigates the item characteristic curves (ICC's) of different groups. ICC's are drawn on the same graph after having calculated their respective theta values (ability levels) and then compared. The central issue here is the probability of a group scoring an item correctly when both groups' abilities are commensurate. If one group's scores are lower than another (and both have the same theta values), then the item can be said to be biased. De Beer calculated these DIF statistics and subsequently had to delete a number of items from the item bank. She investigated DIF for language, culture, gender and grade. Subgroups had to be larger than N = 1000 in order to make meaningful interpretations. All but the white group (N = 658) consisted of the correct sample size. Of interest was determining when and which items would be scrapped from the item bank which was dependent on "…visual inspection and empirical estimation" (De Beer, p. 168). An item was considered to be biased if the difference between the ICC's was greater than 0.5. Thirty five items were discarded in this manner.

De Beer states further that not only are the items important when considering inclusion into the final item bank, but that all items would have to adhere to three more values, such as IRT c-values and a-values and classical test theory (CTT) values. Altogether 47 items were discarded based on these three parameters along with the 35 items discarded due to the DIF results, which is equal to 30 percent of the total item back, and according to De Beer is the norm (a third of the items being discarded). One hundred and eighty eight items were left, and were divided into the pre- and posttest halves of the test, to a ratio of 1:2. Each of the three item types were similarly divided (figure series, analogies and completion). De Beer followed the advice of the literature which suggested that the items in the bank should exceed the number of items any one respondent was to answer by 10:1. This meant that the ratio of pre-test items was in the order of a ratio between 5 and 8 and for the

posttest a ratio in the order of 7 and 10. This translates into 8-12 questions in the pre-test and 12-18 questions in the posttest.

De Beer details the values of the descriptive statistics for the IRT item parameters of the pre- and posttest scores (i.e. a values, b values and c values). The b values indicated that the pre-test items are pitched at the intended Grade 10 level. The items do, however, cover a large spectrum of ability levels even though the above-mentioned is the intended target group (previously disadvantaged groups). De Beer next discusses and gives evidence on the reliability of the LPCAT. The test information function (TIF) is a measure of the reliability of the test. Two advantages of IRT test information are that one can calculate the standard error of measurement at various levels of ability (but consider that this changes at each level and thus has to be calculated for each ability level) and also that the TIF is related to the measurement of the effectiveness of the test itself. "The [TIF] graphically indicates the amount of information at various ability levels, when specific items are included in a test" (De Beer, p. 174). TIF indicates the amount of information that the test yields at each level of ability. It is obvious after some consideration that the classical test theory methods of determining test reliability (such as test-retest, parallel forms of tests and split-half method) are not suitable for CAT since the test itself interactively selects items. De Beer also indicates the rather simple method of determining the standard error of measurement.[128] De Beer made use of the microCAT test evaluation programme EVALUATE to obtain the TIF's as estimates of the reliability of the pre- and posttest LPCAT scores. The graph (De Beer, pp. 177-178) illustrates that most of the TIF is yielded for the theta value of 0.0 (Grade 10 level) and yields less information for lower and higher theta values for the pre-test. Similar values are obtained for the posttest (only slightly lower than the pre-test values). This translates into the need to make use of more items for those candidates at either end of the ability scale.

De Beer's next step was to computerise the items and in so doing make use of the MicroCAT Testing System.[129] Feedback on the items are indicated by a "tick" for a correct answer and a "cross" for an incorrect answer. The LPCAT -1 is an independent form of the test, presupposing the candidate's level of reading is that of at least Grade 6. The training session includes explanations of how to go about finding out the patterns involved in the various items. Of importance to note is De Beer's use of standardised training[130] which is not adaptive (as with the rest of the test) and this is done primarily to make comparisons directly, as well as to allow everyone the same level of entry (regarding familiarisation). De Beer states that this aids in the face validity[131] of the test.

Because of the nature of the LPCAT, the pre- and posttest results are directly comparable, the accuracy of measurement is used as cut-off criterion and each individual is administered different items. The LPCAT-2 makes use of exactly the same items, the only difference being that it does not include text, rather, the directions are read out to respondents in their home language. "The only difference between the two versions is the initial entry level and the method of working through the practice examples and training section …either reading it independently…or being provided with the verbal instructions…" (De Beer, 2000, p. 288). Entry levels are set at 0.0 for the LPCAT-1 and -1.0 for the LPCAT-2, yet candidates with any ability are able to complete either of the tests. The software used for the LPCAT administration takes the following into consideration when the test is given to candidates:

---

128 The standard error of measurement is calculated as follows: the reciprocal of the square root of the TIF is taken and is expressed as the range between which most ability levels will fall. Thus one can choose the ability levels one wants to include in the test by selecting items from the bank which will yield a TIF at the desired level. This will result in the range that one wants included.

129 Only two keys are used in the LPCAT, namely the ENTER and SPACEBAR keys. This test presupposes no computer literacy on the part of candidates.

130 The idea of standardised training is not always seen by other researchers as true dynamic assessment, since by its very nature, dynamic assessment includes training that is specific and individually assessed and administered.

131 Zolezzi (1995) states that face validity is often ignored in dynamic assessment tests and that "[a]ppropriate tests must **appear** to measure many of the prerequisite skills…." [own emphasis] (p. 11).

- A counter number, which keeps track of the number of items answered
- The unique item number of the item administered
- The answer chosen by the candidate
- The correct answer
- The estimated ability level on the theta scale
- The accuracy of ability estimation
- Time that has elapsed since the start of the test (note that De Beer states that the LPCAT is a power test but is not limited to any specific time).

It stands to reason that in order for the software to be able to compute the above-mentioned test statistics, this information has to be put into the system, such as a list of the items, the variance of ability estimation to be used as termination criterion, as well as the maximum and minimum number of items to be administered.

The LPCAT scores are transformed into T-scores (and are directly transferable into percentile scores). The pre-test score indicates the current or actual developmental level; the posttest scores yield the potential level of performance. A composite score takes into account the ZPD (as used in Vygotsky's terminology), the pre-test score as well as the posttest score. Pre-test scores are looked at when allocating candidates to expensive training courses for example, which necessitates a certain ability level. The ZPD (difference score) is looked at when identifying candidates for affirmative action positions (those who are most likely to improve). Posttest scores are looked at when identifying candidates who (it is assumed by the fact that all have received standardised training) have all had the same opportunity of developing skills and to choose those who have performed well. De Beer advises that although the individual scores can be assessed in this manner, it is more feasible to investigate the overall performance score, as this reflects **all** the above-mentioned scores. A crucial line of reasoning takes precedence at this moment and that is the distinction between those candidates who evidence a large ZPD and those who evidence a small ZPD.

Of prime importance is the initial level of functioning, since it is clear that someone with an initially low pre-test score and who scores a high posttest score cannot be equated with someone's initially high pre-test score and only marginally better posttest score. It stands to reason that an increase in ZPD for a high pre-test scorer and higher ZPD increase for a low pre-test scorer are not measurably the same at all. This is where the idea of the composite score is highlighted: "[t]he idea underlying the composite score is that the amount of credit given for the size of an examinee's ZPD (or LPCAT difference score) should be **adjusted** on the basis of the maximum ZPD that he or she could **theoretically** have attained from his or her pre-test level of performance" [own emphasis] (De Beer, 2000, p. 188).[132] De Beer mentions that initial levels of performance are an indication of ability to learn and that the ZPD cannot be seen to be the same (or count for the same) for different candidates. De Beer states that the residual score used by other researchers (Budoff [1969] as cited in De Beer, 2000, for instance) is similar to the composite score as calculated here. The main difference, as De Beer points out, is that the **residualised** score starts out from the posttest score and makes a **downward** adjustment to the pre-test score, whereas the **composite** score is a more conservative estimation as it starts with the pre-test scores and works **upwards** to the posttest score. De Beer firmly states that the manner in which the composite score is calculated is more indicative of the interpretation of Vygotsky's ZPD as opposed to the residualised score calculation. Vygotsky's ZPD is described by himself as "…the distance between the actual developmental level as determined by independent problem solving [pre-test] and the level of potential development [posttest] as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). Clearly one can see that the direction of thinking is from the pre- to the posttest evaluation and not the other way round.[133]

132 The composite score (C) is worked out as follows: C = I + D2 /(3-1) or on the T-scale C = I + D2/(80-1). The difference score is expressed as a proportion of this maximum difference. Credit is given for the proportion of the difference score and one adds this proportional credit to the initial score to give the composite score.

133 De Beer's approach to working out the composite score is much more detailed than the working out of difference scores evidenced in other studies concerned with in this survey, in which many cases deal with the difference score purely as a posttest score minus a pre-test score.

By using the CAT method with IRT, and composite scores, one is able to compare candidates directly with different initial ability levels and with differing ZPD's. De Beer emphasises the point made by Grigorenko and Sternberg (1998) that not enough is written about the reliability and validity of the different dynamic assessment tools available. The aim of her study was to include this pertinent information which is so often neglected.

Next De Beer explains the procedures involved in evaluating the validity of the LPCAT, and proceeds to examine the results of the validity studies. This is briefly examined below.

4.3.19.3 Main findings

Procedure for the validation studies for the LPCAT

De Beer emphasises the validation of the LPCAT, by comparing its predictive properties and/or correlational properties with other cognitive tests, and so she does not really focus on the actual scores obtained during the pre- and posttest phases. Thus far the process of validation of the items in the LPCAT has received attention. What follows is an overview of the procedure followed for the **validation** of the LPCAT in various contexts (gender groups, educational groups and language groups). The computerised form is, of course, also important as opposed to the paper-and-pencil test format used for the validation of the items. De Beer duly discusses the theory behind the need for validation and the primary role accorded certain aspects of the validation procedure. Content and face validity are highlighted by De Beer in this section of the research.

For the validation studies, De Beer made use of three samples for LPCAT-1, two samples for the LPCAT-2 and specific groups, and selections of groups for further validation studies on the LPCAT. In order to determine the criterion validity of the LPCAT, various cognitive tests were used as criteria. These are mentioned below. Face validity of the LPCAT was determined by the nature of the items used in the battery, namely universal figural items with no text. No language is used and reliance on prior learning is of no consequence. There is no time limit on the test, it is available in all eleven official languages and makes use of only two keys on the computer keyboard, making it accessible to most people. The content of the test was said to have been investigated by a panel of experts within the field. Construct validity is assured owing to the one-dimensional "factor structure" yielded by the factor analysis (De Beer, 2000). Items were homogeneous, yielding high internal consistency results which is commensurate with the factor analysis.

Validation of LPCAT-1

For all samples tested, the results consisted of pre-test scores, training, posttest scores and composite scores. Criterion scores were collected for all groups used in the validation process and were also different for each group. These strategies afford greater variety of criteria tests and so provide a better base for comparison.

Three samples used included 92 first year Technicon students in the faculties of science and engineering, 223 first year Technicon students (from a different Technicon), also from the science and engineering faculties, and the third group consisted of 37 Grade 9 high school pupils from an urban high school.

1. The first sample group is not statistically significant due to the non-random allocation of voluntary students to be tested. Criterion measures included the GSAT, matriculation examination results in English, Mathematics and Science and end of first year Technicon results. There were considerably more males than females in this group but an even spread of African and English/Afrikaans speaking students
2. The second group is considered by De Beer to be representative of all first year science and engineering students at this particular Technicon. The gender distribution was more even for this group as opposed to the first group. Language spread was also even. Criterion measures for this

group included the GSAT-CAT[134] (a computerised version of the GSAT[135]), the SAT (un-normed but revised subscales of the test), matriculation examination results in English, Mathematics and Science as well as end of first year Technicon results[136]

3. The third group's sample was randomly selected from the total Grade 9 class from an urban high school. Although small in number (N = 37), language and gender distribution was even. Criterion measures for this group included two results from the school year (an average of four terms' results as well as a year mark).

Validation of LPCAT-2

The validation procedure for the LPCAT-2 followed in much the same manner as described for the LPCAT-1. For the sake of clarity, the groups follow on in number sequence (as carried out by De Beer). Two groups were used for the LPCAT-2 (which was the version of the test which did not include any text). The two groups consisted of a fourth group with 194 adult learners, mostly low-literacy adults, and the fifth group consisted of 144 Grade 8 pupils.

4. Owing to the homogeneity of the fourth group, who were mostly male and spoke predominantly African languages, and who had received on average eight years of education, the results were reported for the total group. Criteria scores for this group included results from the PPG test (Paper-and-Pencil-Games) which is a test measuring figural, verbal and quantitative aspects and is, according to De Beer (as also previously stated by Claassen, 1996 [in De Beer, 2000]), related to scholastic achievement up to the fourth year of school. Level 1 literacy and numeracy scores were also obtained for this group
5. The mean age of the testees for the fifth group was 13 and included almost all the Grade 8 pupils of the class. The sample was slightly skewed towards English/Afrikaans speaking students and there were slightly more females than males. Criterion scores for this group included the GSAT-CAT, the Learning Process Questionnaire (LPQ). Other criterion measures include the Proficiency Test English Second Language (Intermediate level) which according to De Beer is intended to measure English second language proficiency between the 7th and 9th grades, a Mathematics proficiency score obtained via the Basic Numeracy Literacy, which is intended to cover basic knowledge of numbers and associated arithmetical concepts, two teacher-ratings, one for English and one for Mathematics and lastly, the average of four terms' results as well as a year average
6. De Beer's sixth group is a group made up of 109 Grade 9 pupils from an urban high school. The gender and language distribution was even and the reason behind the inclusion of this group "…was to investigate specifically the effect of various types of training on LPCAT-1 results" (De Beer, 2000, p. 210). For this group, three sub-groups were formed after random assignment to these groups. The first sub-group (N = 37) was administered the LPCAT in its standard form; the second sub-group (N = 35) received additional training (18 additional items were administered during the training session) and the third sub-group (N = 37) was administered the LPCAT without any training between the pre- and posttest phases. Criterion measures for these three groups consisted of an average of four term results as well as a year mark
7. Lastly, all six groups were combined in order to investigate the difference scores for all groups more closely as well as to examine the "developmental" changes that may be indicated by the LPCAT. This would result in more validity information.

De Beer captured the data in ASCII format which was later incorporated into the SPSS data package software for further analyses.

134"The GSAT computerized adaptive test can be used successfully for the general evaluation of the individual's ability level, selection with regard to placement, [and] predicting future achievement" (van Tonder, 1994) in http://hagar.up.ac.za/catts/learner/andres/assess.htm.

135 See also Lipson (1992) section 4.3.9 and Nel (1997) section 4.3.14 who also made use of the GSAT.

136 Van Eeden, De Beer and Coetzee (2001) also discuss these results separately in a recent article. The article also details findings of personality factors (16 PF) along with cognitive ability factors as further predictor variables.

Results of the validation studies

Before setting forth the results of the various validity studies, De Beer provides a cautionary note about the small sample sizes used and hence the generalisability of the results. De Beer refers to the pre-test, posttest and composite score as the LPCAT (PPC) which makes for easier reading and is used from this point on.

The spread of the samples (language and gender and race group) makes for easy cross-validation studies and the fact that criterion measures are available for most of the groups, allows access to the criterion validity of the LPCAT. Furthermore, De Beer states that the exercise involves not only criterion measures in general but also concurrent validity and predictive validity results. As the criterion measures were assessed at the same time as the LPCAT, one could investigate the LPCAT's concurrent validity as well as predictive validity because the criterion results obtained later on in the study such as year end marks.

De Beer pre-empts the debate that may arise as to the suitability of the LPCAT to serve as a predictive test, on the grounds that it measures figural and non-verbal aspects whereas academic and other cognitive tests measure static or conventional prior learning, presupposing verbal items in the battery. She states that:

> [t]here is no general answer [as to] how high a validity coefficient should be, and the less overlapping there is in content between the test and the criterion, the lower the expected validity coefficient will be…the latter is important to keep in mind when interpreting the LPCAT results, where performance on a test of nonverbal, figural content is compared to different measure of academic performance (De Beer, 2000, pp. 217-218).

As will be shown later, the LPCAT does indeed correlate significantly for some groups on the criterion variable. De Beer firmly states that "[t]he primary aim of the LPCAT is **not** to predict academic performance but to assess learning potential in the general reasoning ability domain" (p. 218).

Validation results of the LPCAT-1 (Groups 1-4)

Group 1

Non-directional t-tests for independent samples were conducted on all samples in the descriptive results. Regarding the language differences between the two groups, all but three measures were statistically significant. The LPCAT difference score was not significant for this group and the possible reason attributed to this is that there is no difference in how either of the two language groups may improve. All three GSAT scores were statistically significant for both groups, with Mathematics 1 and Grade 12 English not being significant. Regarding the gender differences, most of the results were statistically significant. The LPCAT difference score, GSAT non-verbal test, Mathematics 1 and all three Grade 12 tests were not significant. The males performed better on all the significant tests. The gender groups thus did differ significantly from one another.

Graphically, the distribution of the English/Afrikaans group was more negatively skewed as opposed to the African language group for both the verbal and non-verbal GSAT scores. However, the LPCAT pre-, post- and composite scores as well as the average first year academic scores were more similar for both groups and De Beer concludes on the basis of these results that "…the LPCAT seems to provide a somewhat more equitable distribution of scores for the African language subgroup of Group 1 than do the GSAT standard cognitive measures" (De Beer, 2000, p. 228). Also, not only are the scores more equitable across groups but also within groups. In other words, the LPCAT is able to differentiate between different ability groups in both language groups. The LPCAT posttest score as well as the composite score for Group 1 correlated with the GSAT more so than the pre-test and this supports De Beer's contention that construct validity of the LPCAT has been affirmed and that the two tests measure similar constructs. The low level of correlation between pre-test and GSAT as opposed to the posttest and composite scores attests to the "…dynamic measurement of learning potential" (*Ibid.*, p 229). The correlations between the LPCAT and certain criterion measures for this group (first year Mathematics and first year average) are not encouraging. However, there are significant

correlations between the LPCAT posttest and composite scores, once again more so than for the pre-test, lending support for "…use of learning potential measures that include measure of present (pre-test ability) as well as the results following training (difference score)" (*Ibid.*, p. 230). The posttest and composite scores correlate significantly with Grade 12 Mathematics, Science and English. The reason attributed to the low correlations between the first year scores and the LPCAT is the possible restriction of range[137] of scores. This group has comparable entry level ability, thus making them a "pre-selected" group in terms of range of ability (all having a Grade 12 ability level).

The scatter diagram indicated this restricted range. Of note is that the Grade 12 criterion measures correlate significantly with the first year academic performance, over and above those results pertaining to the LPAT and GSAT. This it true for both language groups. One would have assumed that the LPCAT may have predicted better for the African language group (owing to the previously disadvantaged status of this group). This is not the case.[138]

Of bearing to selection results are De Beer's findings regarding the high correlation between Grade 12 English and first year average performance which is highly significant but not so for the English/Afrikaans group, which underlines the importance of language proficiency when selecting for disadvantaged students. Due to the restriction of range of ability for this group, the magnitude of correlations seems to have been likewise affected (De Beer, 2000). No regression analyses were conducted for this group, because the correlations were not significant. Based on the overall results for this group, De Beer suggests that as the LPCAT does not correlate as well as other cognitive tests on academic results, it should rather be used as additional information and not as an exclusive instrument for academic prediction.

Group 2

See the above-mentioned information pertaining to this group. The two language groups differ significantly on both the cognitive test measures, GSAT-CAT as well as the SAT. However, there is only one significant difference between this group on the LPCAT, namely the posttest score. There are also significant differences on the Grade 12 Mathematics and Science results. This emphasises the fact that the LPCAT does not discriminate as much as the conventional cognitive measures, for instance. Interestingly enough, for this group, the gender groups differed significantly on a number of the criterion scores, more often in favour of the female group scoring higher than the male group. However, no significant differences were recorded for any of the LPCAT results.

Distribution of the GSAT verbal scores between the language groups were similar but still slightly positively skewed for the African language group. The GSAT non-verbal scores were very similar for both groups as were the average first year scores, LPCAT pre- and posttest and composite scores. De Beer concludes that "[t]he LPCAT score distributions are reasonably similar for the two language groups and therefore seem to provide **equitable learning potential measures** for the two language groups in the domain of general reasoning" [own emphasis] (De Beer, 2000, p. 244). Construct validity seems to be supported according to the data.

Correlations between the LPCAT and GSAT suggest that the two tests measure the same construct[139] as well as supporting the concurrent validity of the LPCAT (since as the two tests were administered during the same period). Criterion validity is supported when investigating the correlations between the LPCAT and end-of-year results for Mathematics. LPCAT pre-test and composite scores correlate

137 See Van Aswegen (1997) section 4.3.13.3 in which Huysamen mentions the relationship between restriction of range and predictor/criterion variables.

138 An interesting finding in this data is that the GSAT correlated higher with the African language group as opposed to the English/Afrikaans group for the first year average mark, although this was not significant.

139 Engelbrecht (1999) also found that the GSAT and APIL measured similar constructs for specific groups as did Nel (1997). De Beer (2000) once again makes use of the GSAT and finds that the LPCAT and GSAT-CAT measure similar constructs. It would appear that the GSAT is quite a popular test to use in these learning potential validation studies.

significantly with end of year average results. As with group 1, Grade 12 English correlated significantly with first year average for the African language group but not so for the English/Afrikaans group which underscores the importance of language proficiency, and the effects if instruction are not given in the mother tongue of the student. Regression analyses were carried out in order to determine whether the regression lines for the different groups were similar or not, as well as to investigate the potential over- or under- prediction of some groups compared to others. If the total group regression line is used to predict average first year performance, then females will tend to be underestimated and vice versa for the males. The African language group will also tend to be underestimated and vice versa for the English/Afrikaans group. The regression line for the African language group is quite a flat slope which indicates that the large disparities in scores for this group "…do not result in distinctly different predicted average academic scores" (De Beer, 2000, p. 250).

De Beer states that the pattern of correlations differs from the previous group and since these two groups are both students from Technicons (in their first year) it makes it somewhat difficult to generalise any findings for this group. Three possible reasons (cited by De Beer) as to why the LPCAT composite scores were generally insignificant when correlated with first year academic performance were possible restriction of range on the LPCAT and criterion scores, the fact that averages were calculated differently for different students and the different subjects taken as well as the effect of language (which had been mentioned a number of times thus far).

Group 3

The last group to be investigated in terms of the validity of the LPCAT -1 is the third group of 37 Grade 9 pupils. Apart from the small sample size, this group was beneficial for a number of reasons, namely that the ability spectrum was larger (as is the case with high school pupils, as they are not as homogeneous as say university students), and the fact that all students take the same subjects makes for more direct comparisons between pupils. The make-up of the sample also allowed for multicultural comparisons. There were no differences between the language groups on the LPCAT test scores, and only one significant difference on one of the criterion scores. The same finding is evident for the gender groups, with only one criterion score being significantly different. Distribution diagrams once again indicate that the graph for the African language group is more positively skewed than the English/Afrikaans group when comparisons are made for the Grade 9 average year mark. However, the scenario is slightly different when frequency diagrams are illustrated for the LPCAT pre-, posttest and composite scores. Here, the two language groups are very similar in scores with only a slight skewed depiction of both language groups. In other words, the differences are not as pronounced for the LPCAT as compared with the conventional Grade 9 average year marks. As De Beer points out, the broader the ability range and more reliable the criterion scores, the more informative the correlation scores will be. As this group is more heterogeneous in composition, having a broader ability range yet also having a stronger criterion test score (as all students complete the same courses), will have it such that the correlation for the LPCAT will tend towards zero. In comparison to this contention is the hypothesis that the more homogeneous a group, the higher the correlations between the LPCAT difference score and criterion scores.

This hypothesis is confirmed when investigating the correlation results for this group. All the LPCAT scores (bar the difference scores) are significant when correlated with the four terms' averages as well as the year mark. Concurrent validity of the LPCAT is affirmed in this group, as the fourth term test was conducted during the same time period. The scatter diagram representing the range of ability scores indicates that the ability range is "reasonably wide", noting that the sample size was small.

Correlations between the LPCAT and all five criterion scores were conducted for the two language groups separately. An interesting finding is that the LPCAT scores were all (bar the difference score) significant for the English/Afrikaans group but not so for the African language group for whom none of the LPCAT scores were at all significant. De Beer states in this regard that "[d]espite the fact that it [the LPCAT] may not correlate as highly with academic results for the African language group as for the English/Afrikaans group, it can provide useful additional information for selection or evaluation purposes" (2000, p.263).

Validity results of the LPCAT-2

Two samples were used to investigate the validity of the LPCAT-2 which is exactly the same as the LPCAT-1 except that the instructions are read out to the candidates and not shown on the computer screen owing to lower levels of literacy. The first group was composed of low level literacy adults (N = 194) and the second of 144 Grade 8 school pupils.

Group 4

Most members of this sample were male and spoke African languages. The LPCAT was administered to them as part of a general assessment. Criterion measures included the Paper-and-Pencil Games (PPG), as well as numeracy and literacy tests. In general, LPCAT results were lower on average for this group, but this is not surprising considering the lower mean educational level attained by this group. Scores for the verbal and non-verbal PPG are slightly different with the non-verbal scores being higher on average than the verbal scores. Both diagrams were negatively skewed, the non-verbal graph indicating a ceiling effect. The LPCAT scores were lower in general for this group, as explained before. The LPCAT results were fairly broad and distributed over a broad range of results indicating that the LPCAT can distinguish between a broad range of abilities at a lower level. Correlations with the PPG scores were all significant bar the LPCAT difference score. LPCAT pre-test scores were lower (although still significant) than the LPCAT posttest and composite scores for this group, which De Beer asserts as evidence of support "…for learning potential scores that include both the present level of performance as well as the effect of training" (2000, p. 269). This provides support for the construct validity of the LPCAT for this group, as the two tests seem to overlap on the construct they measure. Correlations with the literacy and numeracy test results were all significant and even on three accounts for the difference scores. The intended use of the LPCAT test battery was to determine potential for disadvantaged students; and for Group 4 (being low level literates) the LPCAT was able to distinguish between the various ability levels within this group. Construct as well as criterion validity for this group is thus supported.

Group 5

This group had quite a comprehensive package of criterion measures available, namely, the GSAT-CAT, academic marks, English and Mathematics proficiency test results, the LPQ results as well as two teacher-ratings. As the students were at high school level the ability levels were wider ranging, and the group was also multicultural, allowing cross-cultural comparisons to be made. Virtually all the results were significantly different for both language groups (African and English/Afrikaans) on all the measures. All the LPCAT (bar the difference scores) were statistically significant. The implication of this is that the ability of both language groups to learn after training is similar and the LPCAT picks up this attribute. In other words, although the two groups differ significantly on the pre- and posttest scores, they are not significantly different in the way they learn. Hence, there is no difference in the difference scores, which "…indicates similarities in the learning attitude of the two groups" (De Beer, 2000, p. 274).

When investigating the gender differences, no LPCAT scores are significantly different, and no differences are marked for the GSAT, LPQ nor proficiency tests. Differences are evident in the teacher ratings as well as all four term average results. The females scored higher in all these areas. Diagrams for the GSAT verbal and non-verbal, average year marks as well as the LPCAT pre- and posttest results are slightly positively skewed for both language groups. The LPCAT pre- and posttest scores are on average less positively skewed than the other scores. Once again, the scores are in general lower than the average owing to the nature of the group (Grade 8 pupils, and the LPCAT being standardised on Grade 10 level).

The LPCAT correlated significantly with the GSAT-CAT, bar the difference scores, which attests to the LPCAT's construct validity. Composite scores receive higher correlations with the non-verbal section of the GSAT, and also the LPCAT posttest correlates higher with the GSAT than with the pre-test scores. Criterion validity for this group is wholly supported when viewing the results of the significant correlations between the LPCAT, GSAT-CAT and proficiency scores with those of all four terms' results and average year mark. All correlations are significant (except the difference score). Also the

posttest scores are higher than the pre-test scores, indicating the effectiveness of training encompassed within the procedure.

GSAT results correlate higher with the criterion measures than the LPCAT, which De Beer states is to be expected due to the similar content that the two tests measure. Clearly there are noticeable differences between the language groups, the English/Afrikaans group evidencing highly significant correlations between all measures and criterion scores, unlike the results for the African language group which only evidence a few significant correlations. English proficiency scores are higher for the African group, which De Beer states as underscoring the importance of language proficiency in education. Regression lines for the different groups were similar bar the line for the African language group, for whom academic prediction would be overpredicted if scores of 35 and higher are obtained, whilst scores below 35 would lead to academic performance being underpredicted for this group.

The GSAT-CAT and LPCAT (PPC) scores correlate significantly indicating that the LPCAT measures the same general reasoning ability, providing support for the "…LPCAT as a measure of learning potential within the general reasoning ability domain" (De Beer, 2000, p. 287). However, crucial differences between the language groups are evident when it is shown that for the English/Afrikaans group, all LPCAT (PPC) scores correlated significantly with all the academic results, but only two such correlations existed for the African language group. Language once again plays a dominant role in the results as the GSAT verbal score was the only score from this test battery which correlated significantly with academic performance for the African language group.

Integration of all five groups

The main finding, which is perhaps a prominent finding, is that the difference scores between the two language groups did not differ significantly for all the groups. Secondly, the distribution of scores attested to the ability of the LPCAT to test across ability ranges. Thirdly, and this follows on from the second main finding, is that the LPCAT seems to provide more equitable measures of general ability as opposed to other measures. The LPCAT is able to measure similar constructs to other batteries, while also being able to incorporate a learning potential element. Slightly negative results which indicated low correlations for two of the sample groups could well be explained as, states De Beer, by the facts that students' lack of language proficiency may have impeded results, that they were a preselected sample, and lastly that range, in terms of ability, was restricted.

The main reason why De Beer carried out the regression analyses for the different groups was not to indicate the predictive validity of the LPCAT (which she reiterates is not the sole intention of the instrument), but to investigate and compare the different groups' total and individual lines and to relate this to the relationship between the LPCAT and criterion scores.

4.3.19.4 Conclusion and recommendations

Also emphasised in the study by De Beer is the fact that the LPCAT should not be used solely on its own but should be used to serve additional information. In her penultimate chapter, De Beer concentrates on the meaning of the difference score to determine exactly the difference indicated. Using it on its own is inadvisable, "…primarily because an attempt has been made to measure learning potential over the entire range of ability levels…" (p. 292). Some issues do, however arise, such as whether scores greater than zero do, in fact, represent significant differences between pre- and posttest scores; whether the differences are result of training or merely the practice effect; thirdly, whether the difference score measures what it purports to measure, namely, learning potential or improvement in performance; and lastly, to investigate any developmental changes reflected in the LPCAT scores.

The first issue:

Single sample t-tests are conducted to find out whether the differences between the pre- and posttest scores are significantly above zero for both language groups. Results affirm that each group's difference scores are, in fact, significantly above zero except for the third group. Reasons for this group's non-significant result may be the small sample size (N = 37), as De Beer notes.

The second issue:

In order to investigate the effect of training, De Beer sampled the sixth group, which represented the whole Grade 9 class; and students were then randomly assigned to three sub-groups:

- The first group (N = 37 which was the original third group) was administered the LCAT-1 in its standard form
- The second group (N = 35) received the LPCAT-1 and also received additional training over and above that already provided
- The third group (N = 37) received the pre- and posttest but no training (any differences here can presumably be attributed to the practice effect).

The mean differences were not surprising, increasing from the lowest difference for the third group to the highest difference for the second group. The mean difference score for the second sub-group of the sixth group was significantly above zero. This, in essence, means that the practice effect (limited exposure to the test) will not be enough to yield learning potential.

The third issue:

The LPCAT composite score correlated significantly with other measures of academic performance but the difference score did not yield similar results. The main point here is that the difference score alone cannot be used to identify learning potential. With hindsight, it would seem obvious that just such a statistic should not be used in isolation to unequivocally represent learning potential. Yet in many early dynamic assessment studies this is exactly what was used to indicate learning.

The fourth Issue:

The LPCAT evidences the ability to distinguish between different ability levels as is seen in the scores obtained across all five main groups (these groups comprising subjects at varying levels of education and age). The mean scores do reflect an increase as educational level increases. A comment made by De Beer sums up a prevalent attitude towards dynamic assessment, that "[t]he basic idea that learning potential tests should have the **psychometric properties of standard tests**, but that their administration procedure should differ in that a training phase is incorporated and improvement performance is monitored, highlights the main characteristics pursued in dynamic assessment" [own emphasis] (2000, p. 300).

In the seven years taken to develop and test the LPCAT, De Beer accomplished a number of goals set out at the start of the research, namely, to construct a valid and reliable test which was standardised and implemented on local South African testees, in order to ascertain a measure of learning potential, which would not only indicate potential ability within the reasoning domain, but a test that would also measure the same constructs that other static conventional tests measure. In so doing, items were decided upon in consultation with experts in the field; item response analyses were carried out to ensure only unbiased; and adequate items were eventually used in the test battery. Validation studies were able to provide norms for various groups, as well as the fact that a second version of the test was constructed in order to accommodate those candidates who are not literate.

Computer adaptive testing and item response theory are melded in this unique test battery designed for and implemented with multi-cultured South Africans in mind. So far in this research survey on dynamic assessment in South Africa, it is the only such test (computerised test based on item response theory) to be developed and normed locally. The battery is not only administered in a dynamic manner, but the test itself is a dynamic test. Some studies make use of conventional tests in a dynamic manner, but the tests as such are not dynamic. De Beer acknowledges the fact that only three out of ten provinces were chosen from which to draw samples, as well as the need for greater sample size in investigating the effect of training. Uniquely, this test is standardised (a criticism levelled at supposedly dynamic test instruments) but "makes up" for this by the nature of the IRT and CAT built into the test battery which allow for the standardised test to become adaptive to individual scores in the progression of the test itself. (See above for information regarding the use of IRT and CAT already discussed.) Although not intended exclusively as a test of predictability, this function nevertheless is a crucial part of the idea behind the LPCAT. As testimony to its less biased nature (when compared to some conventional tests), the LPCAT yields smaller mean differences between

the different language groups. The LPCAT takes only one hour to implement and is not costly, two major issues involved in dynamic assessment, issues which are often seen as the major hindrances towards the use or not thereof. The LPCAT takes as its underlying philosophy the work and ideas of Vygotsky, ideas that are deeply entrenched in the field of dynamic assessment. The LPCAT also takes note of the new non-discriminatory dispensation in which it must function.

PART B

4.4 Results from other South African research efforts - a brief overview

The following section centres on other South African research projects which have been implemented as shorter research projects.

Perhaps one of the earliest studies conducted in South Africa, to make use of a dynamic approach,[140] is the study by Lloyd and Pidgeon (1961) and it is only briefly looked at here, for its significance as archival material relating to dynamic assessment. Eight hundred and seventeen Indian, Black and White learners (approximately 270 in each group) between the ages of six and twelve were randomly allocated to experimental and control groups and tested on the then National Foundation for Educational Research's non-verbal test 1 and 2. Both groups were pre-tested and posttested on the non-verbal assessment instruments, with the experimental group receiving coaching, which was intended to familiarise them with what was expected in the non-verbal tests.

The entire process took place over a period of three weeks. There were significant differences between the experimental and control groups on pre-testing for both the white and black groups, which the authors state were due to factors such as withdrawal and low-scoring children in the experimental group. Although both groups (white and black) improved their performance from the pre- to the posttest, only the experimental groups' performance was significantly increased. The control groups' overall performance improved but was due to the practice effect,[141] and these gains are very similar across the control groups but differences between pre- and posttest results are varied for the experimental group.

The authors are unable to explain why the Indian experimental group failed to improve significantly between the pre- and posttest administration of the assessment. The greatest gains were made by the black group, whose initial pre-test scores were lower than either of the two remaining groups. The authors state that although non-verbal tests were used owing to their alleged culture-fairness, the fact that the Indian group did not evidence any gain and the black group did, begs the question of whether non-verbal tests really are culture-fair. Lloyd and Pidgeon (1961) do not elaborate on what is now known to be a dynamic method in testing, but do conclude that standardised test scores are not comparable across cultures, even after a period of familiarisation. They do not seem to link up the nature of their design (pre- and posttest set-up) with the results. Nevertheless, it is interesting to see that the design was used 40 years ago.

The inclusion of the Potter and Jamotte (1985) study is valid in the context of this dissertation because it is repeatedly referred to in many later studies within the field of dynamic assessment, with particular reference to alternative admissions policies at tertiary institutions. Their main contention is that African Matriculation examination results are in fact "…dubious indicators of academic merit" (p10). Shochet's (1986) sentiments and the findings in his study are also echoed in this study, the results of which were published in 1985. Potter and Jamotte collected data from the 1980/1981 matriculation year results, including pre-matriculation and matriculation examination results and were confronted with data that was inconsistent and misleading. "Matriculation symbols, being the result of a specific

140 The authors did not know at the time that their approach was dynamic. Of note is that their 1961 study coincides with the time at which American literature in learning potential was becoming available (Budoff in Lidz, 1987a).

141 Which may well have been avoided had the researchers made use of a Solomon Four Group Design.

examination, may at best be regarded as only partially significant in determining a pupil's ability, and should be complemented by other forms of testing" (Potter & Jamotte, 1985, p.10).

Two sets of data were collected from each respective year from the then Department of Education (still in the years of the Apartheid government). Students for the study were chosen on the basis of their applications for bursaries from a private sector company. For the 1980 sample, 63 students were chosen based on their relatively consistent high marks in this matriculation year. Each of these students was tested using a battery of psychometric tests. For the 1981 sample, the same procedure was used, except this time, 159 students were selected upon having their applications studied, whereas the 1980 sample were selected after in-depth interviews were conducted (following receipt of the applications). As part of the selection procedure for these students the following school performance measures were intercorrelated:

- English Junior Certificate, Standard 9 marks and matriculation symbols
- Science Junior Certificate, Standard 9 marks and matriculation symbols
- Mathematics Junior Certificate, Standard 9 marks and matriculation symbols.

(Potter & Jamotte, 1985, p.10)

The 1981 sample results were also complemented with aggregate Standard 9 results, aggregate marks on the Junior Certificate, pre-matriculation and matriculation aggregates but owing to the low level of intercorrelations between these variables for the 1981 sample, the researchers also decided to examine the results of psychometric tests conducted with the students in order to try and enhance the predictability of matriculation results. The statistical procedures used for the studies were a simple correlation analysis and regression analysis. The results were very disappointing since most of the correlations were unsatisfactory predictors of matriculation results.

Of the combined Junior Certificate measures, the pre- and matriculation as well as Standard 9 results in all three subjects (Mathematics, English and Science) only Science pre- and matriculation results as well as Standard 9 and Science matriculation results were in anyway predictive of matriculation results for the 1980 sample. None of the results was satisfactory for the 1981 sample. When regression analyses were conducted, it was found that the only strong predictive relationship was found for Science Standard 9 results and Science matriculation for the 1980 sample. There were no strong predictive relationships in the 1981 sample.

Results from the aggregate pre-matric, matriculation and Standard 9 Junior Certificate were also not promising as predictors. The highest correlation was obtained for the aggregate Junior Certificate and aggregate Standard 9 results, evidencing r =.40. This is not the only worrying factor in this study, but as Potter and Jamotte point out "[a]lso noticeable is the instability of the majority of variables in predicting matriculation symbols from year to year" (p. 13).

On average, the overall drop in results from the pre-matriculation to matriculation results was 15 per cent. Hence a low level of correlation for predictive success for all categories of results as well as the fluctuations in results make for an urgent need to re-look the notion of using matriculation symbols as the only indicators of potential success at tertiary institutions. The findings of Shochet (1986) focus on this very issue. However, caution is offered about the Potter and Jamotte results, or at least the method or manner in which statistical inferences are made from the data (Huysamen, 1999). Huysamen states that this study used a "highly select" sample thus resulting in restriction of range[142] and that the negative results obtained in this study might not necessarily reflect poorly on the quality of matriculation results, but may lie with the reliability and validity of the teachers' rating of students, which is more subjective as opposed to having external evaluators assess students on a more uniform basis (1999). It is because the Potter and Jamotte study is so often cited, that Huysamen has deemed it necessary to qualify some of the findings.

142 See section 4.3.13.3 in which Huysamen mentions the relationship between restriction of range and predictor/criterion variables.

Adams (in Adey, Steyn, Herman & Scholtz, 1994) offers an incisive overview of how the state of education should progress and the direction in which it should be moving, with particular regard to the notion of Vygotsky's zone of proximal development. It is his contention that the role played by teachers in South Africa can be strengthened if they are schooled during their apprentice years in the art of mediation.[143] Adams's article is not the result of an experimental design but more of a treatise on the status of Vygotsky's rather neglected theories in the South African education system. He also mentions Tharp and Gallimore (1988), (as cited in Adams in Adey et al., 1994) whose Neo-Vygotskian theory of teaching as mediation has bearing on this treatise, as has also been highlighted by Andrews (1996). The fact that the ZPD of any child is often thought of his/her only ZPD, is actually incorrect. Learners have many such ZPD's, which are at any time "advancing" as well as "retreating" (Adams in Adey et al., 1994, p. 3). The ZPD can be applied throughout life and if a ZPD is no longer open to change it is said to be fossilised (see also Vygotsky 1978, p. 63).

Adams pays special attention to the "socio-historic" role of teaching and the theory involved in training teachers for the classroom. Teachers must be taught the concept of mediation, a process which itself requires training. Teachers are often isolated in their day-to-day working lives, and this lack of a broader social context makes it such that teaching as defined by "assisted performance" is not recognised and unfortunately "…most teachers are unfamiliar with the concept of assisted performance, and they need to be given guidance on what is required" (p. 11). Adams then relates possible interventions which can assist teachers to become more process-orientated and less product-based in their approach to teaching learners. This notion of process versus product-based intervention is also a key factor in the new OBE[144] system in South Africa. Teachers themselves sometimes lack the very competencies that they should be instilling into their students. One way of rectifying this situation is to make available training programmes, encompassed within current training programmes, which emphasise the notion of the ZPD and mediated learning experience which, in turn, reflects a dynamic approach to education.

De Villiers (1996) also highlights the role of Vygotsky's ZPD as applied to disadvantaged learners in South Africa. De Villiers does not dichotomise the theories of Piaget and Vygotsky (as is often the case) but seeks to integrate the two theories, since only by investigating both can one hope to analyse and understand how learners adapt to new task demands. Internalisation or development of metacognitive processes as well as the social mediation given by teachers and peers are interdependent and not mutually exclusive. The inherent cognitive readiness can only be effectively expressed with the assistance of mediation from a culturally similar and competent mediator.

As Adams (1994) points out, the ZPD encompasses the life span of the learner, and according to De Villiers even the most "developed of thinkers" (1996, p. 136). De Villiers embarked on a qualitative study, of the results of interviews with 54 disadvantaged first year Technicon students from a variety of faculties. Tape recordings of the interviews were made and analysed; and the findings were further elucidated by the technique of participant observation. Themes were extracted from the various interviews; and the summary of findings included metacognitive abilities, and rules for success.

From the interviews, it is evident that students differed as to their awareness of their own learning. Some were aware that they were meeting task demands whereas others were not. Those students who were aware of their own learning were able to alter their own learning processes in order to increase the chances of success. A possible reason why some disadvantaged students were unable to recognise their own "type" of learning was their way of learning in secondary school, a method based on rote learning, rather than understanding. The current OBE system is at present tackling this very issue. In other words, product-based learning was emphasised in school, whereas process-based learning was expected of them at tertiary level.

Some students were able to recognise "rules of success", rules which now differed from the secondary school environment. In order for students to learn these new rules, the role of mediation becomes

143 See Coosner (1999) section 4.3.17 who designs a programme to achieve this very goal.

144 See Coosner (1999) section 4.3.17 who also discusses mediation and OBE.

important. Mediation of new rules by competent peers as well as knowledgeable lecturers can serve to make the transition from secondary to tertiary all the more painless. These rules for success are implicit and not evident to all learners. Although qualitative analyses may inherently lend itself to bias, the interviews showed that the main difference between "…good and poor learners is often the ability to evaluate and adapt to new task demands" (De Villiers, 1996, p. 138). De Villiers advises lecturers as well as students to be more reflective of their teaching and learning styles and notes that as with Adams (1994), the role of lecturers (like teachers) should be more attuned to a mediatory role than one of imparting information. In other words, for disadvantaged students, in particular, the implicit rules should move into the "public" arena where they can be internalised and implemented. The role of a dynamic approach, specifically Vygotsky's ZPD becomes even greater in South African higher education.

Lopes et al*.,* (2001) conducted a predictive validity study on the Ability, Processing of Information and Learning Battery (APIL-B) and, in general, found the results to be positive. The APIL-B as a dynamic assessment battery has been used and studied in the arena of tertiary selections and admissions (see studies by Nel [1997] as well as Engelbrecht [1999] and findings by the Potchefstroom, Rand Afrikaans and Port Elizabeth Universities).

Lopes et al. (2001) applied the APIL in an industrial context and thus add to the research of dynamic assessment in areas other than tertiary education with a move towards commercial research. The study consisted of a sample if 235 successful applicants to a large insurance firm, who were rated by managers on a graduated five-point scale of performance and thus gave an indication of each employees ability and potential. This served as the criterion variable. The reason why this research was undertaken was to illustrate that there is indeed a test available on the South African market which has been developed locally and takes into consideration the implications of the Employment Equity Act (1998) which, in essence, states that tests should be shown to be scientifically valid, reliable, be applied fairly to all testees and that it should be unbiased to all groups concerned (Lopes, et al., 2001). The question posed is whether or not the APIL can be used to predict learning potential.

The ratings given by managers of the performance of various employees may be construed as biased and in order to overcome this a qualitative method of assessment was used by the authors, namely, the Elliot Jaques "Critical Incident Approach"[145] on a randomly selected 37 participants. The various scores derived from the APIL are made up from both static and dynamic test procedures. The entire APIL battery was not used[146] but sufficient dimensions were tested to make a sound conclusion. The ratings given by managers were treated as nominal[147] data due to the fact that women on average received higher ratings than did men. Furthermore, according to Chi square analyses, there was no difference in how managers rated the various race groups. Canonical discriminant analysis[148] was applied to the data and according to the original five-point scale, the number of "correct" classifications made by the APIL were in fact not as accurate as was hoped for and therefore it was decided to collapse the classification into two categories, namely, ratings 1-3 and ratings 4-5. Of course, it is obvious that the fewer categories there are, the more accurate the classifications become. Nevertheless, the proportion of correct classifications improved.[149]

145 This entailed interviewing the manager about his choice of rating for each of the 37 chosen employees. This would at least partially ensure that ratings were unbiased.

146 The study of Nel (1997) in section 4.3.14 also makes use of a shortened version of the APIL - which is interesting to note, as this may be indicative of the time needed to administer this battery.

147 Here the authors refer to differences between the scores but assign no quantitative value to either score.

148 Canonical discriminant analysis looks **at two sets of variables** and "…is an additional procedure for assessing the relationship between variables" (http://www.statsoft.com/textbook/stathome.html).

149 This is the third study which "collapses" the classification categories as fewer categories yield better results. See also Nel (1997) section 4.3.14.3 and Engelbrecht (1999) section 4.3.16.3, who similarly collapse their categories.

APIL scores for the black employees were consistently lower than for the other groups, but as Lopes et al. (2001) state, language proficiency may have played a role in these lower results. Intercorrelations between the various sub-tests of the APIL were shown to be high, thus evidencing that the battery measures the same variable. When the original five point rating scale was used, the accuracy of correct classification by the APIL amounted to 36%[150] but when the categories were collapsed, this increased to an average of 72% for the two categories. Lopes et al. (2001) conclude from the study's findings that "…the APIL-B is a reliable instrument when applied to job applicants within a financial institution" (p. 68). These, then, are positive results for the dynamic assessment battery developed and used in South Africa.

In an attempt to bridge learning potential assessment with cognitive[151] skills training, Taylor (1999) discusses the APIL and TRAM1[152] and TRAM2 batteries focusing on them as potentially very useful instruments of assessment in South Africa, because they do not rely on prior knowledge, and because they also include non-dynamic aspects in the tests. "Learning potential tests also seek to achieve perfect understanding of the instructions, but then evaluate individual differences in terms of learning variables rather than knowledge variables" (Taylor, 1999, p.3). The APIL consists of eight dimensions, namely,

- Conceptual reasoning (fluid intelligence)
- Speed of information processing
- Accuracy of information processing
- Flexibility of information processing
- Automatisation which is defined as the amount of improvement over a period of four sessions in which learning as well as study sessions take place and this represents a "curve of learning score"
- Automatisation which is defined as the total amount of work done over the four sessions
- Memory and understanding which is evident though the learning curve score
- Transfer.

TRAM1 and TRAM2 have five and six such components, respectively. Taylor states that although the tests do not contain tasks that relate to those encountered in real life, they do nevertheless correlate significantly with criteria such as training courses. Taylor delineates a set of reliability coefficients for a variety of criterion variables such as the Mental Alertness Test and various in-house rating scales that have been devised by different companies with which he has conducted these validity and reliability studies.

Taylor differentiates between two types of learning potential, Type A and Type B. This is a crucial distinction to make, as Taylor adheres to the Type B learning potential approach which differs from the classical Feuersteinian mediated and individuality centred approach to assessing potential, which is characteristic of Type A. Type B is characterised by conventional learning opportunities in which training and study sessions take place and feedback is given to testees. Both the APIL and TRAM batteries are Type B learning potential tests. Taylor acknowledges the disadvantages of classifying the APIL and TRAM as type B tests, but asserts that the advantages outweigh the disadvantages of making use of Type A learning potential tests, such the Learning Potential Assessment Device (LPAD). Issues such as costs, timing, reliability, validity and logistics of training administrators,[153] are inhibitory factors when administering Type A learning potential tests. The learning potential

150 The authors state that even 36% must be compared to the norm accepted twenty years ago when the amount of variance explained between measures of cognitive ability and job performance were in the order of 9%.

151 In fact, research conducted by Swanson (1995) has shown that dynamic assessment of information processing ability (focusing on cognitive abilities) can be relatively successfully implemented, thus merging both psychometric and dynamic principles.

152 See Van Aswegen (1997) section 4.3.13 for a validity study on the TRAM1.

153 Taylor (1999) refers to these issues cumulatively as a "non-starter" (p. 7). Costs and other issues simply cannot be ignored. Dynamic assessment in terms of learning potential tests cannot afford the luxury of individually assessed, administrated and scored testing procedures.

characteristics of Type A tests he adds, may not emerge when conventional tests are used but can emerge when Type B tests are used. Although a "deeper" understanding is gained from Type A learning potential tests, Type B is favoured for the above-mentioned reasons. Furthermore, the APIL and TRAM tests were developed with two main aims: for ease of administration, (in group format) and in order to adhere to objective scoring techniques. Taylor differentiates between trainability and potential, stating that trainability tests are usually used to select employees for training in a particular working context, where there is a large component of manual work (blue collar work) and most trainability tests have to developed for specific work scenarios.[154] On the other hand, learning potential tests are "generic" and are not aimed at specific job types and contain no material which signifies work of any particular field. However, according to Sehlapelo and Terre Blanche (1996), "…trainability …is quite similar to that of learning potential" (p. 54). Nevertheless, these authors maintain that learning potential testing is rising in popularity but warn that it should not be construed as a technical solution to a problem which is socio-political in nature.

Regarding the development of cognitive skills which are more often than not lacking in disadvantaged learners, (skills which are erroneously assumed to exist and hence which are tested for in conventional tests), Taylor places specific emphasis on two aspects of cognitive logical thinking, the analytical and the conceptual. A programme which has been developed by Taylor is known as CogLab and seeks to address these cognitive deficits and the main intention of the programme is the preparation of individuals for placement in first levels of management as well as for tertiary education. The programme contains tests and the material is based on work and practical issues (thus deviating from the more abstract tasks often encountered in Feuersteinian type tasks). Time is given to those who need it and hence the CogLab programme is not a typical standardised test.

The APIL and/or TRAM can be used in combination with the CogLab programme. Based on scores received on the APIL or TRAM, individuals are chosen to pursue the CogLab programme. Basically, this involves those who evidence greater learning potential and hence are more likely to succeed in the programme. In the limited studies thus far conducted on the CogLab programme, results have been encouraging. This is a good example of dynamic assessment which is coupled with remediative activities.

Another learning potential test, The Conceptual Reasoning Test[155] (CRT) has been used, among others, by Nunns and Ortlepp (1994) as part of a battery of tests to ascertain predictive validity of measures with first year psychology students (both advantaged and disadvantaged students) at the University of the Witwatersrand. The CRT formed part of the Arts Faculty Ratings and states that "[t]he Conceptual Reasoning Test assesses the ability to reason logically without the use of words or Mathematics, and provides a measure of **potential** intellectual ability, rather than achieved intellectual ability" [own emphasis] (1994, p. 205). The CRT proved to be significantly correlated with academic success in psychology for the whole sample as well as the disadvantaged sub-sample. It was the only test to correlate significantly with the disadvantaged group of students while other more static measures such as the Mental Alertness Test (MA) and Reading Comprehension Test (RCT) (of the Intermediate Battery as developed by the HSRC) did not correlate significantly for this sub-sample.

The MA[156] and RCT[157] did, however, correlate significantly for the group as a whole. The CRT is, as the authors state, a test which evaluates learning potential and one which indicates which students have the potential to achieve within "accelerated" conditions. It is a measure independent of prior schooling, but larger samples are perhaps required to more fully investigate the predictive validity as

154 Von Hirschfeld and Downs (1992) define trainability tests (TT) as "…a form of selection in which a job applicant is taught how to perform a task or function that is critical to success in that job" (p. 25).

155 See section 4.3.18 for more results on the CRT.

156 See section 4.3.6 , 4.3.8 and 4.3.18 for information on how other studies have made use of the Mental Alertness Test (MA) as part of static predictors of academic success when compared to dynamic tests.

157 See section 4.3.18 for further information on the Reading Comprehension Test (RCT) as used in predictive studies along with more dynamically attuned tests.

well as cultural bias of the CRT (Nunns & Ortlepp, 1994). Nevertheless, the findings for this limited sample are positive.

Shochet (1994) equates dynamic assessment with interactive assessment and used the two terms interchangeably when discussing the results of a study. He concluded that cognitive modifiability (CM) (as originally proposed and operationalised by Feuerstein [1979] as cited by Shochet [1994]) serves as a moderator effect in the prediction of academic success of disadvantaged students at a university. CM significantly moderated the predictive validity of static test measures for the sample of disadvantaged students. Shochet maintains that interactive (dynamic) assessment goes beyond merely coaching or hinting and its aim is to assess structural modifiability.[158] The hypothesis tested in this study is a tenet espoused by Feuerstein, namely that "[t]he predictive validity of [a] traditional measure should reduce with an increase in modifiability and improve with a decrease in modifiability" (1994, p. 210). Shochet sets about administering a static measure of cognitive ability to a sample of disadvantaged students to obtain a static measure. Dynamic assessment proceeded with a test-teach-test approach using two static measures. The gain score (difference score) represents the level of cognitive modifiability. It is hypothesised that the predictive validity of the traditional assessment will be moderated by CM for disadvantaged students.

The Deductive Reasoning Test[159] (DRT) was used as the static measure of intellectual ability. The DRT along with the Pattern Relations Test (PRT) was administered dynamically (pre- and posttest assessment after a session of mediated learning experiences) as dynamic measures and gain scores of both tests were added together to give a composite gain score. Criterion measures included academic results obtained at the end of their first year at university either in the form of number of credits obtained during the year and/or average results from all subjects. A t-test for paired differences concluded that there was a significant difference between pre- and posttests scores. Shochet employed the same measure he employed in his 1986 study (see section 4.3.1.3) namely, moderated multiple regression analyses (MMRA) which "…regresses the criterion on the predictor variable, the moderator variable, and the cross-products of the moderator and predictor variables. The second technique compares sub-groups ['less' and 'more' modifiable students] formed by partitioning the sample at a particular cut-off score of the moderator variable" (1994, p.212).

A significant result was obtained from the MMRA which evidenced the function of CM as a moderator. The CM as moderator effect significantly improved perdition of the criterion score (number of credits obtained), but did not significantly improve prediction of the alternative criterion score (average results obtained). However, the model of prediction becomes significant for both criteria. The advantage of using the MMRA technique allowed Shochet the opportunity to investigate two different groups (sub-grouping) of disadvantaged students, namely, "less" modifiable and "more" modifiable students (as it has been reiterated in this survey that disadvantaged students are not an homogeneous group and can therefore not be classified as modifiable on the basis on their disadvantaged status only). As expected, the traditional static measure (DRT) correlated significantly with both criterion measures for the less modifiable group. The results for the more modifiable group were not significant. This supports the hypothesis mentioned above. Shochet is cautious in his interpretation of results, stating that less modifiable students may have achieved a high enough DRT score and due to this, room for improvement was not as great for them as for the more modifiable students. These statistical artefacts are considered. This notion is counteracted by the fact that there is no negative correlation between DRT and CM, a negative correlation would have resulted had "…CM [been] a statistical function of the baseline measure of DRT" (p. 212). There was also no differences between the distribution of results for the less modifiable and more modifiable groups.

These results point to the finding that static measures are not able to differentiate between students who posses more potential than others and that with the inclusion of dynamic measures, not only is

158 However, some tests are considered dynamic if such training does occur without any other mode of dynamic assessment involved, such as the identification of cognitive modifiability and remediation.

159 See research by Shochet (1986) section 4.3.1, Zolezzi (1992) section 4.3.8, Zolezzi (1995) section 4.3.10 and University of the Witwatersrand research, section 4.5.2.8 for other results using the DRT as well as the PRT.

the validity prediction of success increased but one is also able to further define sub-samples of students, by differentiating between more and less modifiable students (valuable information in any context, especially tertiary institutions, where subsidies should rather be spent on those who possess a greater likelihood of success during their studies). The study thus provided empirical evidence which supported the notion of an inverse relationship between predictability and modifiability when using static measures. Shochet highlights the fact that this finding should not be ignored, as this aspect of validity and empirical support often result in dynamic results being criticised by conventional psychometrists. Shochet does not turn a blind eye to the fact that learning potential is only one aspect of tertiary admissions and states that educational experience and background form the other "half" of the picture, since modifiability (when used on its own), does not predict or constitute academic success (Shochet, 1994).

Moll (1989) delivered an insightful observation of the three theories of cognitive development which perhaps has the most fruitful application within the field of dynamic assessment, namely, grounded in Piaget, Vygotsky and Bruner.[160] Moll emphasised the importance of matching the method of teaching with the current levels of cognitive development (Piaget), of teaching as a process of social activity and various meanings being mediated by an adult or peer (Vygotsky), and lastly, of teaching that is a mixture of both Piagetian and Vygotskian approaches, in that teaching must still remain accessible to the learner (child) but that teaching can in fact change a learner's developmental level. Although no formal study was conducted by Moll, the contribution to the dynamic assessment field is important for foregrounding this synthesized approach.

In a similar vein, Bradbury and Zingel (1998) investigated the best manner in which to facilitate the learning of second language learners from disadvantaged educational backgrounds. Their main purpose was to "…assess the nature of peer interaction in a multicultural group and its effectiveness with respect to possible effects on learning" (1998, p. 233). The context of their investigation is grounded in the theories of Vygotsky, Feuerstein and Freire. Paulo Freire's view towards learning is that peer interaction ("equal dialogue" or "peer collaboration" as cited by Bradbury & Zingel, 1996) takes place in a "horizontal" manner, and that the learning process should, in fact, take place through peer interaction with co-equals. In this method, authoritarian learning is not emphasised. In contrast to this approach, the authors cite the theories of Vygotsky and Feuerstein, in terms of mediated learning experiences in the presence of informed peers or adults. In order to investigate these notions and theories further, the authors conducted a study in which school children, aged 11 years were observed in their interactions with certain learning tasks.

The findings reveal that the nature of peer interaction differs depending on group composition. In this empirical study, low performing learners managed to cope on their own as a group, but performed poorly in certain tasks, as opposed to when they were placed with more competent peers. In this case their performance improved. Two kinds of groupings can be evidenced from this research depending on the nature of learning that will take place: if the aim of learning is to develop group participation skills, it is advisable to place similar children together (in terms of performance); but if the aim is to develop task competence, it is advisable to place poor performers with more capable peers. In this way, the theories of Freire as well as Vygotsky and Feuerstein (in terms of peer mediation and interaction) can be validated and in so doing "…balance both the social and cognitive development of learners" (p. 239).

Sibaya, Hlongwane and Makunga (1996) elaborate upon Feuerstein's Learning Potential Assessment Device (LPAD) as a means with which to assess giftedness,[161] intelligence and other cognitive abilities in specific groups (black South Africans). Two main aims of their research include investigating informal methods of assessment procedures used in South Africa by psychologists and secondly, developing a dynamic assessment procedure (DAP) in order to assist psychologists,

160 Jerome Bruner's theories draw on both those of Piaget and Vygotsky emphasising "…action as a starting point for the formation of abstract, symbolic thinking…", stressed by Piaget whilst also valuing the role played by social interaction emphasised by Vygotsky (Lloyd, 1995, p. 27).

161 See Hoffenberg (1988), section 4.3.2, and Gaydon (1988), section 4.3.3 for research on gifted children.

trained in western approaches, to assess people from other groups. Findings from this qualitative survey revealed that many psychologists prefer to make use of standardised intelligence tests that have been normed on white samples, because there are hardly any such tests normed on so-called minority groups (hence the wish to develop a dynamic assessment test). "In light of the problems associated with current psychometric testing, it seems logical to explore the applicability of a dynamic assessment, using intelligence tests" (Sibaya, et al., 1996, p. 110).

The DAP follows the test-teach-test approach and the authors discuss the multitudinous benefits associated with dynamic assessment, but do recognise the issues that plague this field of assessment, such as timing, costs and efficiencies in administration yet unanimously state that "[t]he advantages of using dynamic assessment procedure (DAP) outweigh its disadvantages" (p. 112). Sibaya et al., (1996) make mention of a number of static intelligence tests (such as the Raven's Progressive Matrices) which can be adapted to a dynamic method of assessment. However, it must also be noted that they do not propose the development of a new and unique dynamic test but highlight the role of static measures that can be administered in a dynamic manner. (This difference in dynamic assessment definition and practice has already been highlighted.)

In an attempt to induce cognitive change through learning, Miller (1998) states that the University of Natal instituted a programme in 1996 for first year psychology students. Potential is realised, as Miller asserts, more so if students are allowed to modify their learning behaviour cognitively through sustained efforts and engagements within structured learning situations. The course (programme) functions as a learning system, where knowledge is conveyed by means of tutorials and discussions; and emphasis is placed on understanding of material rather than information being fed to learners as is most often the case. Homework assignments are completed[162] and returned, after which students receive detailed feedback. This process allows disadvantaged students the opportunity to familiarise themselves with the tasks and cognitive requirements of tertiary education.

The (1998) study focused on second year psychology students and investigated whether or not they benefitted from the foundation programme. Results were mixed but the findings cannot be interpreted as simply as that. Miller states that when it comes to investigating foundation or intervention programmes, many extraneous variables enter the equation. For instance, it is incorrect to assume that ability levels remain the same (not to mention that it is the ability levels that are targeted for change in the first place). Some students performed less well as the programme continued whilst others achieved in performance scores as the programme progressed. This is not a reflection of lack of ability but more likely a reflection of unwillingness to engage with the tasks on hand. Commitment and diligence are factors that also need to be taken into consideration.

Foundation courses are usually aimed at students who require specific academic development, but it is not always possible to identify those students. The main emphasis that Miller makes is that different groups (composed of low and high performing students) cannot be dealt with as a homogeneous group, since they clearly are not. Initially, low performing students can achieve or even at times out-perform initially high performing students, whereas initially low performing students never perform any better than they did to start off with (motivation, diligence and potential are factors which may/may not necessarily reside within each disadvantaged student). "At best, the results of this study indicate that a preliminary assessment of the specific abilities a programme is designed to address would serve to identify, on an individual basis, those students who potentially could benefit from the programme" (Miller, 1998, p.174).

162 This is similar to the idea underlying the Teach-Test-Teach (TTT) programme, see section 4.5.2.1.

PART C

4.5 Results of the informal interviews - information gathered from Technicons and Universities

This section of the study comprises information gathered from the informal interviews conducted either telephonically or via electronic mail and in some cases through face-to-face interviews with the respective authorities involved in student selections and admissions at Technicons and Universities in South Africa. In many instances, the appropriate authorities were those involved in foundation programmes, bridging programmes and alternative admissions programmes at the various institutions. Only those tertiary institutions that made use of dynamic assessment in some form or another, whether in the past or at present are discussed here.

It is salient to note that there are many institutions that currently implement alternative programmes as stated above. However, these programmes did not meet the criteria as set out by the definition of dynamic assessment in even its broadest sense. These bridging programmes, alternative admissions initiatives and extended foundation programmes are valuable but do not form part of this research study and are thus not included.

Of value is the fact that on numerous occasions persons who were interviewed had either not heard of dynamic assessment, or did not understand the full meaning behind the term. Those who did know of dynamic assessment were either directly or indirectly involved in the procedure. Although not a statistically significant finding at all, it is worth documenting this sort of information. This highlights the diffuse definition of the term.

As discussed in Chapter Two, dynamic assessment as a term may yield a wealth of meanings and interpretations. However, even the most commonly understood meaning was in many instances misunderstood. The following two sections centre on the additional information gathered from these informal interviews with due acknowledgement to the various institutions from which the information was obtained. Lastly, a finding which at first may not seem to be problematic but upon further consideration makes for unease, is that at times a large number of persons involved in this field of research do not always know of other's work in the field. Of course, it is impossible always to be aware of every shred of information, but it is deemed acceptable scientific practice to familiarise oneself with as much of the information as possible. Once again, dissemination of information seems to be paramount.

4.5.1 Technicon contacts and work being done

As with Universities, Technicons generally make use of some form of alternative assessment vehicle or assessment battery. These are, more often than not, run in parallel with assessment of matriculation examination results. These alternative admissions tests, although noteworthy and commendable, do not all constitute dynamic assessment and thus are not all included here. Two Technicons offer what can, broadly speaking, be seen to be dynamic assessment.

4.5.1.1 Technicon Pretoria

The programme run at this Technicon is not strictly a dynamic assessment procedure, but is included here because of the very dynamic nature of its selection and admissions approach. This Technicon makes use of an internally designed admissions programme, which students attend only after selections. Students are placed in a programme only once they have been selected at the Technicon, so the dynamic nature of the programme becomes evident only after selections have been made. This, then, does not aid prospective students. In other words, learners who may be disadvantaged in terms of education are not all given the opportunity **before** or **during** selections (initial selections are still negotiated based on matriculation examination results, a criterion which has been less than satisfactory in many instances) but only once they have been chosen, and this is the crucial difference between this institution and say, for instance, the TTT programme previously conducted by the

University of Natal (see section 4.5.2.1). The title of the selection test that is used is the "Technicon Pretoria Potential Assessment" (TPPA) which, strictly speaking, is not a potential test (H. Kriel, personal communication, September 5, 2001) in the dynamic sense of the word (see section 4.2.2 on ambiguous use of dynamic concepts).

The TPPA as a selection test is followed by an intensive learning programme lasting as long as 25 weeks in some instances. The selections and admissions are a two-fold process: one in which the Technicon assesses the students for entry into specific courses and, then, based on performance on the TPPA, these students are instructed to attend workshops throughout the year.[163] The workshops that are attended by every student regardless of performance on the TTPA entail working on a computer. A programme entitled the "English Word Power" test can be construed as a dynamic as it has a pre-test assessment phase in which initial functioning in English proficiency is assessed, followed by remediation (remediation is offered by means of examples that are shown to students on the screen, rules are taught and grammatical rules are explained) which can last from 10 hours to 25 hours depending on each student's schedule. The test is individually tailored to meet each student's pace. After the workshops have been completed a posttest assesses any differences that the mediation may have resulted in.[164] It is standardised assessment, and human mediation is minimal. The English test is divided into a grammar and spelling section. This test has only been introduced this year (2001). After three years, correlations can be calculated to see how it has faired in preparing students for studies at the Technicon and may also be able to furnish researchers with predictive validity results. At present, it is used only for remediation purposes and not as a selection and placement instrument. There is, as yet, no published information about the programme run at this Technicon. Once again, it must be noted that this procedure veers towards a dynamic manner of assessment, in that students (only once they are selected) are offered remediation and cannot be construed as a pure dynamic assessment procedure.

4.5.1.2 Natal Technicon

According to Sue McKenna (personal communication, September 5, 2001), the Placement Test in English for Educational Purposes (PTEEP) - as develop and also used by the University of Cape Town (see section 4.5.2.2 for an in-depth overview of what exactly the PTEEP is and what statistics are currently available for it) are being used for the first time this year. This battery of tests is primarily used to broaden admissions to this Technicon. However, the PTEEP is only written once matriculation examination results are assessed and based on these results, the test is written by lower performing students. There is thus still a reliance on matriculation examination results as an indication of academic ability. The PTEEP is used as a "broadening access and placement tool", but because this Technicon intends merging with ML Sultan Technicon which makes use of the TELP test (see section 4.6) the PTEEP will be done away with. It must be noted that McKenna states that the PTEEP and TELP are said to be very similar in nature, and that the PTEEP, according to the pre-test, posttest definition of dynamic assessment is not necessarily considered a dynamic assessment test. This Technicon also makes use of the SPEEX (as has the Technicon Pretoria, with which they have also been in contact).

The PTEEP's main function at this Technicon is to counteract the cultural differences between students who arrive at the Technicon seeking admittance. A research group has been put together to oversee the investigation into dynamic tests that are available which will replace the older psychometric tests used in the past. The emphasis is now on language competency.

---

163 The Technicon Pretoria also makes us of the Situation Specific Evaluation Expert (SPEEX) test which is an updated version of the Potential Index Battery (PIB).

164 According to the Pretoria Technicon, results have so far shown that some high pre-assessment scorers actually perform less well in the post-assessment phase, presumably owing to the fact that students have to think about rules of spelling and grammar, aspects that have not been considered before and hence cause confusion.

4.5.2 University contacts and work being done

Although three universities have maintained steady publication of research in this field, other universities have also contributed to the pool of research available.

4.5.2.1 University of Natal - Teach-Test-Teach programme - (TTT)

According to Bradbury and Griesel (in Adey et al*.,* 1994), materials, which have been developed at the University of Natal as an alternative access programme, focus on "…the nature of texuality (and the socially constructed nature of knowledge) and the possibilities of mediating this through distance education materials" (pp. 323-324). The authors state that the TTT programme's first phase took place from 1988 - 1990. During this phase, students were selected for Arts, Social Sciences and Law. The data gathered from the initial trial made available the necessary information to implement an entrance examination from 1991 - 1993. The programme known as the TTT (Teach-Test-Teach) programme centres on the dynamic notion that teaching[165] should precede assessment so as to allow everyone an equal footing. However, a problem inherent in this approach is that one must be cognisant of the background from which different students come and the new environment which they seek to enter. Craig (personal communication, March 6, 2001) states that "the **big** difference between TTT's adaptation of a dynamic assessment protocol and other more typically Feuersteinian approaches is that the latter uses - in principle - **contentless** remediation , whereas our task was to assess adaptation to content and form of university tasks…". This was so, as the TTT programmes' specific intent was to cater for the university population, so students would be chosen based on their success in the programme.

The TTT programme made no use of any specific battery. Craig explains that "(w)e used many ideas, but no standard battery. I think that is what dynamic assessment is, ideally - demands, i.e. the working out of a specific gap between specific learners and specific tasks" (A. Craig, personal communication, March 6, 2001). The crux of the research endeavour into the TTT programme can be summed up from Craig's description of the task demands and students' abilities, namely that central to the view of Vygotsky's framework for dynamic assessment, "…was the insight that measures of **potential** for learning [have] to capture (through task design/educational intervention and assessment of learning):

1. What the learner does know/can do, i.e. what he/she brings to the learning situation
2. The nature of the task/task demands
3. That which will bridge (1) and (2) or that which will scaffold[166] the learners 'learning', i.e. the process of mastering the task demands.

[The TTT programme] focused specifically on the necessary analyses to bring about and sharpen our grasp of (1) and (2)".

(A. Craig, personal communication, March 6, 2001)

The concept of dynamic assessment has also been referred to as "developmental assessment" (J. Bradbury, personal communication, March 7, 2001) which is perhaps a more explanatory term.

---

165 Bradbury states that "[t]he primary assumption of the TTT approach to selection is that, where prior learning opportunities have been inadequate or inappropriate, assessment for selection must be based **on teaching before testing**" [own emphasis] (1993, p.8).

166 Refer to Yeld and Haeck (1997); Coosner (1999) for more information on scaffolding. Bradbury (1993) qualifies the term by stating that "[t]he learning-teaching process is a process of 'scaffolding' understanding within the zone of proximal development to enable the student to fully develop his/her potential ability" (p. 8). Day and Cordon (1993), for instance, have found that scaffolded instruction might be a superior method of instruction and diagnosis in education. Das and Conway (in Haywood &Tzuriel, 1992) refer to scaffolding as "…an adjustable and temporary support" (p. 109). Schaffer in Daniels (1992) describes the term as designating all the strategies that adults use in order to help children's learning efforts through supportive interventions, aiding them to achieve goals that would otherwise be out of their reach.

The TTT programme has since 1995 "diminished rather than increased" (*ibid*.) for those involved in the project at the time. These include Craig, Griesel, Miller and Bradbury. Of interest is the fact that additional queries were made at the University of Natal as to their selections and admissions of disadvantaged students and the possible use of dynamic assessment. At this stage (2001), it did not seem as if dynamic assessment was receiving any measurable attention. Further queries as to the status of the TTT programme yielded information that suggested that the programme was abandoned owing to costs and time. An attempt to circumvent these problems was made by distance learning (Skuy, Zolezzi, Mentis, Fridjhon & Cockcroft, 1996). This (along with reliability and validity issues) is contentious in the field. Regarding resources and intense administration of the programme, Griesel states that "…we deliberately need to create quality learning opportunities which are resource-intensive.[167] Some have followed this route, and … assessment [then] takes on a very different nature from that of once-off selections testing" (2000, p. 5).

There are two viewpoints on dynamic programmes. On the one hand, if one is to introduce and implement a dynamic programme (with all that it entails in its purist form[168]) then it is wise to see the programme through from beginning to end, knowing that the end results will be trustworthy and "reliable" in the dynamic sense of the term. On the other hand. time and money are factors that are too large to be ignored and if, at the cost of streamlining the process, a few tenets (held strongly by purists) are done away with, then these are the consequences one would have to live with. "It is logistically difficult to apply a procedure [dynamic assessment] to large numbers of prospective students that involves teaching prior to testing and to process the kind of qualitative data yielded by this approach" (Miller & Bradbury, 1999, p.31). Once again, it is reiterated that a compromise will have to be found if dynamic assessment is to flourish in South Africa as a means of assessment. Sternberg (1991) states, with reference to his particular test approach (implied by the triarchic theory [1985]), and which is similar to the notion expressed above, "[that] [s]tandard and cheaper products are financially safer, and thus achieve a level of priority that more expensive, riskier projects are not likely to achieve" (p. 267).

Griesel (1991) states that in response to the selection crises that were facing South African tertiary institutions in the late 1980s and early 1990s, "…selection as an education response [at the University of Natal]… [was held] as a first and basic commitment" (p. 19). Griesel refers to the TTT programme as one in which students are first given the opportunity to "…engage in typical university learning tasks in order that their 'potential for success' can be assessed" and that once these students make their way to university, this "potential" is then consolidated, and just as Craig highlighted the learning process involved in assessment; educational development and learning processes are "…inseparable from the selection component of the TTT programme" (p.19).

Essential to the Teach-Test-Teach programme are two component principles, namely, that the assessment component (teach-test) which entails the idea that those with the potential to master learning tasks when given the opportunity, will do so and secondly, that the educational development component (the teach aspect[169]), which according to Griesel (1991) must allow for the process of consolidation of the abilities shown in the "teach-test" phase. It is perhaps this second principle which differentiates this TTT programme from other programmes, which emphasises the need to consolidate what has been learned, by means of a process of remediation. This aspect of remediation is closer to

167 This is precisely what Lidz (1987) emphasises when she states that "[a]ssessment alone has a limited value unless it can guide intervention" (p. ix).

168 This sentiment is substantiated by two primary aims of the TTT programme, the latter which finds particular emphasis here, namely, that the programme "…has been two-fold: selection and educational development" (Bradbury & Griesel, 1991, p. 1). The development is needed as part of the assessment. This is, in fact, critical as underprepared (in contrast to unprepared) students need to be afforded the chance to engage in tasks before a decision can be made as to their suitability for university education.

169 Daniel (1997) places emphasis on dynamic assessment's "…strong connection with education because it focuses on learning as a variable and because, in some of its forms, it is as much a teaching-remedial procedure as an assessment technique" (p. 1042). This notion is highlighted by the TTT programme.

the notion of Feuersteinian dynamic assessment than many other dynamically attuned studies the focal aim of which is to select and admit via a dynamic assessment process, but which, more often than not, neglect to remediate after the process. The TTT programme centres on the notion of providing opportunities to students "…that **alter** the very abilities which are treated as fixed in conventional tests…" [own emphasis] (Griesel, p. 21). This sentiment is reiterated by Craig (2000) when she refers to "…tasks that **will elicit change - tasks that will provoke cognitive adaptation**" (p. 7).

The TTT programme typically recruited disadvantaged learners from the surrounding communities who were identified by the community; placed them in a two-week selection programme (later changed to an entrance examination); and then proceeded with "educational development work". The two-week programme afforded the students an opportunity to engage in tasks which dealt with problem solving, meta-cognitive levels of problem solving as well as the "epistemic level of problem-solving" (Griesel, 1991 p. 23; Bradbury & Griesel, 1991). These are factors that were considered imperative for success in university studies. Discipline-specific tasks were used in the process. These two-week sessions included work which was to be completed after-hours, daily work and an examination.

A technique referred to as "defamiliarisation" in which students are presented with material the content and form of which are not overly reliant on prior learning, was used, so as to overcome any potential barrier of erroneous learning which may have taken place at an earlier stage (Bradbury & Griesel, 1991). "New content-knowledge and new ways of operating on that content and developed through a systematic programme of tasks for guided action [was developed]" (Bradbury, 1993, p.11). Of importance here is that students were to become engaged with the written text and "…through reading effectively to construct their own understanding from text " (*ibid*.). A meta-cognitive level of thought processing is involved here, not merely learning how to solve problems but effectively to learn what mental faculties are involved in certain problems and when to utilise these mental "actions". In other words, emphasis on epistemic assumptions was paramount (Craig, 1991). Adult cognition is, in part, concerned with the epistemic nature of knowledge, i.e. understanding the limits of knowledge, its assumptions and underlying nature. It is this epistemic knowledge which Craig asserts is not a part of most African educational experiences and it is this aspect which the TTT programme strives to address. The theory of knowledge and how knowledge is gained is fundamental to this approach. The TTT programme in a way seeks to empower students, more so than merely equipping them with additional know-how. The course is also not narrowed in the sense of concentrating on only one subject at undergraduate level but seeks to employ the whole range of subject matter, so as to familiarise students with broader "interdisciplinary tasks" (*ibid*.).

Once again, this allows all students relatively equal opportunities to complete the tasks. An interesting finding and one which supports the notion of potential as highlighted by Griesel (1991) is that of the four groups that can be delineated based on the programme. One group (those for whom degree studies are clearly not suited according to matriculation examination results) encompasses students who excel during their second year of study, at times overtaking those students who are initially selected for degree studies. "Data on student performance in the first year of their academic studies (1989) bear testimony to the spurious relation between matriculation results and university performance, and show unequivocally that a dynamic procedure of assessment/selection is superior to selection on the basis of matriculation [examination] results where the results fall outside the top range scores, as is the case with students selected thought the TTT programme" (Griesel, 1991, p.27).

Griesel furthermore emphasises the important point that potential cannot be measured as such, nor as a means towards predicting academic performance without making allowance for the opportunity to engage in a learning process. "To separate selection from the very educational processes whose success it is designed to predict, merely compounds the educational challenge universities face" (1991, p. 28). A decade has passed since this statement was made, and yet it rings true today and will most likely ring true for yet another decade. The most fundamental concept behind the TTT programme is its insistence on teaching and consolidation of what has been learned. It is not to be construed as a test but an instrument of teaching. It is a teaching and learning programme with part of its outcome being the selection of potential students. Of all the studies surveyed in this research, the

TTT programme is perhaps one of the more advanced and closest in terms of Vygotskian theory and implementation.

That cognition is formed socially, is emphasised by Bradbury (1993), who states that "[t]he relation between social and cognitive realities is therefore one of mutual construction and change" (p. 4). Here reference is made to Vygotsky's "mind in society" concept or "society in mind" as being equally valid. The individual mind is a socio-historic construction, notably one that cannot be viewed otherwise. The TTT programme's initiative was based distinctly on a process-driven path, its overt goal being that of learning provocation and of maintaining a research process which assumes the process-nature of the task. This cannot be more aptly put than Bradbury (1993) who states "…[the programme had] been operating from within a cognitive framework to both investigate the question of student selection, and engage in educational development **both prior to and post** selection" [own emphasis] (p.7).

A further differentiating feature of this TTT programme was its insistence on admitting students to the university in order to find out if they would be able to cope with the demands of tertiary education. Many studies apply dynamic assessment precisely to find out whether students will be able to manage, but many programmes are not as lengthy and detailed. However, one must take cognizance of the fact that much of the research undertaken in this field is by way of theses, studies that are conducted within limitations, of both cost and time. This TTT programme was an initiative with considerable financial backing when compared to these aforementioned studies. The aim of the TTT programme was underscored by what Miller has to say regarding prediction and selection, namely, "…we must not look to a predictive entrance test to achieve accuracy but to the nature of the educational intervention that we provide" (Miller, 1991, p. 4). Furthermore, the programme sought to provoke new learning which would accelerate students' abilities to allow them to meet the demands placed on them at university (Bradbury, 1993).

Although the programme was utilised for selections and admissions, its core function was in allowing the manifestation of potential, latent within individuals who may not have been afforded the opportunity of tertiary education but nevertheless possess the potential to succeed. "The fundamental assumption is that education provides learning opportunities that **alter the very abilities that are assumed or treated as fixed in the construction of psychometric/edumetric tests**" [own emphasis] (Griesel, 1991, p. 30).

"The TTT programme was initiated not merely as an alternative approach to the problem of student selection but as an alternative paradigm within which to view and understand the problem" (Miller, 1991, introduction). According to figures obtained early on in the programme, initial poorer performance in first year as opposed to second year studies "…should not be understood as indicative of a permanent condition…requiring constant attention throughout a student's academic career" (Miller, 1991, Part Two p. 1). Students selected for the programme showed very good results with 85% of second year courses being passed. 30 students were originally selected from the two week programme in 1989 (from a total of 98 students who attended the two week programme) and 66 students were chosen for the 1990 year, of which 90 attended the two week programme. Results were as encouraging for this cohort as they were for the 1989 cohort. Based on these results, Miller states that "[i]nstead of understanding DET students in terms of their low matriculation results or lack of preparation, we **can begin to understand them in terms of their propensity to improve**" [own emphasis] (1991, p. 1). Miller (1991) also draws attention to the fact that the fairly homogenous group known as "disadvantaged" is not actually a homogeneous group, and distinguishes between prepared, underprepared and unprepared students for whom different interventions are necessary. The TTT programme has since 1995 been disbanded, or in any event is no longer functioning in its previous capacity. It is perhaps one of the more insightful programmes entrenched within the dynamic assessment school and is philosophically sound in these tenets. Perhaps a revival would be greeted by enthusiasm from some educationalists.

4.5.2.2 University of Cape Town - Alternative Admissions Project (AARP)

According to the Centre for Higher Education Development (CHED) at the University of Cape Town, The Alternative Admissions Research Project (AARP) is run by this testing service who, in the initial stages of the project (1986), made use of pre- and posttest scores correlated with academic

achievement, but this was abandoned owing to "…time, organisation and budget constraints" (A. Visser, personal communication, July 19, 2001). The project was introduced to address the "gross inequalities" of the past and the conditions still existing in schools in South Africa and to provide a more equitable basis of assessment for all students. The primary aim, initially, was to identify educationally disadvantaged students who were likely to succeed at the university (Badsha & Yeld, 1991).

Two of the three specific tests within the AARP programme, the AARP Placement Tests in English for Educational Purposes (PTEEP) and the Mathematics Comprehension Test are considered dynamic assessment tests.[170] According to a project document from the CHED (no date), the AARP testing procedure is characterised by the following four key points:

- Academic literacy and numerical competency tests which "…tap core competencies … and reflect the demands of tertiary level of study" (p. 2) have been put together by multi-disciplinary teams
- "The AARP tests target potential", which is characterised by carefully structured and sequenced teaching. It involves modelling, includes practice elements and opportunities within the test, "…which have the effect of eliciting far wider ranges of performance than are achieved with traditional testing approaches" (p. 2). The test is a paper-and-pencil test which minimises time and costs, as prolonged contact is not necessary[171]
- The AARP is not a replacement test and hence does not undermine or compete with the matriculation certificate
- Most importantly, the AARP tests are not curriculum-based, in other words, content is not based on academic subject matter and the tests are "…not pitched at a homogeneous group" (CHED report).

According to internal UCT studies, the AARP's English and Mathematics tests have yielded positive results, illustrating that results of 60% and 50% of educationally disadvantaged students, in both the English and Mathematics Tests, will yield higher retention and graduation rates than those achieved by students who are admitted solely on matriculation results. For 40% and 50% of the time, both groups will yield equal rates. "…at no time do students recommended by the project prove to be at higher risk than those routinely admitted to the university on the basis of their [matriculation examination] results" (*Ibid.*). The point with the AARP tests is to test for potential and in so doing, allow the university's admissions department some insight into who will mostly likely succeed and graduate within the minimum of time. This issue is not a new one, as universities are in the process of becoming more privatised and less reliant on government subsidies. The onus thus lies with the university to ensure that those student who are chosen to study will indeed perform.

The AARP's PTEEP test[172] can be characterised as a dynamic assessment[173] test based on the following criteria:

- The theme/content that is used in the assessment is novel to all students
- The content is teachable through mediation and tasks are modelled
- Tasks become progressively more complex
- A variety of texts is available.

---

170 The Mathematics Achievement Test is another test within the AARP battery but is a multiple choice item test and relies heavily in previous learning. It is not a dynamic assessment test.

171The fact that timing and costs are taken into account are valid aspects to consider, but one must take cognizance of the essence of dynamic assessment which entails time as a crucial factor. However, the AARP is not alone in trying to minimise time and costs within the field.

172 "The test incorporates a combination of teaching, modelling, and practice elements and opportunities. This is accomplished in the test by providing opportunities for candidates to engage with and manipulate the texts before they are required to produce any extended piece of writing" (www.uct.ac.za/depts/aarp/tests.hrtm).

173 See section 4.5.1.2, for instance, where the PTEEP is not considered as a dynamic assessment test. This reiterates the current confusion as to what constitutes dynamic assessment and what does not.

The AARP's Mathematics Comprehension test[174] can also be characterised as a dynamic assessment test based on similar criteria:

- The content is not based on prior learning [175]
- Students are supplied with the necessary rules and procedures (which is construed as "teaching")
- A learning experience is provided for students, in such a way that understanding is "built up through answering previous questions" (p. 3), and hence this is not based on the actual answers themselves
- The language used in the test is closely monitored.

The Mathematics Comprehension Test is an approach which "…aims to assess the ability of a candidate **to learn** from authentic academic material within the test" (Haeck et al., 1997, p. 71). Learning experience[176] is provided in the test, so that answers to later questions are reliant upon what was learned in previous questions. Haeck et. al. state that one of the indications that scaffolding was successful in this test, was the facility value which decreased with the progression of questions. The facility value is worked out from the total scores of candidates divided by the total possible scores, on a scale from 0 to1. Once again, as with almost every dynamic assessment test, the function of the test is not to supplant matriculation examination results, but to aid in the discovery of potential that may be latent within the individual, potential that is not evidenced in conventional test situations. The AARP has been able to prove (see Haeck et al.) that students gaining admissions into university based on their AARP results have not only done as well as students admitted based solely on matriculation examination results, but have in some cases performed better; and these are students who have not performed well on their matriculation examination results. Test questions are not based on content and there is no reliance on prior learning within the test. Reliability and predictive validity results are discussed at length in the text (see Haeck et al.).

Based on a draft report of the predictive validity of the AARP's PTEEP test (2000), the most crucial aspect of importance to deans and admissions officers is the reliability of the PTEEP's results, which should be relevant to all groups, high and low achievers. In survival analysis studies[177] (which focused on retention times as opposed to performance), descriptive statistics yielded that students who were allowed to enter the university based solely on their matriculation results, evidenced higher matriculation results than the group who were admitted based on both matriculation and PTEEP scores. This is in keeping with national trends, as it has been demonstrated that it is usually low achieving or disadvantages students who complete alternative assessment tests. The PTEEP test itself is a good discriminator of performers. As has been mentioned, high performing students in the PTEEP have either higher retention rates or the same retention rates when compared to students admitted on their matriculation scores alone. However, low scoring PTEEP students have higher rates of exclusion than those admitted based only on their matriculation results. There is currently a doctoral thesis being completed on the full validity and reliability of the PTEEP test (A. Visser, personal communication, July 19, 2001).

---

174 This test "…has been designed to minimise the effect of prior learning in mathematics, thereby enabling all students to demonstrate their ability to learn and apply mathematical knowledge. This is achieved by using topics which fall outside the scope of the school syllabus and therefore are unlikely to have been taught in a formal setting, and by using a text which serves as a guide to answering increasingly complex questions" (www.uct.ac.za/depts/aarp/tests.hrtm).

175 This is especially crucial for Mathematics, seeing as competencies in Mathematics can be deduced from sources other than matriculation results.

176 This is where definitions within dynamic assessment become blurred. The question is whether or not learning (in this manner, as described above) can be equated with mediation and whether or not mediation forms part of the function of a mediator. A related question is whether learning (in this instance), can take over the role played by the mediator and If so whether or not this can still be construed as dynamic assessment. It depends on the definition one uses and this is where the continuum of definitions comes to the fore once again.

177 In the draft report, it is stipulated that many "…well-known survival probability distribution functions were fitted to the empirical exclusion-time data…" (CHED, 2000, appendix) and of these, the Weibull model was utilised.

Mediation (or scaffolding[178]) within the PTEEP test consists of a "…guided summary [which] is furnished in the form of a cloze passage" (Yeld & Haeck, 1997, p. 10). Cloze is a technique in which key words are left out of a sentence, words which are crucial to the core of the meaning conveyed by the sentence. The learner thus constructs a summary of the text and in so doing learns in the process, and when given further instructions, will be able to use what has been learnt in the tasks already completed in tasks ahead. The mediation provided does not result in all candidates performing in the same manner, in other words, scaffolding does not take away from the discriminating function of the PTEEP (which is, after all, a test meant to discriminate between students who are able to benefit from university education and those who would not benefit). "…[s]caffolding has not simply made the test easier for all candidates, but has widened the gap between weaker and stronger candidates" (Yeld & Haeck, 1997, p. 14). Perhaps the PTEEP is, as the authors state, an innovative way of eliciting potential, which is becoming the main function of a selection test today.

Griesel (1999) states that the AARP's achievements are really two-fold, namely, to promote students' probability of success at university while retaining ("protecting") equity at entry levels. This AARP does by admitting and placing students as an alternative to exclusive reliance on matriculation results. Griesel also cites a number of reasons why this project has been successful, reasons which many other institutions may lack, namely,

- Administration support
- Costs, timing and management efficiency
- On-going research and tracking of students' performance
- Sound testing and hence trained testers
- The fact that the AARP programme functions through the CHED within the university.

Yeld and Haeck (1997) state that in order to be an effective selection tool, the AARP (or any other test or programme intended to test for potential) has to:

- Understand the essential academic skills that need to be tested
- Know how the tasks function as teaching agents
- Be implemented by an institution which should provide post-selection opportunities (note the similarities between the TTT programme which also stresses post-selection teaching).

The AARP programme has, since its inception, widened its services to 22 of the 36 tertiary institutions[179] (Griesel, 1999).

4.5.2.3 University of Pretoria - Alternative Admissions Project (AARP)

According to the Registrar's office at the University of Pretoria as well as the unit for student support, the AARP tests are also conducted for students whose matriculation results are not entirely indicative of their potential (as is the case with the University of Cape Town). The AARP, having been developed and co-ordinated by the University of Cape Town, is being implemented at Pretoria at present. The AARP's three sub-tests are used for two types of student populations, namely, for "top" medical students and for "under standard" students who, through assessment of their potential, are allowed the opportunity of being "upgraded" (E. Esterhuizen, personal communication, July 19, 2001).

Unfortunately, further information (regarding survival analyses, reliability and validity statistics) about the AARP as used by this university could not be procured. This information would be most useful as the student profile at this university may differ from the profile of students for whom it was originally intended. Hence, the AARP sub-test may necessitate refining for this particular student population. It is known, however, that the university acknowledges the inadequacies of previous more static type

178 Scaffolding is related to Vygotsky's ZPD (Yeld & Haeck, 1997; Craig [personal communication, March 6, 2001] and Bradbury, 1993) and is also mentioned in the TTT programme.

179 See the University of Pretoria below, section 4.5.2.3 for instance. However, upon informal interviews with many of these institutions, none made mention of the AARP.

tests that assessed prospective students on prior academic content. There has been a move away from this to a more dynamically attuned approach.[180]

4.5.2.4 University of Potchefstroom - APIL test battery

The University of Potchefstroom currently makes use of the APIL battery amongst other test batteries for selection purposes. However, it must be noted, that the research conducted by Engelbrecht (1999) has bearing here, as the results of that study (see section 4.3.16) are not encouraging, yet the APIL is still used as part of the selection tests. However, more information on this matter was unobtainable. Suffice it to repeat that the APIL is used, even though prior research shows its ineffectual use in that particular study.

4.5.2.5 University of Stellenbosch - development of an internal dynamic test battery

According to the University of Stellenbosch, no dynamic assessment as such exists. However, it was made known that this university is in the process of developing a "dynamic assessment test battery" for the "diversity programme", a programme for previously disadvantaged students. The test battery was at this stage (August 2001) being compiled in conjunction with the sociology, psychology and theology departments as well as the student church. There is, as yet, nothing published in the accredited literature. Planned implementation is due to take place in 2002.

4.5.2.6 University of Port Elizabeth - APIL test battery

According to Terence Taylor (personal communication, July 9, 2001), the APIL has been used by student services at the University of Port Elizabeth. The APIL was, in fact, used twice, for the 1998 and 1999 student intake. However, due to revisions of student selections and admissions programmes, the focus of this university turned towards assessment of basic student competencies and not potential (A. Watson, personal communication, July 25, 2001). The 1998 use of the APIL was not for selection purposes, but "[a]s part of the [Special Admission Test Battery's] ongoing research into the predictive validity and 'culture-fairness' of the battery…" (Watson & Ncapayi, 1998, p.1). According to a study conducted in 1998, the APIL was found to have moderate and significant correlations with academic performance and its further use (based on this finding) was made. The sample consisted of 140 applicants who registered with the university. The criterion variable was the average marks obtained during the examinations. Of the seven sub-test scores (which make up the APIL battery) moderate correlations were found for two of the sub-tests and criterion performance for the group of candidates as a whole. Further statistical analyses, per faculty breakdown, yielded significant positive correlations for four of the six faculties (Watson & Ncapayi, 1998). There were, however, a number of negative correlations between the Science and Economics faculties. The results are nevertheless encouraging for further use of dynamic assessment at tertiary institutions in South Africa. Although it must be noted that this is not the only field of application open to dynamic assessment, this area is one in which dynamic assessment is particularly fruitful.

180 The department of English at the University of Pretoria introduced a dynamically attuned test in the1990s to test for potential. The test was developed in such a way that no prior knowledge of English grammar and spelling rules was assumed. The test sought to identify those candidates who would be better suited to mainstream English or practical English. The aim of the analysis was "…to compare **potential** as predicted by the Entrance Test with **performance** in the First Semester and, *de facto*, the likelihood of students passing the ENG101 course at the end of one year" (Gray, 1990, p.1). No mediation took place during the test. Results on the test were then correlated with matriculation symbols, medium of instruction at school and the first semester aggregate mark and the sample size was 96. The results indicated that with a correlation coefficient of 0.75, the "…Entrance Test can be regarded as a reasonably good predictor of results at the end of the First Semester…" (*ibid.*).

4.5.2.7 Rand Afrikaans University - APIL test battery

According to Terence Taylor (personal communication, July 9, 2001), the APIL test battery has been used at this university. According to Deon van der Merwe (personal communication, July 20, 2001), the APIL was tested on students selected at this university. At present, static matriculation examination results, cognitive and non-cognitive predictors of academic success are being used, although the concept of learning potential has, in the past, been addressed in the study of Nel (1997) (see section 4.3.14). These results have been discussed above, yet it must be reiterated that the APIL and GSAT did not explain any more variance of results when used together than when the two are used separately. Although the findings for the APIL are not negative, they are not overwhelmingly positive either, in other words, not positive enough to warrant further use of the APIL at this institution.

4.5.2.8 University of the Witwatersrand - dynamic assessment

According to Shirley Pendlebury (personal communication, July 18, 2001), the WITS School of Education has not made use of learning potential assessment devices and an English language placement test along with a biographical questionnaire that is used during assessments. However, no dynamic component is used (B. Domeris, personal communication, August 2, 2001). The university's admissions department is aware of the APIL but have yet to use it. However, this is an indication that there is at least communication between various bodies, institutions and private ogranisations within the dynamic assessment field.

The Cognitive Research Programme which is run at the University of the Witwatersrand does, however, make use of dynamic assessment (M. Skuy, personal communication, 26 June, 2001). The various Witwatersrand Master's degrees and Doctoral studies (mentioned in Chapter 4) form part of Witwatersrand research activities which have grown from the "…successful implementation of Feuerstein's Instrumental Enrichment (IE) Programme of thinking skills among disadvantaged gifted children from Soweto" (http://www.wits.ac.za/fac/education/cogrp.htm). Other work includes studies conducted in the area of Feuerstein's mediated learning experience. Skuy (1989) states that the dynamic assessment approach is exemplified by Feuerstein's LPAD and remediation is not based solely on cognitive deficits (as with static tests) but is concerned also with emotional enrichment.

Early in the South African research efforts in the dynamic assessment field, Skuy, Archer and Roth (1987) made use of Feuerstein's LPAD in an attempt to discover the usefulness of the test battery in assessing and remediating a learner who evidenced a specific learning disability. It was, according to the authors, the first of its kind, at least in South Africa. The programme, which was a lengthy one[181] due to the individual attention and individually attuned nature of the study, provided positive results. The aim of the study was to bridge "…the schism in special education between process- and content-orientated models [of assessment and remediation]" (Skuy et al., 1987, p. 53). The LPAD aided in assessing the learners' strengths and deficiencies (as opposed to only identifying weaknesses), and showed that mediation can prove useful with learners who are impulsive in judgements, that the supportive mediation as opposed to the stringent examiner approach used in static tests was much more effective in making the learner feel at ease. From these findings, the authors were then able to devise a remediation programme for the learner (based on the subjects taken at school) and successfully implement this programme (which was in itself an extension of the LPAD). The only drawback of the study was a lack of transfer measurement, by which school tasks could be assessed for improvements of cognitive tasks learned during the administration of the LPAD.

In another study conducted by Skuy, Mentis, Durbach, Cockroft, Fridjhon and Mentis (1995), deprivation of sufficient education under the then Apartheid government made for a useful study, in which the authors assessed the effectiveness of Feuerstein's Instrumental Enrichment (IE) across

181 The study was conducted over a period of three months and was divided into two parts, firstly administration of the LPAD and secondly, the devising of an intervention programme. Part one was comprised of 25 hours of assessment, with part two taking 12 hours.

three of the four major racial groups in South Africa. Four groups made up the sample, one coloured group, one black group and two white groups, the last group being split into an Afrikaans and English group. Although the focus of the Skuy et al. (1995) study was on mediated learning experience as well as instrumental enrichment, it is included here as part of the survey.[182] The aim of the study was to determine the effectiveness of a mediated learning experience (MLE) and IE programme run by the authors. This was achieved by ongoing seminars about MLE for teachers and IE packages which was suitable for secondary school children of all race groups. In order to determine the effectiveness of the programme, criteria such as scholastic achievement, cognitive as well as affective measures were used. In total, there were five criterion measures. The duration of the study was a year (once again, the nature of the study necessitates the length of time). A number of workshops were held in which teachers were taught the MLE and IE approach. The aim was for these teachers to eventually bring into their everyday curriculum lessons information they had learned in the workshops. Pre- and posttest measurements were taken in order to assess for significant changes in the learners' abilities. However, nowhere in the study's results is there mention of the fact that improvement due to natural causes/development may take place and that any changes that are evident may not solely be attributable to the MLE and IE programme.

MANOVA results indicated that there were differences between all four groups in the five pre-test measures. Subsequently, another MANOVA yielded results which showed that the different groups performed significantly differently on some criteria measures after mediation. A Bonferroni t-test showed that the Black group improved significantly more so than the other groups, followed in order by the Coloured, and white (Afrikaans and English) groups. The Black group's initially low scores on the criterion measures allowed them a greater chance of improvement, for the mere fact that they started off on a lower platform. However, Skuy et al. point out that "…the improved post-intervention performance of the African children does not merely reflect a regression to the mean, as suggested by the significant improvements achieved by groups who were performing at higher levels" (1995, p. 278). Interestingly, the authors note that the lack of significant improvement scores on the part of the English group are not caused by the initial high scores or "initial superiority of functioning" *(ibid.)*, yet they fail to state what exactly the reason for this "lack of improvement" is. It is possible that this may be due to a ceiling effect. If they perform well on the pre-test and not have much room for left improvement, how can a better score ever be significant? Skuy et al. (1995) offer a tentative reason, stating that the nature of the quality of the mediation may have differed for this group. However, this is inconsistent with what they stated about the fact that all teachers had attended the workshops. Surely then, the teachers all received the same information. Nevertheless, the mediation proved most valuable for the black group and the study's results were positive.

Lack of consistency in teacher implementation may have skewed the results somewhat and this is acknowledged by the authors, and it would have been more beneficial to the integrity of results, had the researchers made use of a control group, so as to account for any natural cognitive improvements that the learners may have made and changes that may be due to these developments may have been erroneously ascribed to the intervention programme. Inherent cognitive changes at secondary level account for substantial improvement no doubt.

The field of application for Feuerstein's MLE is not only limited to the cognitive sphere of education but is attuned to the emotional and cultural side of education as well. MLE's multidimensional nature allows for the constructs to be applied to cross cultural "aspects" which without resolution may remain forever in conflict (Skuy, 1997) and that "…the concept of mediation as reconciliation and promotion of harmony is also integral to [the definition of mediation]" (p. 120). All three spheres, cognitive, emotional and cultural, are intimately bound up in the meaning of mediation. Thus MLE can aid in the cultural transformation taking place within South African education at the moment.

182 As has been mentioned, the two avenues of research are dynamic assessment as characterised by the LPAD, and test-teach-test type of methodology and secondly, as characterised by instrumental enrichment. The latter have (to an extent) been researched in South Africa, but did not form part of the parameters of this survey into dynamic assessment.

Remediation may be effective in increasing scores on the Raven's Progressive Matrices for different groups of academic students, using intervention techniques such as Feuerstein's MLE for example (Rushton, Skuy & Fridjhon, 2001). In a recent study conducted by Rushton et al. two identifiable groups of engineering students were evident, based on results on the Raven's tests. They state that a major question to be answered by the study would be to find out if a training method (based on MLE presumably) would increase performance "…through mastery of subject-specific knowledge or if it increases general problem-solving skills that would apply to other subjects…" (p.90). They state that such training has led to successful results with psychology students who were also tested on the Raven's battery and, who also evidenced two distinct groupings based on the results of a study conducted by Skuy, Gewer, Osrin, Khunou, Fridjhon and Rushton (in press) as cited in Rushton et al., 2001.

A study conducted by Skuy, Zolezzi, Mentis, Fridjhon and Cockcroft (1996) utilised a dynamic test-teach-test process among other static measures in order to ascertain the most effective predictors of university success for both advantaged and disadvantaged groups of students. The Pattern Relations Test was administered in its traditional form, which served as a baseline from which any effects of modifiability could be assessed when administered the enriched version of the Pattern Relations. Mediation took place following the administration of the traditional form of the test. Criterion measures included results in four subjects, Mathematics, Accounting, Statistics and Business Studies.

The authors acknowledge the small sample, N = 26 (18 of which were classified as disadvantaged) and state further that more revealing statistical methods could not be used, such as regression analyses for this reason. Although they remain confident in the reliability and validity of results, general conclusions cannot be made. Of interest is the fact that matriculation examination results were not predictive for either of the groups. The fact that the authors acknowledge other findings (such as Shochet, 1992 as cited by Skuy et al., 1996) in dynamic assessment as being positive predictors of success, made the results of this study all the more surprising, since the dynamic assessment predictors did not correlate significantly for either of the groups on the criterion measures. Other significant correlations were found with other non-cognitive predictors however.

It is very encouraging for the dynamic assessment field to see that local researchers are forging ties with international ogranisations, institutions and researchers (Skuy, Kaniel & Tzuriel, 1988[183]). Just as the research of Hoffenberg (1988) as well as that of Gaydon (1988) undertook to investigate the use of the LPAD with gifted children from disadvantaged communities, so too has international research been underway in similar contexts, namely with gifted Israeli children in disadvantaged socio-economic status conditions. Although the Skuy et al., 1988 study was not a South African study, South African collaboration is noted. The findings for this particular overseas study yielded positive results. The continuing research of the Cognitive Research Programme at this university in terms of dynamic assessment research output is admirable.

4.6 Alternative admissions and dynamic assessment

Although the focus of this study is not an analysis of the various alternative admissions programmes run at the various tertiary institutions in South Africa (such as foundation programmes, bridging courses and the like), because many institutions use some form of **non-static** measure of assessment when allocating student numbers to various faculties, some of these programmes have been included in the general discussion. Most of these programmes can, however, not be classified as dynamic assessment. Alternative admissions programmes are geared towards allowing students to prove themselves more fully in an environment which allows more opportunity for them to compete for entry. It is clear, from the many informal interviews, that most, if not all tertiary institutions are more than willing to accommodate as many students as possible, especially students from disadvantaged educational backgrounds. Non-static measures of assessment do not mean they are dynamic; they merely mean that traditional forms of assessment are either done away with or are modified and/or

183 The work of Skuy has received recognition from overseas researchers such as Sternberg (1998) as well as Lidz and Macrine (2001).

supplemented with new tests which are tailored to the new environment in which tertiary institutions find themselves.

There are as many examples of these types of programmes as there are tertiary institutions and not all can be mentioned. Perhaps one of the most wide-ranging projects currently run at a number of tertiary institutions includes the Tertiary Education Linkages Project (TELP) which is a co-ordinated project spanning many historically disadvantaged universities in South Africa. This project consists of a TELP standardised assessment test which is also written as part of admissions requirements at these institutions. TELP touches on the following important key focus areas within tertiary education:

- Student Academic and Social Development
- Staff Academic and Administrative Development
- Programme and Curriculum Development
- Research
- Management and Administrative Development.

(www.uncf.org.za/telp/components/comp_def.htm.)

Among other attempts at fairer selections which are not traditional and seek to offer greater opportunities for all students to be admitted to tertiary studies, Dawes, Yeld and Smith (1999) proffer a method in which the aggregate matriculation examination result obtained by each student is ranked within that particular school. In so doing, the school is ranked in terms of results. They have termed this indicator the Place-on-Examination (PoE) indicator. The study's results correlate with future academic performance. According to the TELP headquarters in Cape Town, TELP is a programme which is being co-ordinated at the University of the North West and the University of the Western Cape among others. The TELP test battery is not a dynamic test battery in the strict sense of the word, according to the University of the Western Cape's admissions department. However, there are certain aspects which lend themselves to a more dynamically attuned manner of testing which is illustrated by a departure from static and conventional tests used in the past. The TELP test assesses writing skills, grammatical skills and cognitive skills. In its assessment of "at risk students", interviews are conducted and an assignment is undertaken by the student in his/her mother tongue (usually Xhosa/English or Afrikaans). Based on the TELP test results, the assignment as well as the interview (non-cognitive assessment), an all-round assessment is made for each candidate. This university also makes use of a mentoring system, in which each new student is allocated a senior student mentor to help with a myriad of issues. The reason why the TELP is embroidered upon here (although not classifiable as a dynamic assessment test) is to illustrate the growing trend at tertiary institutions to move away from once-off static tests and a move towards a more holistic assessment experience, one in which the student is allowed certain opportunities to prove him/herself in the academic context, without being overly reliant on prior academic knowledge.

Another example in which an institution utilises specific tests of their own is the UNIFY[184] project run by the University of the North. The authors of a recent article which assessed the issues facing South African student selections concluded by stating that,

- More equitable tests should consist of tests which do not rely on prior academic knowledge, but should assess "programme related skills"
- Relevant matriculation results are good predictors of academic success if used correctly[185]
- Single session tests are more practical than "test-teach-test" approaches

184 Zaaiman, Van der Flier and Thijs (1998) discuss the issues surrounding selection at previously disadvantaged universities and state that the UNIFY "…selection tests were aimed at testing subject-related problem-solving skills and insight, while requiring little subject content knowledge" (p. 97). The UNIFY project is an example of an alternative foundation programme and is not a dynamic test procedure. Yet, as with many dynamic testing procedures, it does not rely on prior academic knowledge. The UNIFY programme is a one-year foundation programme "…which prepares students for entry to the Faculties of Mathematics and Natural Science, Health Sciences and Agriculture" (Zaaiman, 1996, p.113).

185 "Used correctly" is quite a vague sentiment and it would be preferable if the authors would elaborate on how best to go about using matriculation results "correctly" when so many other findings have suggested that matriculation results are simply not reliable predictors of academic success especially with disadvantaged students.

- Dynamic tests have not been shown to yield greater predictive validity than single testing sessions.

(Zaaiman, et al., 1998)

The third point is contentious as it is interpreted as meaning that dynamic tests are considered dynamic only if tests are administered in a "test-teach-test" format, which is clearly not the only definition of what constitutes a dynamic test. Tests can be dynamically administered by making use of hints and lessons without utilising a "test-teach-test" approach. When used on their own, dynamic tests often do not contribute more than static measures. However, when dynamic tests are used in combination with static measures, greater predictive validity is yielded and correlations with success are greater than when used separately, as has been pointed out a number of times already.

Likewise, the University of Durban-Westville makes use of the ART model which is not a dynamic assessment test battery, according to the Registrar's office. However, it is included because of the manner in which testing for prospective students is approached. Tests are administered at the start of a semester and if a candidate should fail the test, he/she has to attend workshops throughout the year. After attending this programme another examination is written and the candidate is then assessed for improvement on the test. It is assumed that the programmes that are attended serve to teach what is necessary for the candidate to succeed during the second test session. This approach is indicative of a dynamically assessed approach. This ART model is administered by the languages department. Practical sessions are held every week. There are two English Language development courses, the UDW 100S and UDW 300S respectively. These serve the purpose of a foundation course.

4.7 Unimplemented work as it stands

In the effort to obtain as much information and current research on the topic of dynamic assessment in South Africa as possible, a few research studies were located on various databases in South Africa. In particular, there are four documents ranging from 1987 - 1999 which deal explicitly with learning potential and dynamic assessment - all of which incorporate these words in their respective titles. These sources were supposedly to have been undertaken by researchers at the Human Sciences Research Council (HSRC). Upon inquiring about their whereabouts, it was concluded that these documents did not exist as they either never got off the ground or were abandoned early on in the research process. On the one hand, it may indicate apathy towards the topic, but on the other hand, it at least conveys to the interested researcher/practitioner that this area of research has not be relegated to the bottom of the testing theory barrel. The same scenario arose with a number of other titles, details of which can be found in APPENDIX 1. Clearly, had these projects been completed, it would have added tremendously to the pool of already available research in the field of dynamic assessment.

There is a dynamic assessment instrument that has been developed by the HSRC[186] but has as yet not been published. The material used in the test is colourful and is mainly administered in pictorial form and makes use of stimulus material with which many children can identify across cultures. It consists of a number of sub-tests which tap certain cognitive concepts and is applicable to learners in the first four grades of school. The test is combined with a remedial programme[187] and seeks to benefit the learner in terms of identification of cognitive functioning and processes rather than cognitive products. It also relies on new content and does not rely on prior knowledge. Although the relationship between administrator and testee is informal, there are nevertheless guidelines that have to be adhered to for the sake of reliability and validity. Perhaps most notable in this dynamic test

186 As this test battery has not been published, core information cannot be divulged at this stage. Suffice it to say that this is considered a dynamic assessment test by its developers. Only information of a general nature can at this stage be addressed.

187 Taylor (1999) mentions this particular test and adds that he has "…attempted to minimize the skill demands that the instrument places on the administrator (who will normally be a teacher) and improve the reliability of scores over existing batteries" (p. 6).

battery, is the fact that learners are given multiple opportunities at completing the tasks before more difficult tasks avail themselves. This allows each learner the opportunity of not "falling behind" as such and consequently becoming demotivated. Cognitive theories of eminent theoreticians such as Sternberg are used as foundation for what the test seeks to measure. Although the test is said to be time-consuming, a number of sub-tests can be left out if only a shorter period of time is available. Due to the nature of the test (as with so many other dynamic assessment tests) administrators have to be trained. At the time of writing this, the test was said to be in field on trial runs.[188] Unfortunately, more information about this test cannot at this stage be divulged.

4.7.1 Further studies

Despite efforts to obtain copies of certain studies conducted within the field of dynamic assessment, not all research studies could be located for a number of reasons. However, these studies form part of the pool research and are thus important. According to Engelbrecht (1999), the study conducted by Kotze (1996) was an investigation into the predictive ability of the Ability, Processing of Information and Learning Battery (APIL). The results show that the APIL did not prove to be a very successful predictor of academic success and "…did not deliver significantly to the explained variance of academic potential of first year students" (1999, p. 130). Yet, according to another study (anonymous researcher as cited by Engelbrecht) conducted in 1997, the APIL did reveal positive results in terms of predictability. The findings in terms of the APIL are contradictory. Correlations obtained in yet another APIL study, conducted by Watson and Ncapayi (1998), also yielded positive results, showing that correlations between the APIL results and academic results were significant. For further information on the APIL refer to section 4.4.

A Doctoral research project has been completed and submitted to the University of Cape Town during the completion of this present study. It deals specifically with the reliability and validity of the AARP's PTEEP used by the Centre for Higher Education Development at that university (see section 4.5.2.2). Although survival analysis is available for the PTEEP test, this study focuses on the predictive validity of the test as opposed to the predictive use of matriculation results. Unfortunately, this thesis was not available for inclusion here. Details of other such studies can be found in Appendix 1.

Chapter five condenses the information gathered in Chapter 4.

188 Information subsequent to this yielded that the test was no longer being used and that the project had been halted.

CHAPTER 5 Implications of results for dynamic assessment in South Africa

"Reasoning at every step he treads
Man yet mistakes his way,
Whilst meaner things, whom instinct leads
Are rarely known to stray."

- Cowper

5.1 Introduction

This chapter reviews the results from both the literature sources and those gathered from informal interviews. The main emphasis is placed on the former as this comprises the bulk of the research data. A brief summary of the results is detailed with implications for further research to be conducted in South Africa.[1] From the conclusions drawn from the various sources, an attempt at implications for dynamic assessment research in South Africa has been made. The findings of these research studies serves as a guide to future endeavours in this field and highlights those areas needing urgent attention as well as those areas which are commendable and have offered insights.

5.2 Summary of results from literature sources

The following section now briefly concentrates on the conclusions reached by each researcher and, from these implications for further research in South Africa is investigated. The main conclusions are highlighted in each section.

5.2.1 Results of the Shochet (1986) study

Shochet's study is one of the very few research initiatives that were conducted in the mid 1980s in South Africa and these serves as the basis upon which subsequent research results are often built. Most research in this field references his "pioneering" work. He investigated manifest and potential performance at university and sought to show how enriched conditions could benefit and enhance the chances of disadvantaged students' success.

This is the first of many such studies with the focus being on the disadvantaged learner in South Africa. Shochet conducted this study during a period which still implemented regressive educational policies for many black learners in the country. Shochet's two hypotheses tested in his study were:

- Both advantaged and disadvantaged students would have different predictors of university success
- Prediction of university success will be significantly enhanced by enriched conditions, implemented by Feuersteinian measures.

For advantaged students, matriculation examination results explained the greatest amount of variance. For disadvantaged students, matriculation examination results were not at all predictable. At this point, enriched predictors did not correlate significantly for the disadvantage students. Shochet investigated this further in the second hypothesis. The enriched conditions did not enhance predictions for either advantaged students or disadvantaged students. However, this hypothesis was investigated via a main effects scenario. In other words, Shochet investigated hypothesis two once again, looking for moderator effects, effects that would improve upon the predictors for certain isolated

1 For in-depth details as to conclusions in each study, refer to Chapter 4. For the sake of brevity and clarity, only the main issues within each conclusion are discussed so as to provide a sense of unity when viewing the results one after the other.

groups of students. Shochet did find a moderator effect for the disadvantaged students and also discussed that as modifiability increased, predictability of traditional scores decreased.

Therefore, advantaged and disadvantaged students evidence different predictors for university success. Disadvantaged students' scores are enhanced by receiving Feuersteinian enrichment but this is not so for advantaged students, whose matriculation examination results are still the single best predictor. Although the enriched conditions did not manifest in main effects, a moderator effect was established as enhancing prediction. **These are positive results for dynamic assessment in South Africa. Implications of this research have already been addressed as subsequent studies have built on Shochet's findings as discussed at length in Chapter Four.**

5.2.2 Results of the Hoffenberg (1988) study

The Hoffenberg study (as with the Gaydon [1988] study - see below) formed part of the early studies conducted within the field, and had as its focus gifted children from disadvantaged learning environments. This sample, then, in this study is not the usual type of sample featured in later South African dynamic research. Hoffenberg's results are indeed disappointing and do not aid the case for dynamic assessment in South Africa, in particular for gifted learners. The negative findings must be viewed within this specific context.

A control and an experimental group were used in the study, the experimental group receiving mediation on four sub-tests of Feuerstein's Learning Potential Assessment Device (LPAD) and the control group, who were not exposed to mediation on the LPAD tasks. No significant differences were found between the two groups, bar on one sub-test. It was expected that because of a lack of previous learning experiences, the mediation given during the LPAD tasks would compensate for this.

Two hypotheses were:

- The group receiving mediation on the LPAD tasks would perform significantly better on these tasks
- Transfer measures would be significantly higher for the experimental group as opposed to the control group.

Although there was a tendency towards higher scores for the experimental group, only one sub-test was found to be significantly so. Mediation thus seems to have had the effect of some improvement but this may also be due to chance. No differences were found between the two groups on the transfer measures either, although higher scores were evident for the experimental group, but not significantly so.

Thus neither of the two hypotheses was fully supported. These poor results led Hoffenberg to conclude that making use of the LPAD as a screening device in identifying gifted learners should not be used. Mediation, then, did not appear to have had any effect at all. Reasons cited for these results are varied, foremost among them is the poor mediation administered (the same reason as cited by Gaydon), and a possible Hawthorn effect, (where improvement from pre- to posttest scores is inevitable even though she made use of a Solomon Four Group Design to negate any effects of pre-testing). Among the control group, there was a tendency for the pre-tested half of the sample to perform less well than the unpre-tested half, and this may have been due to tedium. Different task demands were called for in some of the transfer tasks as opposed to the LPAD tasks, which may been counterproductive, making proper measurement of transfer scores more difficult and less satisfactory. It would seem that a few technical and methodological aspects may have been disruptive in the Hoffenberg study, aspects that might well have had less effect if done away with or modified in another similar study. Too many factors militating against the study seem to have been partially responsible for the poor results. **Although there were tendencies for the mediated group to improve after mediation, in essence, the results of this study are not encouraging in terms of using the LPAD in the context of identifying disadvantaged gifted learners.**

5.2.3 Results of the Gaydon (1988) study

Gaydon's study is one of many such studies to emanate from the University of the Witwatersrand during the late 1980s. Unlike other studies, though, her focus here was on gifted students who found themselves in disadvantaged circumstances. She draws on results obtained in the same setting by Hoffenberg (1988) (see above). One's attention is drawn here to the fact that "disadvantage" does not mean "below average". Gifted students may not necessarily be able to effectively utilize their gifts if the opportunities are never afforded them. The aim of the study was to ascertain the effectiveness of both traditional and dynamic assessment tests in determining prediction of success within a programme run for gifted learners. Which was the more effective predictor? It was hypothesised that dynamic assessment results would enhance prediction.

Feuerstein's Learning Potential Assessment Device was employed with mediation for an experimental group and administered to a control group without mediation. Along with other static measures, three subtests from LPAD were included. These measures were correlated with the criterion measure (performance on the programme as indicated by scores in various tests). To account for any "testing" effects, Gaydon also made use of the Solomon Four Group Design. Unlike other studies investigating disadvantaged students, school results did play a role in predicting success on the gifted child programme along with sub-tests of the LPAD. This is not surprising, as these learners are academically superior to students in their group.

Sub-tests of the LPAD accounted for a significant amount of the variance for both pre-tested and unpretested samples. Transfer scores are useful predictors for this particular programme. Even though the findings were positive in terms of prediction being enhanced by the inclusion of LPAD sub-tests, a recommendation that mediation skills be improved was noted, in the hope that better mediation might yield better posttest scores and thus be better predictors for success in the programme. Prediction is enhanced by the learning phase allowed during the LPAD administration. However, along with certain tests and LPAD subtests, only 50% of the variance on the criterion score was accounted for, which led Gaydon to speculate as to whether the LPAD is indeed a useful means of improving manifest performance of disadvantaged children. The LPAD did serve to identify potentially gifted students that may otherwise have been left out of the programme.

The study was inconclusive about whether mediation was indeed effective, as no relationship was found between modifiability and results on the criterion. Gaydon says that the mediation administered was of poor quality and that it was administered to a group and not individually (this was done to combat time and costs though). Results are slightly more positive than those found in the Hoffenberg (1988) study (see above), yet neither, it would seem, are "glowing" in terms of advocacy of LPAD use. **Nevertheless, exposure to training (via mediation) within testing sessions is advised based on the findings, noting that such exposure to learning allows for better transfer scores, from pre- to posttests results.**

5.2.4 Results of the Murray (1988) study

Findings in this study come as no surprise when one reviews other studies of a similar nature. However, at the time at which the study was conducted in South Africa, the results may have been construed as valid and significant. In essence, the results showed that learners (from educationally and socio-politically disadvantaged backgrounds) yet from a wide spectrum of academic achievement, benefitted from mediation after having been administered the LPAD. However, academically superior learners showed greater improvement than did less academically superior learners. Murray indicates that this goes some way towards illustrating the fact that inherent intellectual ability is positively related to potential.

The better one's starting position the greater the likelihood of later success.[2] Inherent ability and academic ability are not the same construct.

The study by Murray analysed responses on the LPAD qualitatively. It was seen that after mediation was offered, learners from both academically high and low groups were able to benefit from mediation. Although there were a few aspects to the mediation that needed to be looked at (such as language of mediation and group administration of the LPAD), **the study provided further impetus for the recognition and use of dynamic assessment in South Africa.**

5.2.5 Results of the Boeyens (1989a) study

This theoretical study by Boeyens in 1989 can be construed as one of the forerunners of dynamic assessment investigations in South Africa. Apart from the Shochet (1986), Gaydon (1988), Hoffenberg (1988), and Murray (1988) studies as well as the beginnings of the TTT programme at the University of Natal, not many other studies in this field can be found. In fact, it is always revealing to scour the reference lists of researchers in this field at this early stage in dynamic assessments' development in South Africa to see who was aware of other researchers' work in the field.

In Boeyens' (1989a) study, the work of Shochet (1986) is acknowledged but not that of Hoffenberg (1988), Gaydon (1988) and Murray (1988). However, retrospectively, this study offers insight into how the field was progressing early on. The theoretical study of Boeyens does not lend itself to the usual research design but serves as a peremptory study for the empirical results to come later. Nevertheless, it is deemed important as an early start in the field and warrants a conclusion in this section. Boeyens investigates the nature of possible dynamic assessment tests and what such assessments may constitute. He based his ideas of what should be included in an assessment on the fact that the tests should be contentless, that learning should take place during the assessment and throughout learning during the test, and that Vygotsky's zone of proximal development should be incorporated as not only a theoretical concept but as a practical tool. The extent to which the learner is able to benefit from the test instruction and proceed with the tasks is an indication of his/her learning potential and further mental development.

Boeyens makes use of the letter series approach in the development of his learning potential test. Of course, this was not an original idea but was taken from other researchers. Various relationships exist within the letter series, relationships which have to be solved by the learner in order to progress to more difficult letter series. **The theoretical underpinnings of this study paved the way for the practical aspects of the study to follow.**

5.2.6 Results of the Boeyens (1989b) study

Vygotsky's notion of the zone of proximal development is affirmed in this study, in which learning potential (the difference between pre- and posttest scores) effectively predicts the degree to which learners will benefit from mediation or instruction. Three groups of learners emerge in this study, foremost among them were those who successfully integrated the lesson and made flexible use of the information received during instruction. Learning potential scores are good predictors for these students who benefitted from an enriched learning situation. A second group did not integrate the instruction administered and hence their learning potential scores yielded no predictive validity, but their ability tests did yield predictive validity. A third group, perhaps the most interesting from a

2 Such a finding can be confusing especially when one reads of results which show that difference scores for academically inferior learners are often greater than difference scores for academically superior learners. The question posed is how one puts the two together. High pre-test scores will yield lower difference scores due to a ceiling effect (see De Beer [2000]), but these same learners can benefit more from mediation than low pre-test scorers by virtue of their inherent capacity. The point is this: inferior intellectual ability is more often than not measured by academic tests. This is why "inferior" learners are seen to improve so well on post-test scores, more so than their "superior" academic counterparts. One must not confuse inherent ability with academic ability as this is where the confusion lies.

theoretical point of view, showed that by applying the rules of the instruction to the tasks in a very rigid manner, lower scores were obtained as they were not flexible in their approach to items. For this group, traditional ability tests correlated more so with academic success than did the learning potential scores. Boeyens equates this rigidity with a style of learning within school, where rote learning was at the time emphasised. Although this group applied what they had learned, they did so in a very meticulous fashion, applying rules without ascertaining the usefulness of these rules for each specific situation and thus making inappropriate choices with regards to which rules to implement.

The findings of this study, are positive in that they make the case for learning potential (dynamic assessment) strongly. **Boeyens' main contribution to the field, in this study, was the identification of three types of learners who may or may not benefit from learning potential tests and by highlighting the fact that not all learners necessarily benefit from such assessments. Hence, learning potential testing is not necessarily the answer to problems identified by some static tests. Learning potential testing should be utilised according to specific protocol and specific targets one has in mind such as learners who will most likely benefit from such mediation.**

5.2.7 Results of the Henley (1989) study

The aim of this study was to investigate the contention that Feuerstein's theory of mediated learning experience is an underlying component of cognitive modifiability. The experimental study showed that children who had received mediated learning experience (by way of child-adult interactions) were better equipped to score favourably in academic tasks but learners who received little mediated learning experience performed similarly to the high mediated group in the LPAD tasks. The transfer scores for the high mediated group were also higher than for the low mediated group. This "delay" in performance regarding the high mediated group (since by definition these learners should have outperformed the low mediated group) is explained by the fact that the high mediated group showed positive effects of mediation only after a brief period of time. The study also showed that even though children who do not receive early child-adult mediation may not perform as well as their high mediated counterparts; cognitive modifiability is still within their grasp (as cognitive modifiability is evidenced in adolescence[3]). In other words, both groups were able to benefit from a mediated learning experience, just not to the same degree. Distal factors such as the socio-economic status of learners are less influential than the more proximal factors such as having received mediated learning experience at an early age, on cognitive tasks. Early mediated learning experience was determined on the basis of a questionnaire designed by the researcher and must be interpreted with caution, as there is no way to support or refute the fact that the questionnaire is measuring the construct of mediated learning experience. **The study provides support for Feuerstein's notion of mediated learning experience as an underlying component to cognitive modifiability, but it also showed that other less fortunate learners in terms of educational disadvantage can benefit from mediated instruction.**

5.2.8 Results of the Zolezzi (1992) study

Although two of the three stated hypotheses were supported by a moderator effect and not a main effect, the study's results are encouraging. The fact that the sample consisted of mainly disadvantaged students may have impinged slightly on the validity of the results, not to mention the small sample size (N = 26).

The study shows that learning potential is a better predictor of academic success for disadvantaged students than either traditional measures of general intelligence as well as school marks, but that (having discovered this through a moderator effect) the degree of modifiability plays a role in the extent to which learning potential is a predictor of academic success. In other words, the less

3 Refer to section 2.4 on the physiological plasticity of the brain - plasticity that may be evidenced in adulthood as well as in younger people.

modifiable disadvantaged students were more predicable on traditional measures than the more modifiable disadvantaged students. Also, the fact that advantaged students predicted entirely differently as opposed to disadvantaged students indicates that learning potential assessment is better able to predict for disadvantaged students. This lends support to the idea of making use of learning potential assessment for disadvantaged learners. The third hypothesis was supported by the findings, indicating that the learning process assessment, when used with learning potential, is a better predictor for both groups of learners than traditional measures used alone. This also supports similar findings, where learning potential assessments are used in a complementary fashion rather than when used exclusively.

Zolezzi states that the measures used in this study go beyond prediction and take on a new paradigm, namely, that an educational-modifiable approach should be used and not merely product-based assessment approaches. Another important finding and one that concurs with the findings of Boeyens (1989b) (who differentiated between three different groups of learners) is that the term "disadvantaged" cannot be used as a blanket term. Not all disadvantaged learners are modifiable. There are disadvantaged students who would predict better with static measures than with dynamic measures. In other words, one must select those disadvantaged students who would most likely benefit from dynamic assessment first before administering such assessments. **Dynamic assessment is better able to predict for modifiable disadvantaged learners. The results are thus encouraging.**

5.2.9 Results of the Lipson (1992) study

In this study, the GSAT and LPAD were assessed for their respective predictabilities regarding learners' academic success in certain school subjects. Although meditation improved posttest scores on the LPAD, the author asserts that this cannot unequivocally be stated as being the sole reason for improvement, seeing as there was no control group in the study which may have been able to throw light on the results, if this control group did not receive mediation (the practice effect may have contributed to the improved results on the posttest assessments, but this cannot be proved or disproved). The small sample size (constraining the types of statistical procedures that could be performed) and the fact that the sample did not fit the normal distribution of the population on which the GSAT was developed may have skewed results to a degree (as their results could not be scaled or converted to a measure of intelligence). The GSAT was found to be a better predictor of academic success in pre- and posttest assessments as opposed to the LPAD, a possible reason being that the criterion variable and GSAT measured similar constructs, underlying cognitive processes which the LPAD did not measure. However, certain sub-tests of the LPAD did correlate with certain school subjects.

Lipson states that the criterion measure should be a more broad-based one, encompassing aspects other than those emphasising school subject learning and cognitive skill. Although the findings would appear disheartening, Lipson believes that the LPAD and any criterion according to which one is going to measure the success (or not) of the LPAD should be more processed-orientated and less product-orientated than the present curriculum evidences. However, this means a total restructuring of the criterion variable in order to suit the nature of the research, which is not really the aim of the study in the first place. Also of importance is the fact that the sample cannot be said to be disadvantaged. In fact, the sample came from a private school (typically these students are more advantaged) and it has already been seen that advantaged learners do not predict as well on the LPAD as do disadvantaged learners. In light of the above information, the results of this study are relatively predictable. **The conclusion is that the GSAT (a static measure) is a better predictor of academic success than the LPAD. However, for reasons mentioned above, the LPAD cannot be said to be a useless predictor either.**

5.2.10 Results of the Zolezzi (1995) study

In this study, Zolezzi was successful at operationalising Vygotskian and Feuersteinian theory into learning potential by devising a test battery. It must be noted that the battery consisted of static measures that were merely administered in a dynamic manner and were not in themselves dynamic

tests. The dynamic assessment battery significantly enhanced prediction of academic success. It was also shown that both advantaged and disadvantaged groups have different predictors of academic success. Most importantly for the study, evidence yielded that there was no relationship between current level of ability (as assessed by static measures) and learning potential. Standardisation was important to the study as this had a direct bearing on reliability as well as validity. Mediation was thus the same for all learners. Zolezzi seriously considered time and costs, hence his choice of static tests which were chosen on the basis of the time it took to administer and record results. The study was able to show that one can take cost and time into account along with standardisation of mediation in a group-administered format and evidence successful enhancement of academic success, whilst concurrently differentiating between advantaged and disadvantaged groups. Also, more importantly is the fact that the newly devised battery is normed on university students, a population for whom it is applicable. The sample size was small and lack of individual mediation (which is essentially the goal of mediated learning experience), compromised the study somewhat. **Nevertheless the results are positive and encouraging.**

5.2.11 Results of the Andrews (1996) study

The Andrews study sought to determine whether or not the group administration of the LPAD would lead to modified cognition as well as to determine whether administering the LPAD in a group format would in any way lead to a detection of individual differences in the degree to which each learner was modified. A matched experimental and control group was selected after 21 students were administered the LPAD pre -test.

Of the three LPAD sub-tests, two evidenced improved results from the experimental group during posttest administration but this was not evident for the control group. One sub-test did not yield the same pattern of results due to problems associated with the mediation given and the nature of the sub-test. The experimental group was itself divided as to the quality of improvement which followed mediation. There were learners who improved significantly, those who improved but not significantly and those who, in fact, performed less well in the posttest than in the pre-test. This points to the discerning nature (or discriminatory ability) of the LPAD to identify degrees of modifiability in different learners. For two LPAD sub-tests, group administration with mediation did result in improved scores for learners and the results on the posttest scores enabled the researcher to conclude that the group administration was indeed able to differentiate between different levels of modifiability. A third sub-test was unable to provide this information, as the posttest administration of the test was a more difficult version of the one offered in the pre-test session. This confounded the results to a degree and lent only limited support for the above two questions.

Factors such as language, layout of test tasks, lack of further qualitative analyses, confusion of certain LPAD terms and feedback that did not take place as a result of the administration were aspects that Andrews considers affected the study in a negative manner. **Nevertheless, the findings of the study are largely positive, and as Andrews cautions the reader, dynamic assessment will more fully adapt to the local teaching situation in South Africa, if teaching itself were to become more process-orientated and less reliant on rote-type of learning. Costs are, in his opinion, still inhibitory.**

5.2.12 Results of the Tayler (1996) study

The aim of the Tayler study was to determine the predictive validity of the Learning Ability Battery (LAB), a test which is classified by Tayler as a learning potential test but not as a **true** test of learning potential. Here one can clearly see that the definition of dynamic assessment is based on a continuum. Tayler states that the LAB can be defined as a learning potential test based on four criteria that she stipulates, but that it is not entirely a learning potential test.

Most notable was the acknowledgement of the test's lack of pre- and posttest scoring system (which leads it away from the dynamic approach) and also the area of application, the test being used in industry and not in the field of tertiary admissions and selections as with most other dynamic assessment instruments. The LAB was found to be a valid and reliable predictor of groupings in which

employees are placed initially in order to undergo further training. Based on performance on the LAB, employees were placed in groups characterised by the degree to which lower educated people would benefit from further education.

The LAB is one of the few locally developed and normed tests, thus not relying on overseas tests that are not normed on local populations. This makes the use of this battery all the more appealing. It nevertheless retains many of the "types" of questions that are included in many other available tests. The fact that employees were given practice examples before having to complete the rest of the questions in the battery, served to equalise the chances of all who participated; and emphasising a process-based approach as opposed to a product-based approach and allowing a two-way communication process between learner and examiner means that the LAB can be defined as partly dynamic in nature. **The findings are positive in that the LAB was found to be a valid predictor of further employee training, although certain items in the test will have to be looked at again. It must be emphasised too that a dynamic approach was used in the study. It is this aspect that allows for the study's results to be included in this survey.**

5.2.13 Results of the Van Aswegen (1997) study

This study investigated the validity and usefulness of the Transfer, Automatisation, Memory and Understanding Learning Potential Battery (TRAM-1) to predict future performance of poorly qualified workers in an adult education programme. Once again, as with the study conducted by Taylor, this battery's field of application lies strongly within industry. This is further evidence that dynamic assessment *per se* is not limited to certain areas of application within tertiary education.

Six dimensions of cognitive functioning are measured by the TRAM-1 and the Van Aswegen results show that five of them are effective predictors of learning performance amongst this population. The total TRAM-1 score and the total criterion score (variables that were used to validate the findings of the TRAM-1 such as college subject results in the adult education programme) were the most highly correlated of all variables. Individual correlations, although still significant, varied in their strength of correlation. All subjects included in the adult education programme correlated with the TRAM-1 results, but one sub-test of the TRAM-1 did not correlate with the criterion data and, based on this, it was suggested that it be removed. The TRAM-1 is a dynamic assessment battery with learning taking place throughout testing, but this learning is standardised and improvement scores are calculated on each dimension in the test. The test tasks are themselves contentless, so as to allow everyone the same chances of success. **The results of this study, although bearing on a specific population within a specific context, are positive. The TRAM-1 test battery as an example of a dynamic assessment test proves to be valid and reliable**.

5.2.14 Results of the Nel (1997) study

The main finding in the Nel study was the fact that the Ability and Processing of Information and Learning Test Battery (APIL) when used together with matriculation examination scores, language proficiency sub-tests as well as the GSAT, explain more variance than when any one test battery was used on its own. There was a high correlation between the GAST and the APIL, indicating that the two batteries were measuring the same construct. The fact that they tended to measure similar constructs led Nel to conclude that if given a choice between the APIL and the GSAT, the latter would be chosen for to time and cost effectiveness. The main recommendation given in the study is that the APIL should be used when testing students who are educationally and culturally disadvantaged as the APIL is able to add valuable information that would otherwise remain unknown.

According to discriminant analyses conducted on the data, all four predictors, APIL, GSAT, M-scores and language proficiency sub-tests accurately classified learners into two groups: those who would be successful and those who would most likely be unsuccessful in tertiary education. These tests were not able to differentiate learners into three groupings, namely those who failed or passed and those who would not get access to write the final examinations. The APIL, forming part of a larger test battery, was investigated, but a main emphasis in the study was to ascertain the usefulness of the test battery as a whole as well as to investigate the APIL. No single predictor was able to predict any

student's success at further education accurately. Only when used in conjunction with other tests could the variance explained increase. Nel notes the importance of non-cognitive factors as well. This supports the contention that dynamic assessment, although a very promising alternative to static testing, cannot be used as a sole alternative. An issue which does need to be taken seriously, however, is the fact that if, as Nel states, the APIL and the GSAT measure similar constructs,[4] can this be taken to mean that the APIL is not as culture-fair as it is professed to be? Alternatively, the same constructs may be measured, but it can also be said that the APIL does so in a dynamic manner. This point, perhaps, needs further clarification.

Another reason why the GSAT might be chosen over and above the APIL if the choice did come down to choosing one test is that the APIL is intended for testing of small groups at any one time. This clearly becomes a problem with regard to selection and admissions tests. The method of scoring the APIL is vulnerable to error, as it is scored by hand, unlike the computerised version of the GSAT. Moreover the APIL is a lengthy test to administer and fatigue can play a negative role. In sum, the GSAT is the preferred test to administer, but if deemed necessary (on the basis that the GSAT scores are not accepted for whatever reason) the APIL should be administered in order to allow the candidate a second chance at selection. If an institution wishes to follow the culture-fair testing route, it is advisable (according to Nel) to administer the APIL first. However, using both tests together does not yield any more information than when the tests are used separately; but when used with other predictors (as in this study) more information will be available from both than when only one test is used. **The results of this study are not negative in any way, but do provide constructive criticism for the APIL, a critique that is perhaps best dealt with in further studies. Only in this way can dynamic assessment progress; and the study provides the impetus for further evaluation of the APIL.**

5.2.15 Results of the Gewer (1998) study

In most assessments of non-verbal reasoning ability, which are more often than not qualitative in nature, the likelihood of disagreement is always a factor, and in this study, inter-rater reliability did become slightly cumbersome as an issue. Non-verbal reasoning of educationally disadvantaged students was investigated by means of a dynamic assessment approach. The quantitative study design enabled the researchers to ascertain the effectiveness of mediation that was given to an experimental group as opposed to a control group which did not receive any mediation on the same tasks.

The non-verbal tests, the Raven's Coloured Progressive Matrices (RCPM) as well as the Rey-Osterrieth Complex Figure Test (ROCFT) were administered to both groups, and this served as a pre-test score. After mediation (given only to the experimental group) the subjects were again tested, using the Set Variations 1 from the LPAD (seeing as this sub-test was similar to the RCPM and could serve as a transfer test). The ROCFT was administered, which served as a pre-test score, with mediation given to the experimental group, and the posttest score was ascertained from results from a sub-test of the LPAD which was similar to the ROCFT. This then served as a transfer test.

An issue that comes to the fore later in this survey is that this study's findings were hampered by untrained test administrators. A qualitative method of analysing the RCPM and the ROCFT was used in order to add value to the quantitative results. However, student testers were not familiar with the matrices, to such an extent that that inter-rater reliability was non-significant. There was thus little agreement between raters as to the responses from the learners. There were improvements in the experimental groups after mediation was given, but no such changes were evident for the control group (for whom the tests were static indicators of present functioning and not potential). Yet the results were not significant for the overall scores on the RCPM. The results for the ROCFT are similar, but for the ROCFT's transfer test there was no statistical procedure available in order to investigate whether the changes were significant or not, as this was a qualitative test. For this to be

4 For certain study directions, it was found that the GSAT and APIL both correlated with academic success in the Engelbrecht (1999) study.

investigated, the qualitative results had to be looked at. The qualitative method was based on a known qualitative scoring method.

Gewer places heavy emphasis on the qualitative findings of the results, as he states that information yielded in this manner offers new information not reflected in the quantitative results. In-depth analyses of the qualitative findings were more insightful than those offered by the quantitative results. This supports the notion that dynamic assessment is more process-based than product-based. Findings such as the experimental group who were less impulsive, more motivated, and who demonstrated better visual-spatial orientation are the results of qualitative analyses. This information can be used to relate back to areas of teaching that need to be looked at, as well as learners' cognitive functions (abilities) which are/are not performing as they should. Gewer states that this would provide further guidelines to intervention processes. The fact that the learners' first language was not used, that the time taken to complete the study was lengthy, and that inter-rater reliability was poor, seems to hamper the study to a degree. **However, the findings are positive and provide information on not only quantitative results of a study using dynamic assessment, but provide qualitative findings as well. Most studies of this nature more often than not concentrate on the quantitative results, and thus miss out on the potential for the diagnostic and remedial aspects of qualitative interpretations of results.**

5.2.16 Results of the Engelbrecht (1999) study

In an attempt to identify the usefulness of the Ability, Processing of Information and Learning Battery (APIL) as a predictor of academic success among first year university students at two universities, as part of a larger set of tests, Engelbrecht concluded that the APIL served no use in terms of identifying learning potential among the students. Although the results were not negative (in terms of a total lack of correlation), the APIL was unable to predict for academic success for the entire sample. However, it must be noted that certain sub-tests of the APIL were able to predict for certain study fields/directions. The APIL was not used in its entirety and a shortened version was used, consisting of only three sub-tests of the total available eight sub-tests. Some findings were at odds with findings from the original validation studies conducted on the APIL (Taylor, 1997 as cited in Engelbrecht, 1999), which may be important for future use. A main aim, that was not achieved in the study, was the development of APIL norms for first year university students. Based on the poor results, however, this could not achieved.

The APIL along with the GSAT and SAT was not able to yield any more information than either of the other tests when used alone, but was able to yield more information when used jointly with other tests. Hence, there is a use for the APIL in further refining the choice of student selections. It was also found that the APIL and GSAT[5] did measure many of the same constructs which, as Engelbrecht points out, may be misleading to those who interpret this finding with scepticism, as the APIL is intended to be culture-fair and assess learning potential. A possible counter-argument is that the APIL may well measure similar constructs but it does so in a dynamic manner. Once again, the dynamic administration of the second sub-test in the APIL is the only sub-test to be administered dynamically in this study.

The results for this study are not encouraging for the use of the learning potential APIL battery. However, other APIL findings in this survey yielded positive results. Lastly and, more importantly, is that although the results were not positive, this is but one test within the dynamic paradigm and negative results in one study are not enough evidence against dynamic assessment to discredit it altogether.

5 Nel (1997) section 4.3.14.3 also found similarities in the GSAT and APIL. Taylor also found strong correlations with the Mental Alertness Test and the APIL as well as academic performance and the APIL (1999).

5.2.17 Results of the Coosner (1999) study

As with the study conducted by Gewer (1998) who used qualitative methods as well as quantitative methods of analyses, Coosner's study is one of the few studies surveyed which make use of qualitative methods and analyses. She states that a process-based manner of assessment will complement the Outcomes Based Education system (OBE) used currently, and that both OBE and dynamic assessment share many characteristic features, such as emphasising learner potential, redressing the examiner-learner relationship and emphasising understanding more than rote learning. Hence, dynamic assessment is suited to an OBE initiative.[6]

The aim of the study was to develop a pilot dynamic assessment strategy which had as its main emphasis the theoretical underpinnings of Feuerstein's Mediated Learning Experience (MLE), as well as his Structural Cognitive Modifiability (SCM) theories. Application and implementation in the school classroom in the context of the South African schooling system were perhaps the most notable aspects of the research. The more locally developed tests that are available, the better. The programme is said to last for approximately four hours.

The fact that this pilot study (leading, it is hoped to a published one) was developed locally adds credence to its reliability and validity within the South African context. The results of the study show that teachers who received training on the new strategy found it to be practical and implementable, with a few considerations from teachers being valuable in making the programme even more understandable and applicable. Not only was this finding positive, but the teachers themselves who had undergone the training felt that they had benefitted, in terms of having felt more learner-centred and self-reflective. Changes in teacher perception are also a major milestone. After training, some teachers remarked that they now held different views of certain learners in their classrooms. A more positive stance and attitude can only benefit both teachers and learners alike.

Being taught the theory of MLE and SCM, teachers were better equipped to deal with certain learner problems, which prior to training, were assumed unalterable. The programme also had the effect that negative attitudes towards education could be changed and that such a strategy was not only valuable for educationally disadvantaged learners but also for immigrants and those speaking different languages. The dynamic approach which characterised this strategy also made it possible for teachers to pay individual attention to learners and in so doing offer each learner more valuable aid. Teachers were also shown that some learners who do not perform well on static measures or in a static environment did, in fact, show learning potential, potential that would otherwise have gone unnoticed. Once again, states Coosner, this dynamic strategy is best utilised in combination with other static measurements. **The findings of the study are very encouraging even though there are no quantitative results available in order to verify the findings of subjective teacher comments scientifically (this being the nature of a qualitative study). Coosner states though, that this is a major issue plaguing the field of dynamic assessment: how to ensure reliability and validity in an approach which de-emphasises such aspects.**

5.2.18 Results of the De Villiers (1999) study

The results of this study proved valuable and were on the whole positive. De Villiers, by reason of the results, is able to conclude that although dynamic assessment may be a more costly method of assessment, Vygotsky's theory of proximal development (ZPD) can be successfully operationalised and implemented within dynamic testing in South Africa. Even though matriculation examination marks were the overall best predictor of academic success for the disadvantaged group of tertiary students, measures of ZPD were able to yield more information about less modifiable and more

6 Of interest is that in no other study is OBE mentioned. A common critique of why dynamic assessment is not used or why it cannot fulfil its full potential as an approach is that its process-based manner is not suited to a product-based education system. However, as Coosner states, OBE itself is process-based. Perhaps the time is therefore ripe for implementing more dynamic assessment procedures.

modifiable students. Learner modifiability, then, acted as a moderator variable, when investigating predictive power of various variables.

De Villiers' results also show that two learning potential tests the Conceptual Reasoning Test[7] (CRT) and the Learning Potential Test (LPT), when administered with mediation to an experimental group - yielded higher results in the posttest than did the results of the control group who did not receive mediation. Response to mediation is a useful tool in measuring accuracy of valid prediction. In other words, not only do dynamic assessment tests yield greater predictive validity for the disadvantaged group, but by investigating the responses towards mediation as well, further accuracy can be assigned to the prediction of tertiary success. This is an important element, as "disadvantaged students" are not an homogeneous group, and not all disadvantaged students necessarily possess the potential to succeed. In this manner, one is not only allowing for a less discriminatory procedure of assessment, but also encouraging accuracy of selection. In other words, one can ensure that those who are seen to possess potential are selected and so this can increase the odds of success of further studies. This is a better alternative than the quota system that has been mentioned, in which a certain number of students from groups are selected according to their group affiliation. Using the quota system in this manner would mean that many students would gain access based solely on their disadvantaged status and not learning potential status.

These are positive findings but costs and timing are still an issue. One must consider though, that the costs of selecting high-risk students might be higher at the end of the day, if they do not succeed and/or drop-out. Initial outlay of costs may be a more prudent course to pursue. It must also be borne in mind that dynamic assessment (as is evidenced by De Beer [2000] for instance) need not be an impediment when it comes to costs. Vygotsky's ZPD was operationalised through the use of dynamic testing and in so doing yielded important information for educators at tertiary institutions. **Dynamic assessment should not only be able to offer an alternative towards testing, but must also indicate to what extent each learner is indeed modifiable and hence provide a profile as to their low-risk and high-risk status. This study successfully shows that dynamic assessment can do exactly that**.

5.2.19 Results of the De Beer (2000) study

The study by De Beer is indeed an excellent example of how effectively a locally developed and normed dynamic test can be implemented in a variety of contexts. Upholding most tenets of dynamic assessment, such as reliance on non-verbal contentless tasks, culture-fairness, mediation, as well as pre- and posttest assessment (both the pre- and posttest forms of the Learning Potential Computer Adaptive Test [LPCAT] are two separate adaptive tests) in order to yield difference scores (by taking into consideration performance prior to and after mediation as well as the difference score itself), the LPCAT is also psychometrically defensible in terms of reliability and validity. A thorough validation study was conducted on several multicultural groups, and normed and tested on a large sample base. What makes this study so notable, is the fact that a dynamic test was developed and did not, as in many studies do, rely on already existing static tests and alter their mode of administration by including an element of mediation and providing pre- and posttest assessment scores.

De Beer has shown that dynamic assessment does not necessarily incur unnecessary costs, does not have to take up an undue amount of time and can, in fact, be administered in a group format. Most important of all is the use of a computer which aids tremendously in administration. It is owing to this technological advance in dynamic testing that the above-mentioned aspects no longer play such a large inhibitory factor in dynamic assessment. Item response theory (IRT) and computer assisted testing (CAT) are invaluable tools in this new dynamic assessment battery. There is minimal human intervention (notwithstanding the fact that two versions are available, one in which mediators aid testers who are not literate enough to undergo assessment), individual testing is precluded by the group administration (yet it is individually scored and individually attuned), but remediation is not an emphasised element in this assessment battery. A compromise has been reached in terms of

7 Considered a dynamic test because of the hints and lesson (mediation) used in the test.

dynamic assessment tenets on the one hand, and timing and costs, on the other. It has been shown throughout this study, that these two issues (pure dynamic assessment and the efficiency implications) are difficult to bridge, if at all possible.

The LPCAT is not only applicable to disadvantaged persons (although this is the main target population), but to anyone at any particular level of performance. Measurement accuracy and testing efficiency, then, are two major obstacles in dynamic assessment research. This recently published research has not been referred to thus far in much South African research owing to its recent addition to the published literature in this field. The LPCAT's theoretical underpinnings are firmly grounded in Vygotskian theory and, in fact, go as far back as the pioneering theory of Binet (an intelligence researcher in the early twentieth century) who is not frequently mentioned with regard to dynamic assessment research in South Africa. Binet's emphasis on development is also an underlying principle of importance in the development of the LPCAT. **The introduction of a locally developed and normed learning potential test, that rests squarely within the domain of dynamic assessment and is at the same time psychometrically sound and suitable for South African's need for equitable testing, is a positive move in this field in South Africa and may well be a leading example in other countries with similar circumstances.**

5.2.20 Results of other South African research projects

Lloyd and Pidgeon (1961), very early on, reveal the usefulness of a dynamic manner of testing by implementing a form of coaching, and although no mention is made of the term "dynamic", the study evidenced results which are similar to results that were to come much later in the last two decades of the twentieth century in South Africa, even though the researchers themselves made no mention of the dynamic nature of their design. **The dynamic methodology has at this early stage been recognised for its usefulness.**

Adams (1994) emphasises the need to look more closely at mediated learning experience and also highlights the need for teachers to become more informed of the potential of identifying zones of proximal development in learners. Hence a dynamic approach towards education is espoused largely in terms of remediation. **Dynamic approaches towards education will, then, for obvious reasons, go hand-in-hand with dynamic assessment.**

De Villiers (1996) reiterates this notion of implementing and integrating the zone of proximal development in South African tertiary education. Rules (ways of studying at tertiary institutions as opposed to high schools) should be externalised so that disadvantaged students, in particular, are able to internalise them after a process of mediation from more competent peers and teachers (lecturers) has taken place. **Mediation is thus encouraged as a method or approach when aiding disadvantaged students in terms of studying.**

Lopes, Roodt and Mauer (2001) conclude that the APIL learning potential battery is indeed a reliable instrument when applied within a specific context. **Noting one or two factors which may need to be assessed for future use of the battery, the findings are encouraging for the further use of this test instrument.**

Taylor (1999) made use of the Ability, Processing of Information and Learning Potential Battery (APIL) as well as the Transfer, Automatisation, Memory and Understanding Learning Potential Battery (TRAM) test batteries as part of a greater initiative, known as the CogLab programme, which aims to place employees in various first level managerial positions; and, in order to ascertain their potential for success within the CogLab programme, the APIL and/or TRAM is administered. Those achieving higher results are considered more likely to succeed in the programme. Here, a cognitive approach informs both static and dynamic approaches towards assessment. Dynamic assessment is applicable in a wide variety of contexts where past disadvantaged schooling has resulted in a lack of skills training making it difficult for potential learners as well as employees to achieve their potential. **The APIL and TRAM assessment instruments coupled with the CogLab programme seek to address these issues and have done so with relative success.**

Nunns and Ortlepp (1994) show that, with a small sample, the Conceptual Reasoning Test (CRT), when used as an instrument with which to assess potential, achieves positive results. **The CRT is, they state, able to identify students who possess the potential to succeed within accelerated conditions and do so without reliance on prior schooling.[8] This can be seen to encourage further use of the CRT as dynamic test instrument.**

The fact that modifiability, when used on its own as a measure of potential, does not predict for academic success is perhaps the most crucial conclusion made by Shochet (1994). If the field of dynamic assessment is to move forward and become more accepted within mainstream testing, not only should it withstand scientific rigours, it should be utilised in combination with other static and non-cognitive assessments such as educational experience and background. Non-cognitive factors have been given more status in the past decade and just as static tests, non-cognitive assessments and dynamic assessments do not yield the best results when used separately, they must do so when used together. **Dynamic assessment, then, is only one aspect of assessment, an aspect that has become more important but one that cannot work without other more traditional manners of assessment.**

The field of dynamic assessment is irrevocably intertwined with the work, theories, thoughts and philosophies of Vygotsky and Feuerstein and no discourse on dynamic assessment can be complete without copious referrals to these two eminent researchers' work. Moll's (1989) article as well as Bradbury and Zingel's[9] (1998) article captures the essence of what each theorist has to offer the practical teaching world in terms of dynamic assessment theory. **Theoretical discourse and discussion are just as imperative to this field as to any field in psychology and argumentation and debate can only but spur this field on to greater and more widespread achievement. Theory and practise inform one another, and to date, most studies have concentrated on practise but practise needs to be informed by theory, so that theory can enhance practise.**

Sibaya, Hlongwane and Makunga (1996) conclude that owing to the almost complete lack of assessment instruments that have been normed on populations other than white samples, dynamic assessment offers a potential for assessment, within a paradigm which espouses tenets that, on the surface, seem to answer many of the nagging issues which psychologists find themselves asking. If assessment instruments are to be used in tertiary settings mainly for the purposes of selection and admissions of previously disadvantaged students, then perhaps it is a wise idea to do away with biased tests. (Many are still being used, many of which have not been normed on this population.) This study recommends the introduction of non-biased tests, which include dynamic assessment. The authors view dynamic assessment as a very good starting point, seeing as this method already adheres to many tenets and conditions set forth by new legislation in South Africa on testing. **Being mindful of the pitfalls of this approach, the authors state that dynamic assessment is by far the best option now available.**

Miller's (1998) study focuses on cognitive change through learning and states that, in order to assess a foundation programme's utility, it would do best to understand that groups of disadvantaged learners (just as is the case with advantaged learners) cannot be classed into a homogeneous group. Practitioners must understand that low performing students might not necessarily be changed at all during a programme which seeks to change their cognitive processes and that other non-cognitive factors play a large role in how learners perform in these foundation programmes. Being disadvantaged does not unequivocally mean that potential is resident. It is important to identify potential among a sub-group of disadvantaged learners. **The conclusions in this study point to**

8 Miller (1992) takes issue with this very point, and his opinion on the matter cannot be ignored. For, if, as Nunns and Ortlepp (1994) and numerous other researchers believe prior schooling need not carry that much weight, then it follows that if "prior school learning is irrelevant for the prediction of university performance, then we must seriously ask why bother with school" (p. 100).

9 See Chapter 2 for information on the differences and similarities between Bruner, Vygotsky and Piaget in the context of dynamic assessment.

evaluating foundation programmes to see whether or not they can accurately identify learners who could benefit from such programmes in the first place.

5.3 Main findings of documented literature

The following section summarises the various findings evidenced from both sections 4.3 and section 4.4. Not all results are clear-cut positive or negative results; and with dynamic assessment research, its nature lending itself to such "grey" areas of interpretation this is hardly surprising. However, an attempt has been made, for the sake of clarity and economy of thought, to assess these documented results in a quantitative manner. Within this framework, research results have been classified into one of three groups, namely, generally positive results, generally negative results and both positive and negative results. Apart from the Boeyens (1989a) and the Potter and Jamotte (1985) studies which deal with theoretical underpinnings and static measures respectively and do not fall into any of the three categories mentioned this section considers the results of the remaining 29 studies.

Generally positive results

The findings of Adams (1994), Andrews (1996), Bradbury and Zingel (1998), Coosner (1999), De Beer (2000), De Villiers (1996), De Villiers (1999), Gewer (1998), Henley (1989), Lloyd and Pidgeon (1961), Lopes, Roodt and Mauer (2001), Moll (1989), Murray (1988), Nunns and Ortlepp (1994), Shochet (1986), Sibaya, Hlongwane and Makunga (1996), Taylor (1996), Taylor (1999), Van Aswegen (1997), Zolezzi (1992) and Zolezzi (1995) can all be considered generally positive and after limitations of all studies are taken into account as well as the few negative findings (negative in terms of not supporting the null hypothesis in most instances), these studies prove that dynamic assessment is most assuredly an approach not to be ignored. Future research in this direction is encouraged. Of all the studies then, 21 of 29 are generally positive, and that reflects 72% of all the studies. This is a very encouraging and overwhelmingly positive result.

Generally negative results

The findings of Engelbrecht (1999) and Hoffenberg (1988) can be considered generally negative and after limitations of both studies are taken into account as well as the few positive findings (positive in terms of partially supporting the null hypothesis in these instances), these studies do not evidence encouraging results for the further use of dynamic assessment as outlined in the studies. Of all the studies then, 2 of 29 are generally negative, and that reflects 7% of all the studies. This is still a very good result in terms of low figures of negative results.

Both positive and negative results

The findings of Boeyens (1989b), Gaydon (1988), Lipson (1992), Miller (1998), Nel (1997) and Shochet (1994) can all be considered both positive and negative in terms of results evidenced within the dynamic assessment field. After limits of all these "grey area" studies are taken into account as well as both positive and negative findings (positive in terms of partially supporting the null hypothesis in some instances and negative in terms of not supporting the null hypothesis in other instances), these studies offer encouragement with due caution attached. Of these studies then, 6 of 29 evidence both positive and negative results and that reflects 21% of all the studies. The encouraging conclusions in these studies (even though negative findings did result) can also be considered as partially supporting research in the field of dynamic assessment.

In essence, these findings confirm the importance of dynamic assessment research in South Africa.

5.4 Summary of results from informal interviews

The section investigates the main findings of dynamic assessment that has been used in the past or that which is presently being used at tertiary institutions.

5.4.1 Technicon Pretoria

Although matriculation examination results are still the main source of information for the admissions department, students who are selected at this institution undergo assessment and remediation as part of an intensive year-long programme to promote student success in tertiary education. A computer programme is used which utilizes a pre-test/posttest set-up with mediation in between these assessments. In this manner, students are allowed to become familiar with the English language and the rules which govern its grammar and spelling and this allows all students, regardless of educational background, the opportunity to learn before being assessed on the posttest. Remediation is ongoing, although human mediation and intervention are kept to a minimum. **Although not construed as a dynamic test in the pure sense, this is a positive sign for the direction in which alternative admissions are heading in South Africa**.

5.4.2 Natal Technicon

Natal Technicon is aware of the work being done at the Technicon Pretoria. Although the PTEEP which is used here is not considered dynamic by some, the fact that it is being used for the first time, is a positive factor. However, owing to the amalgamation of this Technicon with another Technicon, the test will soon be replaced by the TELP test. **What will become of dynamic testing here remains to be seen.**

5.4.3 University of Natal - Teach-Test-Teach programme (TTT)

Although the Teach-Test-Teach programme (TTT) is no longer run at the University of Natal, it was included in this study as an example of dynamic assessment nearing its most pure form, bar the fact that the tests relied on content to be learned within the academic context, as opposed to contentless tests in typical Feuersteinian assessment. As has been mentioned in Chapter 4, cost effectiveness and efficiency are two main reasons why this project was shut down. However, based on the success of the project from its inception until 1995, the many positive aspects of this project should be remembered.

The most striking aspect of the TTT programme is the order in which teaching and testing progressed. As is usually the scenario, TTT implies a test-teach-test format where pre-test and posttest scores are evaluated and a difference score calculated. However the TTT programme stresses mediation and re-mediation as main goals. Once students were admitted to the programme, teaching continued as part of the programme. Not only are playing fields levelled within the selection context, but investment is made in students after selections as well. The potential pit-fall with this type of approach is however, resource intensive backing, resources which are, of course, scarce. This is the excuse many make as a reason why dynamic assessment (stressing mediation and re-mediation in particular) is not implemented at more tertiary institutions at present. However, the argument is as follows: if one is to select students in a fair manner from a heterogeneous pool (both advantaged and disadvantaged) and one hopes that in so doing student retention rates and success rates will be just as high as the norm or even higher, and that the money invested into these ventures[10] is seen to offer good returns by way of student success, is it then not wise and feasible in the long run, to identify and re-mediate properly on the basis of a Teach-Test-Teach format as opposed to a once-off admissions test which is itself dynamically administered. Dynamically administered as opposed to dynamic assessment[11] are two ends of the continuum mentioned a number of times thus far.

---

10 The word "ventures" is used, as these days progressively more tertiary institutions, like so many other institutions previously reliant on government aid and subsidies, are now being pushed into more competitive roles, roles which are making these institutions more business-like and commercially competitive.

11 Campione and Brown make it very clear that the essence of dynamic instruments is their " …need to be **continuously re-evaluated** as the student begins to acquire skill within some domain" [own emphasis] (in Lidz, 1987, p. 87).

This issue has to be approached and solved by each institution and its merits will have to be reviewed carefully. Some institutions simply cannot afford resource intensive programmes, notwithstanding sustaining such programmes. Ultimately, the onus lies with the institution. The TTT programme was one of the very few programmes running at the time (1987 - 1995) and is an example of a workable solution to student selections and continued consolidation of learning and teaching. **Although the programme no longer runs as such, it contributes a positive result to the field of dynamic assessment in South Africa and is a project that should be referenced if any other project of a similar nature is ever decided upon in the future.**

Just as Mervin Skuy (University of the Witwatersrand) has in the past collaborated with overseas researchers in the field of dynamic assessment, so has Ronald Miller[12] (University of Natal), and quite extensively in fact and most importantly has collaborated with Feuerstein among others. Instrumental enrichment, cognitive processes and cognitive changes have been developed, explored, discussed and implemented in a number of collaborative attempts, namely, in Feuerstein, Miller, Hoffman, Rand, Mintzker and Jensen (1981); Feuerstein, Miller, Rand and Jensen (1981); Feuerstein, Rand, Hoffman and Miller (1980) and Miller, Pascual-Leone,[13] and Andrew (1991). **This is most certainly a very positive foundation from which to further pursue dynamic assessment research in South Africa.**

5.4.4 University of Cape Town - Alternative Admissions Project (AARP)

The fact that the Alternative Admissions Project (AARP) is still running at the University of Cape Town attests to its success as a dynamic assessment test battery. Although the test is not characterised by a test-teach-test approach (which so often is the defining feature of dynamic assessment in South Africa), the dynamic assessment components are integrated within two of the three sub-tests in the AARP, namely, the Placement Test in English for Educational Purposes (PTEEP) and the Mathematics Comprehension Test. To date, reliability and validity data have affirmed the continued use of the test battery. A learning component within the test itself allows for a process of scaffolding, which is akin to mediation, although this definition may not seem commensurate with the strict definition of mediation in the Feuersteinian sense of the word, even though the tasks themselves are said to function as teaching agents.

As with almost all tertiary institutions employing some form of what is perceived to be dynamic assessment, the University of Cape Town applies the field within the selections and admissions faculty. Of course this is not the only area of application but is its main area in South Africa at present.

The two sub-tests do not rely on prior learning and are "contentless" as far as academic content is concerned. The tests are able to discriminate between students who harbour potential for further higher education and those who do not. **The AARP's two sub-tests mentioned above can be construed as dynamic assessment. It has been shown to be valuable and psychometrically[14] defensible and is seen to be a positive contribution to the field of dynamic assessment in South Africa.**

12 Miller is not directly involved in testing, dynamic or otherwise at this stage (R. Miller, personal communication, March 8, 2001).

13 According to Pascual-Leone, rates of cognitive development differ across cultures and are an important factor of learning, a factor that Piaget may not have emphasised (Niaz & Caraucan, 1998).

14 Note that the very use of the word may cause consternation with advocates who uphold the tenets of Feuersteinian dynamic assessment. The fact that a dynamic test has to be held up to psychometric validation is itself a contradictory concept. However, in order for tests to be assessed as fair, they have to withstand psychometric rigour. It would be interesting to note what the developers of the TTT programme would have to say about this.

5.4.5 University of Pretoria - Alternative Admissions Project (AARP)

The University of Pretoria makes use of the AARP sub-tests as developed and co-ordinated by the University of Cape Town. Changing from a prior reliance on static type test to a more dynamic approach bodes well for the future use and perhaps further development of dynamic assessment at this university. As with other previous advantaged universities, it has to deal with the continued problem of adequate and fair student selections. This entails admitting students from disadvantaged backgrounds with the potential to succeed at university. **This is a positive result for the case of dynamic assessment.**

5.4.6 University of Potchefstroom - APIL test battery

The fact that a dynamic test (among other types of tests) is being used as part of selections at the University of Potchefstroom is very encouraging. The issue of prior negative results using this same test in an earlier study, however, would necessitate further investigation into exactly why the test is being used. **Nevertheless this is a positive step once again for the field in South Africa.**

5.4.7 University of Stellenbosch - development of internal dynamic test battery

This university is beginning the process of developing and introducing a dynamic assessment battery, due for implementation in 2001. **This augurs well for dynamic assessment in South Africa.**

5.4.8 University of Port Elizabeth - APIL test battery

The results from a study into the predictive validity of the dynamic APIL test battery yielded positive results for the University of Port Elizabeth. The correlations between academic performance and scores on the various sub-tests of the APIL are an encouraging reminder that dynamic assessment batteries are useful in many contexts (since the APIL can be utilised in many contexts). Although the significant correlations that were achieved were not high but low to moderate, the use of dynamic tests in culturally heterogeneous institutions has been proved to be successful and a worthwhile cause. Unfortunately, due to revisions of policies, the APIL is no longer in use by this university as part of the assessment batteries for selection and admissions. The focus has changed from potential assessment to core competency measurement. Some middle ground will have to be found in the future, where both these demands can be catered for to the mutual benefit of both. **In sum, the results are positive.**

5.4.9 Rand Afrikaans University - APIL test battery

At present, the concept of learning potential and its measurement is acknowledged by the Rand Afrikaans University; but perhaps on the basis of the rather neutral findings in the study by Nel (1997) it has not received the attention it should. It is important to consider that just because one dynamic assessment tool might not yield sufficiently positive results in a particular context and time, does not in any way mean that other types cannot be afforded the opportunity of trial. **Hence, although the results might not be entirely positive within this context, the findings do not themselves constitute a general negative finding either.**

5.4.10 University of the Witwatersrand - dynamic assessment research

The most important work within the dynamic assessment field has emanated from the Cognitive Research Programme, headed by Professor Mervin Skuy at this university. Many studies conducted at this university have yielded positive results. Moreover, international collaboration has at least alerted the international research community to the existence of the practice of dynamic assessment research in South Africa. The main focus within the dynamic assessment research at this university has been on the work of Feuerstein, with particular emphasis on Mediated Learning Experience (MLE) as key component in cognitive modifiability. Numerous studies undertaken by this university have highlighted the usefulness of implementing dynamic assessment in the South African context, a

context which is particularly open to alternative assessments, such as learning potential tests. South Africa's cross-cultural society and the educational implications of making use of dynamic assessment make the country particularly suited to this approach and it can only but benefit from this approach. Moreover, this institution has been involved in dynamic assessment research for over two decades and the wealth of information that is available should spur other institutions in similar directions. **The results are very encouraging and it will prove very interesting to see what further research this university has to offer in this field in the future.**

5.5 Main findings of informal interviews

Of the 36 tertiary institutions that were contacted, 10 institutions revealed dynamic assessment procedures that were:

- In place at the time of contact
- Being thought about and considered for possible future use
- Not necessarily dynamic in the strictest sense of the definition but were assuredly veering in the direction of dynamic assessment
- Used in the past, but are no longer being implemented for various reasons. Foremost among which are costs and timing.

This represents a quantitative finding of 27% of tertiary institutions either having been involved in or presently involved with the use of dynamic assessment. More enlightening is the fact that three institutions, in particular, have contributed more (in terms of research) than the remaining seven. These three institutions, the University of the Witwatersrand, the University of Natal and the University of Cape Town have contributed very useful information, information which is mixed as to positive and negative results. Of the remaining 16 institutions, most have or are in the process of implementing alternative admissions programmes, however, not in the field of dynamic assessment. The results can be interpreted as positive from the viewpoint that at least a few organisations are aware of the potential use of dynamic assessment but, on the other hand, can be seen as disappointing in terms of the fact that some have since disbanded dynamic assessment. Perhaps the most important aspect here is that very few people within these institutions know about dynamic assessment, and fewer know the intricate details of what exactly a dynamic approach to assessment entails.[15] There is awareness in some quarters as to the possible meaning of the concept, but other than the fact that people may "have heard of it[dynamic assessment]" to many it still remains an enigma.

5.5.1 Other tests and research findings in the field

A test that has been developed and was at the time of writing being pilot-tested in the field (see section 4.7) by the HSRC, is particularly encouraging in terms of institutional support for dynamic assessment. Most publications are thus far concentrated within degree studies and not too many receive the backing of larger institutions. **This is a positive result in this study. It is hoped that the test's details will become available at a later date and will be widely publicised.**

5.6 Main themes running throughout the results

The most noticeable negative "themes" gathered from the myriad results within the field of dynamic assessment are of a methodological nature and can be summarised as follows:

15 Compare these findings to a similar study conducted by Lidz (1992b) in which she surveyed 120 school psychologists and their knowledge, use and incorporation of dynamic assessment into programmes of cognitive assessment. Her main findings evidenced a wide-spread familiarity with the concept, gained primarily through reading; and those who make use of it implement Feuerstein's model and see it as a welcome change from product-based assessment procedures. However, those who were not familiar with the concept also did not understand the notions and ideas underpinning the concept. Time, lack of competence as well as a lack of match between the school/employment situation were also major issues curtailing the use of dynamic assessment in the United States.

- **Lack of adequate training** - there is a lack of proper instruction to research aids (usually undergraduate or post-graduate students in Psychology) in research projects (particularly emanating from Master's and Doctoral studies), in that these students are rarely taught effectively how to mediate in this method of assessment and even though authors state that these aids have received instruction, when the results are not as positive as expected, state that these students were not trained fully
- **Lack of time** - having sufficient time seems to be a major issue, many authors stating that if given more time, results would have improved significantly. However, the point remains that in the field of dynamic assessment, time is an essential feature. If time is needed, it should be granted and not served as an excuse for poor, inadequate or unexpected results
- **Small sample size** - this is an issue which is frequently encountered in these studies. If workable conclusions are to be made based on statistical defensibility, surely sample sizes should be compatible with the required statistical procedures that need to be carried out on the data. These procedures (often regression analyses of one sort or another) are valuable and would most certainly add credence to any results obtained in the studies. However, these procedures are not conducted in many instances owing to lack of adequate sample size. If the nature of the study and the results are very closely dependent on adequate sample size and this is known from the outset, surely, the sample size should not be compromised. Larger sample sizes are needed, since a small sample size immediately precludes the results.[16] A predicted finding can already be tentatively made which really negates the reasons for doing the study in the first place
- **Use of normed criterion variables (tests) on samples which are not the normed population -** it is easy to sympathise with researchers who really have no choice but to use tests, usually serving as the criterion variable in the study, on samples for whom the test was not normed. One cannot realistically expect researchers to develop and norm a criterion test for their present research endeavours. This may seem then, to be an unfair criticism levelled at the various authors. Perhaps it is time to norm tests on the populations (usually black and disadvantaged students) on whom the study is being conducted. This is more of a general critique but the implications are that skewed results from criterion measures do not help the ultimate findings when learning potential tests are then compared to these criterion measures
- **Lack of control groups -** although many studies do make use of control groups in order to assess for unwanted effects, some studies do not make use of control groups. Control groups can only but benefit findings. Certain studies make use of the Solomon Four Group Design which is a step better (statistically speaking) in these studies. More reliable conclusions can be made with the added knowledge that control groups affirm or disaffirm extraneous variables.

In sum, if researchers are to go to such great lengths, such as designing intricate studies, running varied statistical analyses on the data, devising training sessions and the like, the above factors should be closely looked at, and not glanced over and perceived as oversights to be rectified in later studies. These issues are highlighted, so that future research studies can try to avoid these potential pitfalls in their designs.

The most noticeable positive "themes" gathered from the myriad results within the field are of a general nature and can be summarised as follows:

- **Standardisation of administration** - although individual administration of learning potential tests is the most desirable option, this manner of administration is clearly not viable at the moment in South Africa. Standardisation provides two positive aspects, namely, that a group of learners can be assessed at any one time, and issues of psychometric significance such as reliability and validity can be more easily monitored and measured, thus ensuring equity of test administration

16 The question as to what one can really state about the statistical findings of a sample of less than 20 is posed and whether the findings are generalisable. Lastly, the question as to whether such a study is valid as a quantitative effort or more valid as a qualitative effort can be posed. These issues must be looked at more closely in future.

- **Locally developed and normed tests**[17] - the Learning Potential Computerised Adaptive Test (LPCAT), the LAB (Learning Ability Battery), the Ability, Processing of Information and Learning Potential Battery (APIL) as well as the Transfer, Automatisation, Memory and Understanding Learning Potential Battery (TRAM) and the Conceptual Reasoning Test (CRT)[18] tests are locally developed and normed learning potential assessment devices which are currently available, and in the recent past a dynamic assessment battery developed and normed by the Human Sciences Research Council was devised and normed on local South Africans. There is a slow but progressive move away from administering foreign test material in a dynamic manner and a shift towards locally developed test batteries being used
- **Emphasis on educationally disadvantaged learners as a heterogeneous group -** it has been emphasised on many occasions in the studies reviewed that educationally disadvantaged learners do not form an homogeneous group and are in fact varied as to intellect and potential, just as is the case with advantaged learners. Emphasis has been placed on the need to differentiate between those who truly do posses potential and happen to be classified as disadvantaged and those who too are disadvantaged but do not possess the same levels of learning potential. This is especially important for selections and admissions to tertiary education institutions which simply cannot afford to select and subsidise high-risk students. An informed yet fair decision has to be taken and this decision is almost always aided by dynamic assessment batteries which act in complementary roles
- **Distinction between dynamic assessment as sole alternative and necessary complementary assessment** - researchers have highlighted and acknowledged that dynamic assessment batteries do not always provide the desired amount and type of information needed in certain contexts but, in most instances, prediction of success is aided by dynamic assessment. Without the use of conventional tests, validity of prediction decreases, value of information is not as comprehensive and, in most instances, when used in combination, static and dynamic assessment yield more information that is of a better quality than when either is used on its own
- **Variety of contexts investigated** - by far the majority of studies in this survey had, as their field of application, tertiary and secondary education, in which the main emphasis was placed on the prediction of academic success.

However, work has also been done in a variety of working environments such as commercial and semi-parastatal companies. The field of application for dynamic assessment has traditionally remained in the arena of education, but this is not the exclusive domain of this method.[19] It is versatile enough to warrant even further research in other contexts in South Africa, as it has much to offer. In fact, De Beer (1991) researched the usefulness of the Feuerstein Instrumental Enrichment Programme for learning disabled children within the multicultural context, but this survey has emphasised the role of assessment more so than enrichment.

In sum, dynamic assessment, as investigated in these local studies, has a lot to offer South Africans in many spheres in which testing is part and parcel of selection and/or promotion processes. These

17 In fact, the findings of this study are in agreement with those of Van Eeden and De Beer (in Foxcroft & Roodt, 2001), who state that there are very few locally available dynamic assessment tools, among them the APIL, TRAM and LPCAT but state that this field is still under investigation.

18 Contrary to De Villiers' (1999) use of the CRT as a learning potential test, De Villiers did state that the CRT may not necessarily be considered a dynamic test owing to the lack of a "test-teach-test" approach (personal communication, September, 2001).

19 In fact, the overseas literature is replete with studies of dynamic assessment within varied contexts such as speech and language research, deaf and hearing-related research, language disorders, pre-school assessment, assessment of intellectual handicapped children, assessment of metacognitive functioning, and neuropsychological evaluation, assessing social learning potential, assessing reserve capacity in old age as well as assessing vocational aptitude (Guthke in Hamers et al., 1993b; Haywood & Wingenfield, 1992; Keane in Lidz, 1987; Lidz & Pena, 1996; Lidz & Thomas in Lidz, 1987; Lidz in Flanagan et al. 1997; Vye et al., in Lidz, 1987).

positive highlights are foundations upon which to build further and when looked at with the negative points, allows the practitioner a balanced view of the reality of the situation currently in South Africa.

5.7 Limitations of the study

As mentioned in Chapter 3 and appended in the study, not all available research in the field of dynamic assessment could be accessed due to logistics, lack of availability and timing. Furthermore although contact was made with all tertiary institutions, people spoken to in their capacities as knowledgeable sources of information, did not always yield consistent results. These methodological issues may well have affected the nature of the outcomes of the survey. Given greater latitude regarding time and costs, accuracy of information obtained from these sources may well be of a higher quality.

Chapter six highlights general conclusions.

CHAPTER 6 Conclusion and recommendations

"There is reason to believe that voluntary activity, more than highly developed intellect, distinguishes humans from the animals which stand closest to them."

- Lev Vygotsky

6.1 Introduction

Chapter 1 introduced the idea of dynamic assessment as a distinct possibility as a complementary assessment method, an alternative form of assessment which would seek to alter the nature and process of psychometric testing in South Africa, and in so doing, allow more refined, fair and less biased opportunities for learners/prospective learners/employees during testing for selections and admission in whichever contexts. In order to progress within the field of dynamic assessment, a field which is, comparatively speaking, quite young in South Africa when compared to Europe and the States, the best choices as to test-development can and should only be made on the basis of previous findings with similar tests in the field. Most researchers and practitioners in this field (whether active or not) are not aware of the large pool of data that is available on the topic.

The aim of the study has been to provide a summarised version of such studies that have been conducted locally, investigate and compare findings, provide recommendations based on those findings and to offer a succinct account of what is currently available on the topic. This chapter condenses the overall findings into a coherent and more manageable conclusion and thus dispenses with intricate detail dealt with in the preceding two chapters.

6.2 What the results point to - recommendations for further studies in South Africa

A few main issues are tackled and discussed here which will, it is hoped, inform potential researchers in this area who wish to conduct similar studies in future as to the directions and information that is needed to progress in the field.

6.2.2 Early isolated research as a lesson to be learnt for later collaborative research endeavours

It would seem that early research in this field was often fragmentary and isolated. This statement is based purely on various researcher's acknowledgement of other similar research taking place at the time. Thus, it might not necessarily be the case, but if others' work were known surely it would be safe to assume that it would have received due notice and reference. Perhaps there is a lesson to be leant here, especially for future research in this field in South Africa, namely that if dynamic assessment is in any way to hope for greater resource allocation in its projects, then researchers in the field can only benefit through collaboration in their efforts. The use of dynamic assessment in whatever contexts needs to be identified and made known to all other researchers in the field. Consolidation of all dynamic assessment research results (in whichever contexts) would further support the case for dynamic assessment to be implemented and used on a wider scale than is used at present. Fragmentary and isolated research, conducted without due consultation and more precisely, collaboration with other practitioners in the field, may be the undoing of dynamic assessment research. Collaboration in numbers will go a long way in procuring funds for this type of research (which in this dissertation has been revealed as one of the most pressing issues) than would individual attempts at fund procurement.[1] Collaboration strengthens the case for dynamic assessment. Future ideas such as the establishment of a dynamic assessment society or, at least, the awareness of dynamic assessment within larger psychology bodies can aid in making more practitioners aware of what the field has to offer.

1 The AARP (University of Cape Town) is a good example of a national networked co-ordinated attempt at dynamic assessment testing.

6.2.3 Fields of application

Dynamic assessment has its roots in educational assessment and, as such, is strongly linked to education. The objective of most dynamic assessment research endeavours has been to address educational crises and/or the problems faced by educational institutions in South Africa. But this is clearly not the only area of application, for dynamic assessment should filter into as many contexts as possible, contexts (other than the working and educational environments), which could be aided by this type of research such as clinical and counselling psychology, as well as sociological studies. More research in other contexts is needed too, but the status of education and learning (which has been seriously compromised by past discrimination) necessitates the use of dynamic assessment which, if implemented, will prove beneficial as it is well suited to research among culturally (educationally) disadvantaged learners.

6.2.4 Cautious pursuance of dynamic assessment research in future

Practical limits and constraints inhibit the implementation of pure dynamic assessment and even if implemented can only be sustained for a limited period of time. Psychometric defensibility of any new tests that are developed and marketed in South Africa is paramount to its eventual acceptance, and costs and timing are considered crucial elements in any test administration procedure. Between the extremes of both dynamic and traditional assessment paradigms, lies a middle ground, which if utilised correctly, realisation of what can be achieved will be all the more closer at hand. Researchers, practitioners, psychologists, psychometricians and human resource managers are faced with conflicting messages and conflicting results in the area of assessment.

On the one hand, traditional tests are not perceived in a very favourable light, and this is due to many reasons, such as past testing inequalities, use of oversees normed tests and lack of predictive validity for disadvantaged groups. To do away with the tests is clearly not the answer, yet something must give in order to address culture-fairness issues. On other hand, any new test development is usually greeted with dismay if certain traditional tenets are not withheld. Along with traditional assessments, traditional concepts such as reliability and validity take precedence over what the test may purport to offer in the long term.

Dynamic assessment is seen as a partial solution to this issue, a solution that must be combined with what has gone before in test theory and practise. However, dynamic assessment in its most pure form decries the notion and use of alien concepts such as reliability and validity. One can only sympathise with energetic proponents of this staunch view of dynamic assessment advocating the idea that initial outlay of costs in terms of money and time will be far outweighed by the potential benefits that are to be derived at the end. However, reality is filled with obstacles and such realisations of pure dynamic assessment research are in essence idealistic.

The middle ground that has to be reached, and in fact has in most cases already been reached is perhaps, at this stage, the most viable place in which to situate this method of assessment. This study has shown that, although more good can come from this approach than bad, there are weak areas which have been addressed but yet need to be accounted for and rectified. Dynamic assessment is not a blossoming form of test theory, it has been propounded and researched for many decades but has only received closer attention in South Africa in the past two decades. Dynamic assessment cannot afford to be discredited due to poor research design and implementation. It lends itself more easily to potential methodological faults, is vulnerable due to its precarious handling of psychometric issues and can easily be exploited for unscientific purposes more so than traditional tests. The goal is to overcome these obstacles and aim for a more informed and literate research body. It is hoped that future development of dynamic assessment batteries of static tests which are administered in a dynamic manner in South Africa will be better armed in terms of past knowledge and will be able to build on from what has gone before.

Where do we go from here? The results of the study are, in one word, positive. Positive in terms of many aspects. With cautious interpretation of findings, dynamic assessment can and should progress

in this country, when used in combination with other assessment techniques. This conclusion might not be a novel one, but it is, at this stage, an informed one.

References

About PsycINFO. (2001). Retrieved August 10, 2001, from http://www.apa.org/psycinfo/about.htm.

Adams, H.B. (1994). Teaching as assisting performance in the zone of proximal development: implications for teacher education. In D. Adey, P. Steyn, N. Herman & G. Scholtz (Eds.), **State of the art in higher education. Limited congress edition volume 1. Proceedings of the ninth biennial SAARDHE congress and general meeting held at the breakwater campus, Cape Town** (pp. 1-14). Pretoria: University of South Africa.

Adey, D., Steyn, P., Herman, N. & Scholtz. G. (Eds.), (1994). **State of the art in higher education. Limited congress edition volume 1. Proceedings of the ninth biennial SAARDHE congress and general meeting held at the breakwater campus, Cape Town.** Pretoria: University of South Africa.

Anastasi, A. (1988). **Psychological Testing** (6th ed.). New York: Macmillan.

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

Anonymous. (no date). Centre For Higher Education Development, Academic Development Programme: Alternative Admissions Research Project. Internal document. University of Cape Town.

Asmal, K. (1999). Call to action: mobilising citizens to build a South African education and training system for the 21st century. **Statement made by Professor Kader Asmal, Minister of Education, July 27 1999.** Pretoria: Ministry of Education.

Badsha, N. & Yeld, N. (1991). The Alternative Admissions Research Project (AARP) at UCT. **University of Durban-Westville bulletin for academic staff, 12** (2), 32-38.

Bejar, I.I. (1984). Educational diagnostic assessment. **Journal of educational measurement, 21** (2), 175-189.

Bethge, H., Carlson, J.S. & Wiedl, K.H. (1982). The effects of dynamic assessment procedures on Raven Matrices performance, visual search behaviour, test anxiety and test orientation. **Intelligence, 6,** 89-97.

Boeyens, J.C.A. (1989a). **Learning potential: A theoretical perspective**. Pretoria: Human Sciences Research Council.

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

Bolig, E.E. & Day, J.D. (1993). Dynamic assessment and giftedness: the promise of assessing training responsiveness. **Roeper review, 16** (2), 110-113.

Borland, J.H. & Wright, L. (1994). Identifying young, potentially gifted, economically disadvantaged students. **Gifted child quarterly, 38** (4), 164-171.

Bradbury, J. (1993). The Meta-language of cognition - a paper presented for the Kenton Olwandle Conference.

Bradbury, J. & Griesel, H. (1991). Proposal for the implementation of a foundation course based in the faculty of social science - the development and production of knowledge - submitted to the University Educational development board and the Boards of Humanity and Law, November 1991.

Bradbury, J. & Griesel, H. (1994). Text as impetus for learning. State of the art in higher education: limited congress edition vol. 1. In Adey, D., Steyn, P., Herman, N. & Scholtz, G. (Eds.), **Proceedings of the ninth biennial SAARDHE congress and general meeting held at the Breakwater campus, Cape Town** (pp. 323-331). University of South Africa.

Bradbury, J. & Zingel, J. (1998). Learning through peer interaction in a multi-cultural primary school classroom. **South African journal of education, 18** (4), 232-240.

Bransford, J.D., Delclos, V.R., Vye, N.J., Burns, M.S. & Hasselbring, T.S. (1987). State of the art and future directions. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 479-496). New York: Guilford Press.

Brown, A., Campione, J.C., Webber, L.S. & McGilly, K. (1993). Interactive learning environments: a new look at assessment and instruction. In Gifford, B .R. & O' Connor, M.C. (Eds.), **Changing assessments: alternative views of aptitude, achievement and instruction** (pp.121-211). Boston: Kluwer Academic.

Buchel, F.P. & Scharnhorst, U. (1993). The learning potential assessment device (LPAD); the discussion of theoretical and methodological problems. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Rujssenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp. 83-111). Amsterdam: Swets & Zeitlinger.

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

Budoff, M. (1987b). Measures for assessing learning potential. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 173-195). New York: Guilford Press.

Budoff, M. & Friedman, M. (1964). "Learning potential" as an assessment approach to the adolescent mentally retarded. **Journal of consulting psychology, 28** (5), 434-439.

Budoff, M., Meskin, J. & Harrison, R.H. (1971). Educational test of the learning-potential hypotheses. **American journal of mental deficiency, 76** (2), 159-169.

Campione, J.C. (1989). Assisted assessment: a taxonomy of approaches and an outline of strengths and weaknesses. **Journal of learning disabilities, 22** (3), 151-165.

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

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

Canonical Analysis. (2001). Retrieved February 4, 2001, from http://statsoft.com/textbook/stathome.html.

Carlson, J.S. & Wiedl, K.H. (1978). Use of testing-the-limits procedures in the assessment of intellectual capabilities in children with learning difficulties. **American journal of mental deficiency, 82** (6), 559-564.

Carlson, J.S. & Wiedl, K.H. (1992). The dynamic assessment of intelligence. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp.167-186). New York: Spinger-Verlag.

Carroll, J.B. (1997). Theoretical and technical issues in identifying a factor of general intelligence. In B. Devlin, S.E. Fienberg, D.P. Resnick, & K. Roeder (Eds.), **Intelligence, genes, success: scientists respond to the Bell Curve** (pp. 125-156). New York: Spinger-Verlag.

Centre for Higher Education Development. (no date). **Academic development programme: Alternative Admissions Research Project.** Internal Project Document. University of Cape Town.

Centre for Higher Education Development. (2000). **The Alternative Admissions Research Project. Draft report: the predictive validity of the AARP Placement Tests in English for Educational Purposes (PTEEP).** University of Cape Town.

Changeux, J.P. & Konishi, M. (Eds.), (1996). **The neural and molecular bases of learning.** New York: John Wiley & Sons.

Christman, U. & Groeben, N. (1996). Reflexivity and learning: problems, perspectives, and solutions. In J. Valsiner & H. Voss (Eds.), **The structure of leaning processes** (pp. 45-85). New Jersey: Ablex.

Cognitive Enrichment Techniques Research Programme (the). (2001). Retrieved March 1, 2001, from http://www.wits.ac.za/fac/education/cogrp.htm.

Cole, M., John-Steiner, V., Scribner, S. & Souberman, E. (Eds.), (1978). **Mind in society: the development of higher psychological processes.** Cambridge: Harvard University Press.

Coosner, C.D. (1999). **Dynamic assessment - a practical strategy for school educators.** Unpublished M.Ed. dissertation, University of Stellenbosch.

Cormier, P., Carlson, J.S. & Das, J.P. (1990). Planning ability and cognitive performance: the compensatory effects of a dynamic assessment approach. **Learning and individual differences, 2** (4), 437-449.

Craig, A.P. (1989). The conflict between the familiar and unfamiliar. **South African journal of higher education, 3** (1), 166-172.

Craig, A.P. (1991). Adult cognition and tertiary studies. **South African journal of higher education. 5** (2), 137-144.

Craig, A.P. (1996). Education for all. **South African journal of higher education, 10** (2), 47-55.

Craig, A. P. (2000). Time and learning. **South African journal of psychology, 30** (4), 6-16.

Cronbach, L.J. & Furby, L. (1970). How we should measure "change" - or should we? **Psychological bulletin, 74** (1), 68-80.

Daniel, M.H. (1997). Intelligence Testing: status and trends. **American psychologist, 52** (10), 1038-1045.

Daniels, H. (Ed). (1996). **An introduction to Vygotsky.** London: Routledge.

Daniels, H. (1996). Introduction: psychology in a social world. In H. Daniels (Ed.), **An introduction to Vygotsky** (pp. 1-27). London: Routledge.

Das, J.P. & Conway, R.N.F. (1992). Reflections on remediation and transfer: a Vygotskian perspective. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp. 94-115). New York: Spinger-Verlag.

Dawes, P., Yeld, N. & Smith, M.J. (1999). Access, selection and admission to higher education: maximising the use of the school-leaving examination. **South African journal of higher education, 13** (3), 97-104.

Day, J.D. & Cordon, L.A. (1993). Static and dynamic measure of ability: an experimental comparison. **Journal of educational psychology, 85** (1), 75-82.

Day, J.D., Engelhardt, J.L., Maxwell, S.E. & Bolig, E.E. (1997). Comparison of static and dynamic assessment procedures and their relation to independent performance. **Journal of educational psychology, 89** (2), 358-368.

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

De Beer, N.A. (1991). The application potential of the Feuerstein Instrumental Enrichment Program (FIE) for learning disabled children within the multicultural South African context. **Tdyskrif vir Geesteswetenskappe, 31** (3), 189-198.

De Villiers, A.B. (1996). Disadvantaged students: analyzing the zone of proximal development. **South African journal of higher education, 10** (1), 135-139.

De Villiers, A.B. (1999). **Disadvantaged students' academic performance: analysing the zone of proximal development.** Unpublished D.Phil. thesis, University of Cape Town.

Definition of TELP components. (2001). Retrieved July 6, 2001 from http://www.uncf.org.za/telp/components/comp_def.htm.

Delclos, V.R., Vye, N.J., Burns, M.S., Bransford, J.D. & Hasselbring, T.S. (1992). Improving the quality of instruction: roles for dynamic assessment. In H.C. Haywood & D. Tzuriel (Eds**.), Interactive assessment** (pp. 317-331). New York: Spinger-Verlag.

Devlin, B., Fienberg, S.E.; Resnick, D.P. & Roeder, K. (1997). **Intelligence, genes, success: scientists respond to the Bell Curve.** New York: Spinger-Verlag.

Du Bois, P.H. (1970). **A history of psychological testing.** Boston: Allyn and Bacon.

Duffy, K.G. & Wong, F.Y. (1996). **Community Psychology**. Boston: Allyn and Bacon.

EBSCO Publishing. (2001). Retrieved July 19, 2001 from http://epnet.com/index.html.

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

Embretson, E.S. (1987). Toward development of a psychometric approach. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 141-170). New York: Guilford Press.

Embretson, E.S. & Reise, S.P. (2000). **Item response theory for psychologists**. London: Lawrence Erlbaum.

Engelbrecht, M. (1999). **Leerpotensiaal as voorspeller van akademiese van universiteitstudente.** Unpublished D.Phil. thesis, Potchefstroom University for Christian Higher Education.

ERIC System (the). (2001). Retrieved July 15, 2001, from http://www.eric.ed.gov/about/about.htm.

Eysenck, M.W. (Ed.). (1994). **The Blackwell dictionary of cognitive psychology**. Oxford: Blackwell Publishers.

Evaluation of the Assessment Strategy for Admission at Pretoria University. (2001). Retrieved December 8, 2001 from http://hagar.up.ac.za/catts/learner/andres/assess.htm.

Feuerstein, R. (1972). Cognitive assessment of the socioculturally deprived child and adolescent. In L.J. Cronbach & P.J.D. Drenth (Eds.), **Mental tests and cultural adaptation** (pp. 265-275). The Hague: Mouton.

Feuerstein, R. Feuerstein, R. & Gross, S. (1997). The learning potential assessment device. In D.P. Flanagan, J.L. Genshaft & P.L Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp. 297-313). New York: Guilford Press.

Feuerstein, R., Miller, R., Hoffman, M.B., Rand, Y., Mintzker, Y. & Jensen, M.R. (1981). Cognitive modifiability in adolescence: cognitive structure and the effects of intervention. **Journal of special education, 15** (2), 269-287.

Feuerstein, R., Miller, R., Rand, J. & Jensen, M.R. (1981). Can evolving techniques better measure cognitive changes? **Journal of special education, 15,** 201-219.

Feuerstein, R., Rand, Y. & Hoffman, M. (1979). **The dynamic assessment of retarded performers: the learning potential assessment device, theory, instruments and techniques.** Baltimore: University Park Press.

Feuerstein, R., Rand, Y., Hoffman, M.B. & Miller, R. (1980). **Instrumental Enrichment: an intervention program for cognitive modifiability.** Illinois: Scott, Foresman and Company.

Feuerstein, R., Rand, Y., Jensen, M.R., Kaniel, S. & Tzuriel, D. (1987). Prerequisites for assessment of learning potential: the LPAD model. In C.S. Lidz (Ed.), **Dynamic assessment: an interaction al approach to evaluating learning potential** (pp. 35-51). New York: Guilford Press.

Flanagan, D.P., Genshaft, J.L & Harrison, P.L. (Eds.). (1997). **Contemporary intellectual assessment: theories, tests and issues.** New York: Guilford Press.

Folman, R. & Budoff, M. (1971). Learning potential and vocational aspirations of retarded adolescents. **Exceptional children, 38** (2), 121-131.

Foxcroft, C. & Roodt, G. (2001). **An introduction to psychological assessment in the South African context.** Cape Town: Oxford University.

Foxcroft, C., Roodt, G. & Abrahams, F. (2001). Psychological assessment: a brief retrospective overview. Assessment of cognitive functioning. In C. Foxcroft & G. Roodt (Eds.), **An introduction to psychological assessment in the South African context (**pp. 11-33). Cape Town: Oxford University Press.

Frisby, C. & Braden, J.P. (1992). Feuerstein's dynamic assessment approach: a semantic, logical and empirical critique [Electronic version]. **Journal of special education 26** (3), 281-300.

Gaydon, V.P. (1988). **Predictors of performance of disadvantaged adolescents on the Soweto/Alexandra gifted child programme.** Unpublished M.Ed. dissertation, University of the Witwatersrand.

Gewer, A. (1998). Uncovering potential: dynamic assessment of non-verbal reasoning ability in educationally disadvantaged children. Unpublished M.Ed. dissertation, University of the Witwatersrand.

Gifford, B .R. & O' Connor, M.C. (Eds.), (1993). **Changing assessments: alternative views of aptitude, achievement and instruction.** Boston: Kluwer Academic.

Gipps, C. V. (1994). **Beyond testing: towards a theory of educational assessment.** London: The Flamer Press.

Gould, S.J. (1996). **The mismeasure of man** (2nd ed.). New York: Penguin.

Gray, R. (1990). Analysis of Entrance Test - internal report. Pretoria: University of Pretoria.

Griesel, H. (1991). Selection as an education response. **University of Durban-Westville Bulletin for academic staff, 12** (2), 19-31.

Griesel, H. (1999). **Access and the higher education sector: A South African case study on policy and programme achievement.** Background document for the association for the development of education in Africa (ADEA) and South African national Department of Education (DoE).

Griesel, H. (2000). The problem of selection - draft paper prepared for the Assessment in Higher Education Symposium, University of the Western Cape, 18 September.

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

Guthke, J. (1982). The learning test concept - an alternative to the traditional static intelligence test. The German journal of psychology, 6 (4), 306-324.

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

Guthke, J. (1993b). Developments in learning potential assessment. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Rujssenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp. 43-68). Amsterdam: Swets & Zeitlinger.

Guthke, J. & Wingenfeld, S. (1992). The learning test concept: origins, state of the art, and trends. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp. 64-93). New York: Spinger-Verlag.

Haeck, W., Yeld, N., Conradie, J., Roberston, N. & Shall, A. (1997). A developmental approach to mathematics testing for university admissions and course placement. **Educational studies in mathematics, 33** (1), 71-91.

Hamers, J.H.M., Hessels, M.G.P. & Pennings, A.H. (1996). Learning potential in ethnic minority children. **European journal of psychological assessment, 12** (3), 183-192.

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

Hamers, J.H.M., Sijtsma, K & Rujssenaars, A.J.J.M. (Eds.). (1993). **Learning Potential assessment: theoretical, methodological and practical issues.** Amsterdam: Swets & Zeitlinger.

Hanson, B.G. (1995). **General Systems Theory.** Washington: Taylor and Francis.

Harrison, R., Singer, J., Budoff, M. & Folman, R. (1972). Level of aspiration as a function of learning potential status in the educable mentally retarded. **Psychological reports, 30** (1), 47-57.

Haywood, H.C. & Wigenfeld, S.A. (1992). Interactive assessment as a research tool [Electronic version]. **Journal of special education, 26** (3), 253-269.

Haywood, H.C. & Tzuriel, D. (Eds.), (1992). **Interactive assessment.** New York: Spinger-Verlag.

Haywood, H.C. & Tzuriel, D. (1992). The status and future of interactive assessment. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp. 504-507). New York: Spinger-Verlag.

Haywood, H.C., Tzuriel, D. & Vaught, S. (1992). Psychoeducational assessment from a transactional perspective. In H.C. Haywood & D. Tzuriel (Eds.), **Interactive assessment** (pp. 38-63). New York: Spinger-Verlag.

Hedegaard, M. (1996). The zone of proximal development as basis for instruction. In H. Daniels (Ed), **An introduction to Vygotsky** (pp. 171-195). London: Routledge.

Henley, S.J. (1989). **An investigation of Feuerstein's theory of mediated learning experience with a disadvantaged community.** Unpublished M.A. dissertation, University of the Witwatersrand.

Hinton, G.E. (1992). How neural networks learn from experience. **Scientific American, 267** (3), 105-109.

Hoffenberg, S.R. (1988). **Effectiveness of the learning potential assessment device with high achieving adolescents from an advantaged community.** Unpublished M.Ed. dissertation, University of the Witwatersrand.

Huysamen, G.K. (1980). **Psychological testing in South Africa.** Pretoria: Academica.

Huysamen, G.K. (1996). Fair and unbiased admission procedures for South African institutions of higher education. **South African journal of higher education 10** (2), 199-207.

Huysamen, G.K. (1997). Potential ramifications of admissions testing at South African institutions of higher education. **South African journal of higher education, 11** (1), 65-71.

Huysamen, G.K. (1999). Psychometric explanations for the poor predictability of the tertiary-academic performance of educationally disadvantaged students. **South African journal of higher education, 13** (1), 132-138.

Ittenbach, R.E., Esters, I.G. & Wainer, H. (1997). The history of test development. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp. 17-31). New York: Guilford Press.

Jensen, A.R. (1980). **Bias in mental testing.** New York: The Free Press.

Jensen, M.R. & Feuerstein, R. (1987). The learning potential assessment device: from philosophy to practice. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 379-402). New York: Guilford Press.

Jitendra, A.K. & Kameenui, E.J. (1993). Dynamic assessment as a compensatory assessment approach; a description and analysis [Electronic version]. **Remedial and special education, 14** (5), 6-12.

Kamphaus, R.W., Petoskey, M.D. & Morgan, A.W. (1997). A history if intelligence test interpretation. In D.P. Flanagan, J.L. Genshaft & P.L Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp. 32-47). New York: Guilford Press.

Kandel, E.R. & Hawkins, R.D. (1992). The biological basis of learning and individuality. **Scientific American, 267** (3), 53-60.

Kandel, E.R. & Squire, L.R. (2000). Neuroscience: breaking down scientific barriers to the study of brain and mind. **Science, 290** (5494), 1113-1120.

Kanjee, A. (2001). Cross-cultural test adaptation and translation. In C. Foxcroft & G. Roodt (Eds.), **An introduction to psychological assessment in the South African context** (pp 86-102). Cape Town: Oxford University Press.

Keane, K.J. (1987). Assessing deaf children. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 360-376). New York: Guilford Press.

Kerlinger, F.N. (1981). **Foundations of behavioural research** (2nd ed.). New York: Holt-Saunders.

Klauer, K.J. (1993). Learning potential testing; the effect of retesting. In J.H.M Hamers, K. Sijtsma & A.J.J.M. Rujssenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp. 135-152). Amsterdam: Swets & Zeitlinger.

Kozulin, A. (1990). **Vygotsky's psychology: a biography of ideas.** Cambridge: Harvard University Press.

Laughon, P. (1990). The dynamic assessment of intelligence: a review of three approaches [Electronic version]. **School psychology review, 19** (4), 459-471.

Lidz, C.S. (1981). **Improving assessment of schoolchildren.** London: Jossey-Bass.

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

Lidz, C.S. (1992a). Dynamic assessment: some thoughts on the model, the medium and the message. **Learning and individual differences, 4** (2), 125-136.

Lidz, C.S. (1992b). The extent of incorporation of dynamic assessment into cognitive assessment courses; a national survey of school psychology trainers [Electronic version]. **Journal of special education, 26** (3), 325-331.

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

Lidz, C.S. & Macrine, S.L. (2001). An alternative approach top the identification of gifted culturally and linguistically diverse learners. **School psychology international, 22** (1), 74-96.

Lidz, C.S. & Pena, E. D. (1996). Dynamic assessment: the model, its relevance as a nonbiased approach, and its application to Latino American preschool children. **Language, speech and hearing services in schools, 27** (4), 367-372.

Lidz, C.S. & Thomas, C. (1987). The preschool learning assessment device: extension as a static approach. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 288-326). New York: Guilford Press.

Lipson, L.E. (1992). **Relationship of static and dynamic measures to scholastic achievement of black pupils.** Unpublished M.Ed. dissertation, University of the Witwatersrand.

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

Lloyd, P. (1995). **Psychology: cognition and language development.** Leicester: British Psychological society.

Lopes, A., Roodt, G. & Mauer, R. (2001). The predictive validity of the APIL-B in a financial institution. **Journal of industrial psychology, 27** (1), 61-69.

Luria, A.R. (1978). Biographical note on L.S. Vygotsky. In L. Vygotsky, **Mind in society** (pp. 15-16). London: Harvard University Press.

Mautner, T. (Ed.). (2000). **The penguin dictionary of philosophy**. London: Penguin.

Meijer, J. (1993). Learning potential, personality characteristics and test performance. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Rujssenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp. 341-362). Amsterdam: Swets & Zeitlinger.

Meijer, J. & Elshout, J.J. (2001). The predictive and discriminant validity of the zone of proximal development. **British journal of educational psychology 71** (1), 93-113.

Menzel, R. & Bicker, G. (1987). Plasticity in neuronal circuits and assemblies of invertebrates. In J.P. Changeux, & M. Konishi (Eds.), **The neural and molecular bases of learning** (pp. 433 - 472). New York: John Wiley & Sons.

Merzenich, M.M. (1987). Dynamic neocortical processes and the origins of higher brain functions. In J.P. Changeux, & M. Konishi (Eds.), **The neural and molecular bases of learning** (pp. 337-358). New York: John Wiley & Sons.

Miller, R. (1991). Evaluation report on the Teach-Test-Teach programme (1989-1990). Submitted to the University of Natal, March.

Miller, R. (1992). Double, double, toil and trouble: the problem of student selection. **South African journal of higher education. 6** (1), 98-104.

Miller, R. (1998). A follow-up of the academic performance of English first and second language students. **South journal of higher education, 12** (2), 167-175.

Miller, R. & Bradbury, J. (1999). Academic Performance of English first and second language students: selection criteria. **South African journal of science 95** (1), 30-34.

Miller, R., Pascual-Leone, J. & Andrew, D.J. (1991). Cognitive executive processes and mental capacity on the compound stimulus visual information task in a group of Zulu-speaking children. **South African journal of psychology, 22** (1), 1-9.

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

Moll, I.C. (1989). Roots and disputes of cognitive developmental conceptions of teaching. **South African journal of education, 9** (4), 714-721.

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

Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J. & Urbina, S. (1996). Intelligence: knowns and unknowns. **American psychologist, 51** (2), 77-101.

Nel, A. (1997). **Die voorspelling van akademiese sukses binne konteks van 'n alternatiewe universiteitstoelatingsbeleid.** Unpublished M.A. dissertation, Rand Afrikaans University.

Neuman, W.L. (1997). **Social research methods: qualitative and quantitative approaches.** Boston: Allyn and Bacon.

Nexus database system. (2001). Retrieved May 14, 17, 18 and July 24, 26 2001 from http://star.nrf.ac.za/scripts/starfinder.exe/155/nexus.txt.

Niaz, M. & Caraucan, E. (1998). "Learning-to-learn": a neo-Piagetian interpretation of the potential for learning. **Perceptual and motor skills, 86** (2), 1291-1298.

Nunns, C. & Ortlepp, K. (1994). Exploring predictors of academic success in Psychology 1 At Wits university as an important component of fair student selection. **South African journal of psychology, 24** (4), 201-207.

Online Databases. (2001). Retrieved August 14, 2001 from http://www.hsrc.ac.za/databases.html.

Potter, C.S. & Jamotte, A.N. (1985). African matric results: Dubious indicators of academic merit. **Indicator SA 3** (1), 10-13.

Quartz, S.R. & Sejnowski. T.J. (1997). The neural basis of cognitive development: a constructivist manifesto. **Behavioral and Brain sciences 20** (4), 537-596.

Rand, Y. & Kaniel, S. (1987). Group administration of the LPAD. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 196-214). New York: Guilford Press.

Ruijssenaars, A.J.J.M., Castelijns, J.H.M. & Hamers, J.H.M. (1993). The validity of learning potential tests. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Rujssenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp. 69 - 82). Amsterdam: Swets & Zeitlinger.

Rushton, J.P., Skuy, M.S. & Fridjhon, P. (2001). **Report on study of performance on Raven's Matrices by engineering students at the University of the Witwatersrand.** Internal document.

Sabinet Online. (2001). Retrieved June 12, 2001 from http://www.sabinet.co.za.

Sahakian, W.S. (1981). **History of psychology: a source book in systematic psychology**. Illinois: F.E. Peacock.

Schaffer, H.R. (1996). Joint involvement episodes as context for development. In H. Daniels (Ed.), **An introduction to Vygotsky** (pp. 251-280). London: Routledge.

Sehlapelo, M. & Terre Blanche, M. (1996). Psychometric testing in South Africa: views from above and below. **Psychology in society, 21,** 49-59.

Selkoe, D.J. (1992). Aging brain, aging mind. **Scientific America, 267** (3), 96-103.

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

Shochet, I.M. (1986). **Manifest and potential performance in advantaged and disadvantaged students.** Unpublished D.Phil. dissertation, University of the Witwatersrand.

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.

Sibaya, P.T., Hlongwane, M. & Makunga, N. (1996). Giftedness and intelligence assessment in a third world country: constraints and alternatives. **Gifted education international, 11** (2), 107-113.

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. Rujssenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp. 117 - 134). Amsterdam: Swets & Zeitlinger.

Sijtsma, K. (1993b). Psychometric issues in learning potential assessment. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Rujssenaars (Eds.), **Learning Potential assessment: theoretical, methodological and practical issues** (pp. 175 - 194). Amsterdam: Swets & Zeitlinger.

Skuy, M. S. (1989). **Dynamic assessment of South African children and adolescents.** Unpublished document. University of the Witwatersrand.

Skuy, M.S. (1997). Cross cultural and interdimensional implications of Feuerstein's construct of Mediated Learning. **School psychology international, 18** (2), 119-135.

Skuy, M.S., Archer, M. & Roth, I. (1987). Use of the Learning Potential Assessment Device in assessment and remediation of a learning problem. **South African journal of education, 7** (1), 53-58.

Skuy, M.S., Kaniel, S. & Tzuriel, D. (1988). Dynamic assessment of intellectually superior Israeli children in a low socio-economic status community. **Gifted Education International, 5** (2), 90-96.

Skuy, M.S., Mentis, M., Durbach, F., Cockroft, K., Fridjhon, P. & Mentis, M. (1995). Crosscultural comparison of the effects of Instrumental Enrichment on children in a South African mining town. **School psychology international, 16** (3), 265-282.

Skuy, M.S., Zolezzi, S., Mentis, M., Fridjhon, P. & Cockroft, K. (1996). Selection of advantaged and disadvantaged South African students for university admission. **South African journal of higher education, 10** (1), 110-118.

Smit, G.J. (1996). **Psychometrics: aspects of measurement.** Kagiso: Pretoria.

Snow, R.E. & Lohman, D.F. (1984). Toward a theory of cognitive aptitude for learning from instruction. **Journal of educational psychology, 76** (3), 347-376.

Sternberg, R.J. (1991). Death, taxes, and bad intelligence tests. **Intelligence, 15** (30), 257-269.

Sternberg, R.J. (1997). The concept of intelligence and its role in lifelong learning and success. **American psychologist, 52** (10), 1030-1037.

Students lack language skills. (2001). Retrieved July 17, 2001 from http//:www.news24.co.za/News24/South_Africa/0,1113,2-1-1053215,00.html.

Swanson, H.L. (1995). Using the cognitive processing test to assess ability: development of a dynamic assessment measure. **School psychology review, 24** (4), 672-693.

Tayler, J. (1996). **Assessment of the predictive validity learning ability battery.** Unpublished M.A. dissertation, University of the Witwatersrand.

Taylor, R.L. & Richards, S.B. (1990). Validity of the estimated learning potential and other measures of learning potential. **Perceptual and motor skills, 71** (1), 225 - 229.

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 philosophy, 24** (4), 184-193.

Taylor, T.R. (1999). Assessing learning potential and integrating it with cognitive skills training. Paper presented at the institute for International's Research's conference, Compliant Assessment Techniques on 24 June.

Thomson, R. (1974). **The Pelican history of psychology.** Middlesex: Penguin.

Thorndike, R.M. (1997). The early history of intelligence testing. In D.P. Flanagan, J.L. Genshaft & P.L Harrison (Eds.), **Contemporary intellectual assessment: theories, tests and issues** (pp. 3 - 16). New York: Guilford Press.

Tzuriel, D. (1992). The dynamic assessment approach: a reply to Frisby and Braden. **The journal of special education, 26** (3), 302-324.

Tzuriel, D. & Klein, P.S. (1987). Assessing the young child: children's analogical thinking modifiability. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 268 - 287). New York: Guilford Press.

Valsiner, J. & Voss, H. (Eds.), (1996). **The structure of learning processes.** New Jersey: Ablex.

Van Aswegen, M. (1997). **The standardisation of a learning potential battery for the selection of poorly qualified employees.** Unpublished M.Com. dissertation, University of Pretoria.

Van Eeden, R. & De Beer, M. (2001). Assessment of cognitive functioning. In C. Foxcroft & G. Roodt (Eds.), **An introduction to psychological assessment in the South African context** (pp. 175 - 208). Cape Town: Oxford University Press.

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.

Visser, A. (2001). **University of Cape Town: Alternative Admissions Research Project.** Retrieved July 19, 2001, from University of Cape Town Web site: http://www.uct.ac.za/depts/aarp/test.htm.

Von der Malsburg, C. (1987). Synaptic plasticity as basis of brain organization. In J.P. Changeux, & M. Konishi (Eds.),**The neural and molecular bases of learning** (pp. 411 - 432). New York: John Wiley & Sons.

Von Hirschfeld, S. & Downs, S. (1992). Testing for trainability. **People dynamics, 10** (9), 24 - 26.

Voss, H. (1996). Learning, development and synergetics. In J. Valsiner. & H. Voss (Eds.), **The structure of learning processes** (pp.17-44). New Jersey: Ablex.

Vye, N.J., Burns, M.S., Delclos, V.R. & Bransford, J.D. (1987). A comprehensive approach to assessing intellectually handicapped children. In C.S. Lidz (Ed.), **Dynamic assessment: an interactional approach to evaluating learning potential** (pp. 327 - 359). New York: Guilford Press.

Vygotsky, L. (1978). **Mind in society: the development of higher psychological processes.** London: Harvard University Press.

Watson, A.S.R. & Ncapayi, M.E. (1998). **The APIL test battery. A predictor of academic performance at University?** Poster paper presented at the annual conferences of the society for student counselling in Southern Africa, 15-18 September, Vista University, Bloemfontein.

Yeld, N. & Haeck, W. (1997). Educational histories and academic potential: can tests deliver? **Assessment and evaluation in higher education, 22** (1), 5-16.

Zaaiman, H. (1996). The quantification of educational disadvantage for selection: necessary? Practical? **Academic development, 2** (2), 113-123.

Zaaiman, H., van der Flier, H, & Thijs, G.D. (1998). Selecting South African higher education students: critical issues and proposed solutions. **South African journal of higher education, 12** (3), 96-101.

Zolezzi, S.A. (1992). **Alternative selection measures for university undergraduate admissions.** Unpublished M.Ed. dissertation, University of the Witwatersrand.

Zolezzi, S.A. (1995). **The effectiveness of dynamic assessment as an alternative aptitude testing strategy.** Unpublished D.Phil. dissertation, University of South Africa.

APPENDIX 1

The following research texts and theses were located on various databases, such as the HSRC's library catalogue, the Nexus database as well as Sabinet. Further inquires were made at the various tertiary institutions where the sources were supposedly held. After several unsuccessful attempts to locate these sources, they were either:

1. Located on one database but not on others
2. Were not indexed by the institution owing to not having received the source to index
3. Were in the process of being indexed but were not available for inter-library loaning at the time
4. Were for various reasons unavailable, even though the text was indexed and located on various databases
5. Were indexed as "current" research projects, but in many instances "current" projects were in reality "completed" projects, although not in every instance.

The following four texts were located on the Nexus database system but were not never completed:

- Boeyens, J. (1987). **Investigation into learning potential, transfer of training and other modifiable aspects of intelligence.** RGN - Psychology Testing .
- Claassen, N.C.W. (1999). **The development of a dynamic assessment instrument**. HSRC Group Education. Psychology, education and training.
- Owen, K. (1995). **Development of an instrument for measuring learning potential**. HSRC Group Education. Manual.
- Owen, K. (1997). **Measurement of learning potential.** HSRC Group Human Resources. Psychometrics.

The following (mostly dissertation) texts were indexed on the **Nexus** database system:

- Andrews, S.G. (1997). **Dynamic assessment procedures in a college of education: a case study.** Unpublished M. Ed. dissertation, University of Natal. (This was indexed as "complete" but was not in stock at that university upon inquiry for inter-library loan.)
- Benjamin, L. (2000). **The effects of the observation of dynamic and static measurement on teachers' perceptions of the learning potential of less academic learners.** Unpublished M.Ed. dissertation, University of the Western Cape. (This text was said to be available at this university but could not be obtained but was however indexed as being "current" and not "complete".)
- Boeyens, J.C.A. (1987). **Investigation into learning potential, transfer of training and other modifiable aspects of intelligence.** Human Sciences Research Council. (This was indexed as "completed" but could not be obtained. However, it must be noted that the proceeding two investigations by Boeyens [1989a, b] might have supplanted the information in this particular text.)
- Bradbury, J. (2000). **The questioning process in the development of knowledge.** Unpublished D.Phil. dissertation, University of Natal. (This was indexed as being "complete" but was unobtainable via inter-library loan.) Although the main focus of this study did not directly hone in on dynamic assessment, it did make use of a sample of students that were chosen for the TTT programme at the University of Natal. The study investigated the particular types of questioning which may serve as mediation between "…historically constituted disciplines of textual knowledge characteristic of the human sciences and the worlds of knowledge and understanding of new, underprepared learners" (http://star.nrf.ac.za/scripts/starfinder.exe/155/nexus.txt). There is stark differentiation between underpreparedness and other sources of failure .
- Fortuin, A. J. (1993). **Mediated learning experience as essential component of the training of primary school teachers in South Africa.** Unpublished M.Ed. dissertation, University of the Western Cape. (This text was said to be available at this university but could not be obtained and has been indexed as being "complete".) In this work, cultural disadvantage is defined in terms of Feuersteinian mediated learning experience. An approach to training learners with cognitive disadvantage is the focus of this study (http://star.nrf.ac.za/scripts/starfinder.exe/155/nexus.txt).
- Potgieter, A. (1997). **The prediction of academic success within the context of an alternative university admission.** Unpublished M.A. dissertation, Rand Afrikaans University. (This was indexed as being "complete" but was not located at this university.) Thus study's focus was on determining the predictive validity of the selection test battery as used at the Rand

Afrikaans University. The APIL as dynamic assessment instrument formed part of the test battery used in this research. The APIL together with matriculation results and the GSAT, explained a greater percentage of variance in academic performance, than did the three separately (http://star.nrf.ac.za/scripts/starfinder.exe/155/nexus.txt).

- Skuy, M.S., Hoffenberg, S., Visser, L. & Fridjhon, P. (1990). Temperament and cognitive modifiability of academically superior black adolescents in South Africa. **International journal of disability, development and education. 37,** 29-43. (Not obtainable as an inter-library loan at the time.)
- Skuy, M.S. (Ed.). (1996). **Mediated learning in and out of the classroom**. University of the Witwatersrand. IRI/Skylight Training. (Not a text that has been referred to in most other texts in the field, but one which would prove valuable. This was unfortunately unobtainable.)
- Skuy, M.S. & Mentis, M. (1993). **Applications and adaptations of Feuerstein's instrumental enrichment programme in South Africa**. University of the Witwatersrand. Internal report. (This was indexed as "complete" but was unobtainable via inter-library loan.)
- Skuy, M.S., Mentis, M. & Dunn, M. (1993). **Illustrating Feuerstein's instrumental enrichment programme.** University of the Witwatersrand. Internal report. (This was indexed as "complete" but was unobtainable via inter-library loan.)
- Skuy, M. S. & Shmukler, D. (1987). Effectiveness of the learning potential assessment device with Indian and Coloured adolescent in South Africa. **International journal of special education, 2**(2), 131-149. (This is an important text within the field, one which was unobtainable.)
- Taylor, I.A. (1998). **A predictive bias investigation of a learning potential and a personality test.** Unpublished D.Phil. dissertation, University of the Orange Free State (This was indexed as "current" and hence not yet obtainable.)
- Taylor, N.E. (2001). **An analysis of dynamic assessment as an alternative to static assessment using the group administration of Feuerstein's LPAD.** Unpublished M.Ed. dissertation, University of Natal. (This was indexed as being "complete" and was located but was not transferred in time to the local academic information service.)

In total 13 research study results in the field of dynamic assessment in South Africa were unobtainable owing to various logistic reasons. (This does not include the four HSRC research results as they were never completed, as mentioned above.) 5 of the 13 were co-authored by Mervin Skuy (University of the Witwatersrand), 3 emanated from the University of Natal, 2 from the University of the Western Cape, 1 from the University of the Orange Free State, 1 from the Rand Afrikaans University and 1 from the HSRC. It must be noted however that 2 studies are still in progress. In addition, of these 13 studies, 5 are Masters' studies, 4 are internal reports, 2 are Doctoral studies and 2 are published articles.