A SYSTEMATIC REVIEW OF HIGHER EDUCATION ADMISSIONS TESTING PRACTICES IN ISRAEL: IMPLICATIONS FOR SOUTH AFRICA

Johan Herman Cronje

Submitted in partial fulfillment of the requirements for the degree of
Magister Artium in Research Psychology

In the
Faculty of Health Sciences,
Nelson Mandela Metropolitan University, Port Elizabeth

January 2009

Supervisor: Prof C. Foxcroft
Acknowledgements

I would like to thank God for the gifts that have been bestowed on my life which lead to
the completion of this study.

My parents, Brian, family and friends for their continued support and faith in me.

Prof Cheryl Foxcroft, my supervisor, for her vision, perseverance, and brilliant input
throughout this study.

Andrea Watson for agreeing to act as the independent consultant for the systematic
review and for finding many of the resources employed in this study.

My colleagues at the Centre for Access Assessment & Research who passionately pursue
the cause of fair and meaningful access to Higher Education for students.

Prof Dalena van Rooyen and Dr Blanche Pretorius for their input and interest in this
study.
# Table of Contents

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>Abstract</td>
<td>x</td>
</tr>
</tbody>
</table>

## Chapter 1  Introduction

1.1 Chapter Overview  
1.2 Preamble  
1.3 Background  
1.4 Definition of Key Concepts  
1.4.1 Admission  
1.4.2 Access  
1.4.3 Disadvantaged  
1.4.4 Multicultural  
1.5 Outlay  

## Chapter 2  Higher Education in South Africa: Admissions Practices and Issues

2.1 Introduction  
2.2 The Context of Education in South Africa  
2.3 Government Initiatives to Transform HE
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.1 Acts and National Plans</td>
<td>12</td>
</tr>
<tr>
<td>2.3.2 Restructuring of the HE Landscape</td>
<td>16</td>
</tr>
<tr>
<td>2.3.3 Changes in the Educational Curriculum at School Level</td>
<td>20</td>
</tr>
<tr>
<td>2.4 Admissions Testing</td>
<td>24</td>
</tr>
<tr>
<td>2.4.1 The National Benchmark Tests (NBTs)</td>
<td>27</td>
</tr>
<tr>
<td>2.5 Concluding Remarks</td>
<td>33</td>
</tr>
</tbody>
</table>

**Chapter 3 Issues in Admissions Testing in Higher Education**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>34</td>
</tr>
<tr>
<td>3.2 Admission Criteria and the Role of Testing in Admissions</td>
<td>35</td>
</tr>
<tr>
<td>3.3 The History of Admissions Testing</td>
<td>38</td>
</tr>
<tr>
<td>3.4 Guidelines for National Admissions Tests</td>
<td>43</td>
</tr>
<tr>
<td>3.4.1 Considerations when Implementing a National Testing Programme</td>
<td>44</td>
</tr>
<tr>
<td>3.4.2 Reliability</td>
<td>48</td>
</tr>
<tr>
<td>3.4.3 Validity</td>
<td>49</td>
</tr>
<tr>
<td>3.4.4 Bias</td>
<td>52</td>
</tr>
<tr>
<td>3.4.4.1 Differences in performance across groups</td>
<td>54</td>
</tr>
<tr>
<td>3.4.4.2 Language issues in testing</td>
<td>55</td>
</tr>
<tr>
<td>3.4.5 Coaching effects and Overlap with the School Curriculum</td>
<td>57</td>
</tr>
<tr>
<td>3.5 Concluding Remarks</td>
<td>59</td>
</tr>
</tbody>
</table>
Chapter 4  Problem Formulation and Research Method

4.1  Introduction 60

4.2  Problem Formulation 60

4.3  Research Aims 63

4.4  Method 64

4.4.1  The Purpose of Systematic Reviews 64

4.4.2  The History of Systematic Reviews 65

4.4.3  The Methodology of Systematic Reviews 66

4.4.4  Advantages and Disadvantages of Systematic Reviews 68

4.4.5  Traditional Literature Reviews and Systematic Reviews 71

4.4.6  Reliability 72

4.4.7  Validity 72

4.4.8  Data Analysis 73

4.5  Objectives of This Study 74

4.6  Steps and Procedure 75

4.7  Reliability and Validity 78

4.7.1  Reliability 78

4.7.2  Validity 79

4.8  Data Analysis 79


Chapter 5 Results and Discussion

5.1 Introduction 80

5.2 Narrative Overview 83

5.2.1 The Educational System in Israel 83

5.2.1.1 The school system in Israel 84

5.2.1.2 Higher Education in Israel 85

5.2.1.3 Discussion 87

5.2.2 The Psychometric Entrance Test (PET) 90

5.2.2.1 Specifications of the Psychometric Entrance Test (PET) 90

5.2.2.2 Discussion 93

5.2.3 Assessment Procedure for the Psychometric Entrance Test (PET) 94

5.2.3.1 Discussion 95

5.2.4 Higher Education (HE) Admissions and the Psychometric Entrance Test (PET) 97

5.2.4.1 Discussion 98

5.2.5 Summative Comment 98

5.3 Systematic Review 99

5.3.1 Reliability 99

5.3.1.1 Statistical methods used to investigate reliability 100

5.3.2 Validity 101

5.3.2.1 Construct validity of the Psychometric Entrance Test (PET) 101

5.3.2.2 Predictive validity for general samples 101
<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.2.3 Predictive validity for cross-cultural samples</td>
<td>102</td>
</tr>
<tr>
<td>5.3.2.4 Predictive validity for different age groups</td>
<td>104</td>
</tr>
<tr>
<td>5.3.2.5 Predictive validity for students with disabilities</td>
<td>104</td>
</tr>
<tr>
<td>5.3.2.6 Predictive validity for tests of ability in first language and second language</td>
<td>105</td>
</tr>
<tr>
<td>5.3.2.7 Statistical techniques used to investigate predictive validity</td>
<td>106</td>
</tr>
<tr>
<td>5.3.3 Bias</td>
<td>107</td>
</tr>
<tr>
<td>5.3.3.1 Cross-cultural bias</td>
<td>107</td>
</tr>
<tr>
<td>5.3.3.2 Gender bias</td>
<td>109</td>
</tr>
<tr>
<td>5.3.3.3 Language bias</td>
<td>109</td>
</tr>
<tr>
<td>5.3.3.4 Bias against students with disabilities</td>
<td>110</td>
</tr>
<tr>
<td>5.3.3.5 Statistical methods used to investigate bias</td>
<td>111</td>
</tr>
<tr>
<td>5.3.4 Psychometric Issues</td>
<td>111</td>
</tr>
<tr>
<td>5.3.4.1 Types of test items included in the Psychometric Entrance Test (PET)</td>
<td>112</td>
</tr>
<tr>
<td>5.3.4.2 Methods for equating different versions of the Psychometric Entrance Test (PET)</td>
<td>112</td>
</tr>
<tr>
<td>5.3.4.3 Equivalence of the paper-and-pencil and computerized versions of the Psychometric Entrance Test (PET)</td>
<td>113</td>
</tr>
<tr>
<td>5.3.4.4 Time limits on the Psychometric Entrance Test (PET)</td>
<td>114</td>
</tr>
<tr>
<td>5.3.4.5 Differences in guessing behaviours on the Psychometric Entrance Test (PET)</td>
<td>115</td>
</tr>
<tr>
<td>5.3.5 Coaching</td>
<td>115</td>
</tr>
</tbody>
</table>
Chapter 6   Conclusions and Recommendations

6.1  Introduction

6.2  Conclusions and Recommendations

6.2.1  Theme one: The admission test (domains, test format and language

6.2.2  Theme two: The implementation of a national admission test

6.2.3  Theme three: The fair use of test and school results for Higher Education (HE) admission

6.2.4  Theme four: Reliability of the national admission test must be established

6.2.5  Theme five: Establish the validity of the national admission test

6.2.6  Theme six: Bias within the national admission test must be researched

6.2.7  Theme seven: Psychometric aspects related to the national admission test should be explored

6.2.8  Theme eight: Preparation for the national admission test

6.3  Limitations of the Research and suggestions for Future Research

6.4  Concluding Remarks

References

Appendix A

Appendix B

Appendix C
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Research outputs included in the systematic review and narrative overview</td>
<td>81</td>
</tr>
<tr>
<td>Table 2</td>
<td>Themes extracted from the narrative overview and systematic review</td>
<td>99</td>
</tr>
</tbody>
</table>
Abstract

Internationally, the practice of admissions entry testing in Higher Education (HE) has gained momentum over the past few decades, sparking ongoing research on its effectiveness. On a national level, three factors have signalled a timely evaluation of admissions practices at HE institutions. Firstly, changes in the school curriculum and the new format of the National Senior Certificate (NSC) have impacted on the admissions criteria of HE institutions. Secondly, with the merging of HE institutions an alignment in admissions practices between the institutions involved were necessitated. Thirdly, South Africa has embarked on the development of National Benchmark Tests (NBTs).

The primary aim of this research study was to develop a set of recommendations to guide admissions testing practices in the South African HE context. These recommendations had to place special emphasis on the multicultural and multilingual context of this country. To achieve this aim a systematic review was conducted on HE admissions testing in Israel, as it was identified as a multicultural and multilingual country that had successfully implemented national HE admissions testing. More specifically, a retrospective systematic review was performed on research regarding the national HE admissions test, the Psychometric Entrance Test (PET), used in Israel. The systematic review also contained a narrative overview on the educational landscape in Israel and the specifications of the PET, from which themes were also extracted.

Eight broad themes emerged through the systematic review and narrative overview as being of critical importance to an effective national HE admissions test. These were the components of the test, the implementation of the testing programme, the method in which the test results are used to make HE admission decisions, the reliability
of the test, the validity of the test, bias inherent in the test, other psychometric aspects related to the admissions test, and the effect of coaching or specialized preparation on test results. These themes, together with their sub-components, were used to develop eight recommendations that can guide the development and implementation of the National Benchmark Test (NBTs) in South Africa. Both the themes that emerged during the systematic review and narrative overview as well as the recommendations that were made to guide the development and implementation of a national admission test, represent an important contribution to the field of admission testing and decision-making in South Africa.

Key words: Admissions test, Higher Education admission, Israel, National Benchmark Tests, Psychometric Entrance Test, South Africa, Systematic review.
Chapter 1
Introduction

1.1 Chapter Overview

In this chapter the key issue addressed in this study is discussed in the form of a preamble. It is followed by general background information that provides a motivation for the problem at hand. Some of the key concepts that are used in the study are defined. Finally the layout of the rest of the study is presented.

1.2 Preamble

While various universities in South Africa use tests to guide admissions decisions, the possibility of developing a national admissions test has been on the agenda for a number of years (Griessel, 2003). The purpose of this study is to provide recommendations to guide the development and implementation of a national test for the purpose of admission to Higher Education (HE). The central problem as viewed by this study is that such a project should be based on a solid research foundation of prior experience and international best-practice guidelines. Admission testing is a high stakes endeavor and the potential issues that may arise are compounded within multicultural and multilingual environments. When an environment has been shown to be vulnerable to discrimination and unfair opportunities in terms of access, accurate research becomes critical. The present study sought to provide this research foundation by deriving lessons from an established national HE admissions test in Israel as well as international
implementation guidelines. In the next section the background of the admissions dilemma experienced by HE institutions internationally is discussed.

1.3 Background

Problems around the distribution of valued resources, including access to educational opportunities, preoccupy most societies (Klitgaard, 1986). Education systems internationally have faced a barrage of challenges in the past few decades. Increased democratization and integration within societies have placed an emphasis on equity and redress of past inequalities. The advent of the information age has led to a new knowledge driven economy. The importance of a workforce skilled for the production sector has been replaced by the need for an educated workforce, with specialized skills, much of which is gained through continued post-school education (Griessel, 2003). Other important developments were computers and information technology, technologies that have impacted greatly on people and nations. Computers and the internet have brought more information to empower the individual in different ways, from greater financial independence to more control over their vocational direction. Economies across the world have been linked, multinational corporations mushroomed across the globe and the resultant globalization of the planet has broadened employment opportunities in different countries as well as created a demand for more transferable skills. The changes within the global economy have also affected the employment patterns of workers. Traditionally, employees tended to have a stable relationship with their employers for a period that could last the duration of the employee’s working career. Workers stayed with their employers for decades. The greater fluidity of the international economy has resulted not
only in shorter relationships between employees and employers, but also an increase in the number of older students in the HE sector, seeking to update existing skills, or embracing new skills to effect career changes.

The aspects above, amongst others, have impacted on educational systems. Access to HE has changed from the preserve of the privileged to a right for many. The number of HE applicants has surged, whilst the changes within the HE curriculum have resulted in an ever-widening gap between school outputs and HE admissions requirements (Griessel, 2003). Relying solely on school results to make HE admissions decisions presents an increased barrier to participation for especially those from disadvantaged communities. Admissions authorities worldwide have increasingly been implementing more complex systems to distinguish between applicants. A major additive component to the admissions process has been aptitude and achievement testing. These tests are used in conjunction with school results to admit students, place learners within appropriate programmes and to identify the developmental needs of students. Great demands are placed on admissions tests to bridge the gap between school and HE, by assessing skills that are required for HE studies, whilst not overlapping the domains already assessed by the school curriculum (Griessel, 2003).

Another challenge for admissions authorities has been the increasing numbers of applicants from different cultural groups within a country as well as from international destinations. According to Van de Vijver (2000), admissions tests have increasingly been used to compare students from different cultural or language backgrounds. The advance of testing has not gone unchallenged, especially in the light of growing awareness that test constructs may differ across cultures and languages, that norms developed in one
group may not be transferable to others, and that nuisance factors specific to individual cultures may influence test results (Koch, 2005).

Internationally and nationally, the demands placed on HE are ever-changing, placing a considerable burden on the development and implementation of fair and efficient admissions systems. This setting provides the backdrop for the present study. In the next section some of the concepts that are used in this study are defined.

1.4 Definition of Key Concepts

1.4.1 Admission

Harman defines admissions as “the process from when a potential student develops an interest in entry to higher education until enrolment in a particular institution and course of study takes place” (1994, p.318). Admission is generally granted based on certain criteria, such as school performance and admissions test results, and may extend to other forms of assessment, such as applicant interviews and portfolios.

1.4.2 Access

Access relates to the opportunity that exists for members of a specific population group to gain entrance to study opportunities as fully enrolled students (Harman, 1994).

1.4.3 Disadvantaged

For this study the term refers to educational disadvantage as a result of inadequate access to educational resources, as a result of economic and socio-political factors (Zaaiman, van der Flier & Thijs, 1997).
1.4.4 Multicultural

This relates to environments where groups from different ethnic, cultural or language backgrounds coexist. For testing purposes it also refers to tests that are used for different groups regardless of whether the tests have or have not been translated or adapted for each group. In some cases the researcher may also refer to this as “multilingual” or “cross-cultural” although it is acknowledged that these terms may not always be fully interchangeable (Koch, 2005).

1.5 Outlay

The rest of the dissertation is structured as follows. In Chapter 2 the evolving educational context of South Africa is discussed. Specific focus is given to the social and political developments that have shaped the educational environment and the resultant changes in the school and HE system. The changes in HE admissions practices, and attempts to broaden access in a fair and equitable manner through admissions testing are presented. Chapter 2 ends with a discussion of the National Benchmark Tests (NBTs) Project, which is the result of calls to provide placement and admissions testing through a nationally managed structure.

Chapter 3 highlights important aspects of admissions tests, by looking at those elements that are important for fair testing practices, as well as international guidelines that have been developed to assist test developers and users in this task.

The rationale for the current study is presented in Chapter 4. The methodology which was employed in this study is fairly unfamiliar within traditional psychological research, and is therefore discussed in greater detail.
In Chapter 5 the results of the research are examined, followed by the relevant discussion of the findings.

In Chapter 6 the conclusions and recommendations reached by the study are presented. The limitations of the findings and suggestions for future research are discussed.
Chapter 2

Higher Education in South Africa: Admissions Practices and Issues

2.1 Introduction

In Chapter 1 the various developments that have impacted on Higher Education (HE) admissions worldwide were discussed. Some of the trends in HE in the 1990s included massification, democratization, globalisation, internationalism, access, lifelong learning, outcomes-based education and quality assurance (Blunt, 2005). South African institutions have continuously faced similar challenges, many of which have been compounded due to its colonial past and former apartheid system. These phenomena have impacted on HE from student admission to eventual graduation, including diverse aspects such as, student and institutional governance, curriculum development, lecturing and content delivery styles, course assessment frameworks and so forth. The most important purpose of this chapter is to highlight the impact on admissions practices in HE and the response from institutions to evolve in a changing world. It will touch on some key developments in the country’s education system. Specifically, a contextual overview of education in South Africa is provided, followed by some of the government initiatives to transform the educational system through formal acts, policies and plans, changes in the school curriculum, and the restructuring of the HE landscape.

According to the Department of Education (1997) a successful admissions practice is one that admits a high percentage of those applicants which would succeed, while not admitting as few applicants as possible that would also succeed if given the opportunity. In attempting to realise this ideal HE institutions have adopted various
approaches, but a common development has been the inclusion of a standardised test as part of the admissions process. As a result of the increase in admissions testing a more recent occurrence has been the development of national benchmark tests (NBTs). These will be discussed towards the end of this chapter.

2.2 The Context of Education in South Africa

A historical overview of the educational landscape in South Africa is divided clearly by the first democratic national election of 1994. Before this election the educational system was divided along cultural lines and after the election the newly elected government of national unity set about integrating the various systems. Although some changes had already begun before the election, the development of the country after this watershed election was marked by focussed efforts towards integration of society in all spheres.

Education does not operate within a vacuum. It is influenced by the larger socio-political and historic context that surrounds it (Seymour, 2002). The effects of the apartheid system in South Africa were evident in its impact on education (Skuy, Zolezzi, Mentis, Fridjhorn & Cockcroft, 1996). The Bantu Education Act of 1953 provided the legal foundation for separate educational systems for different ethnic groupings (Act no. 47 of 1953). A total of 19 Separate Departments of Education were created to cater for different groups (Koch, 2005). These operated in isolation, were funded in a separate manner, and provided learners with a different educational curriculum and learning experience (Goduka, 1999). Despite being billed as separate but equal, the funding framework for schools and HE institutions were criticised for favouring white
institutions, whilst leaving especially black institutions under-funded (Naidoo, 1998). In the period 1988 to 1989, for example, R656 was spent on education for a black child, while R2 882 was spent per white child by provincial education departments (Huysamen, 2000). Access to HE was especially difficult for black students because universities set aside for their education were situated in the former homelands. The cost of HE for those learners, the majority of whom were from families who were economically disadvantaged by the apartheid system, combined with other obstacles, meant that furthering one’s education was a distant dream for many. Generally contact between learners from different ethnic groups was very limited, with black students being prohibited from entering white institutions. In the late 1980s this legal barrier was removed. The separate Departments of Education continued to exist, each administering a different matriculation exam for its learner base. The different exams, which were disproportionate in difficulty levels, made accurate comparisons based on matric performance impossible for university applicants from different ethnic groups (Zaaiman, van der Flier & Thijs, 1997). This, together with the poor schooling system of black students and the high academic entrance requirements of white HE institutions limited the opportunities for black students to gain access to highly regarded white HE institutions (Cooper, et al., 1992; Naidoo, 1998). The late 1980’s marked a period of social unrest, during which the marginalised majority black population revolted against the apartheid system. During this period the education of many black pupils and students was disrupted.

With the abolition of apartheid and the elections of 1994 came a renewed interest in the importance of education as a means to uplift all communities and to break the cycle of poverty experienced by millions of previously disenfranchised South Africans. The
reason for this was that the newly elected government recognized education as one of the pillars of not only a democratic society, but also as fundamental for a diversified economy and sustained economic development (Department of Education, 1997).

In 1995, President Nelson Mandela appointed the National Commission on Higher Education (NCHE) to investigate the means of transforming, reconstructing and developing the HE sector. The NCHE found that the world economy that South Africa had been reintroduced to, after years of international sanctions, had a rapidly increasing demand for highly skilled workers (National Commission on Higher Education, 1996). This was due, in part, to the decline experienced by the primary sectors of the economy and the growth of the service sector (Mochela, 2006). The advent of the information age led to a new knowledge economy that relied on specialist training (Department of Education, 2001). In order to remain competitive in the global village South Africa had to move with the times by boosting participation rates in further and higher education. This, among other things, resulted in the spotlight falling on HE together with issues related to access and admission to educational opportunities. The NCHE asserted that a fundamental principle within a transformed HE system had to be to “ensure equity of access and the possibility of success to those – irrespective of race, colour, gender, age or class – seeking to realizing their potential through higher-level education and learning” (1996, p.10).

The fall of apartheid saw many black students resuming their education and the country witnessed a growing demand for admission to HE. The removal of former stumbling blocks resulted in an increase in the number of black learners applying to especially historically white universities. Former white English medium universities
experienced a growth in applications by black scholars between 1993 and 1999 of 10 000 (94%), whilst the number for white Afrikaans medium universities surged by 56 000 (1117%) at the same time (Council on Higher Education, 1999). Enrollment figures at public HE institutions grew from 473,000 in 1993 to over 741,000 in 2006 (Department of Education, 2006; Subotzky, 2008). This constituted an increase of over 56 percent in 14 years. Throughout, the onus has been placed on the HE sector to increase the participation rates of non-traditional and disadvantaged students in particular on the one hand, as well as to produce graduates in fields of study that meet the needs of the labour market on the other (Harman, 1994; Watt & Paterson, 2000). In 1999, Huysamen and Raubenheimer captured the challenge of HE by stating:

Admissions authorities at South African universities are faced with several dilemmas. On the one hand, socio-political considerations dictate that learners from educationally disadvantaged backgrounds be admitted at a higher rate than before. On the other hand, these learners’ educational background may fail to prepare them to cope with the demands of tertiary education. (p. 171)

Government, being mindful of this situation, realised that the problems experienced by HE were the result of an inefficient educational system as a whole (Department of Education, 1997). What was required was more than just a focussed intervention, but rather a large-scale restructuring and integration of education from primary and secondary school level through to HE. Some of the major developments that arose in line with this are discussed in the next section.
2.3 Government Initiatives to Transform HE

In this section some of the educational reforms in South Africa over the last decade are highlighted, especially in terms of their impacts on HE. The aspects that are covered include government Acts and national plans, the restructuring of HE, and the transformation of the school system.

2.3.1 Acts and National Plans

Although government has produced a comprehensive legislative and planning framework for education, two stand out as landmarks for their future impact on HE. These are the Higher Education Act of 1997 and the National Plan for Higher Education of 2001.

In 1997, government introduced the Higher Education Act (Act no. 101 of 1997) to regulate HE. The Act set the agenda of the new post-apartheid government for HE. The purpose of the Act was:

To regulate higher education; to provide for the establishment, composition and functions of a Council on Higher Education; to provide for the establishment, governance and funding of public higher education institutions; to provide for the appointment and functions of an independent assessor; to provide for the registration of private higher education institutions; to provide for quality assurance and quality promotion in higher education; to provide for transitional arrangements and the repeal of certain laws; and to provide for matters connected therewith (Act no. 101 of 1997, p.2).
From the above it is clear that the Act encompassed many aspects of HE and that the intention was to effect global change in line with the vision of the new government. It was a large step towards transforming HE from a divided and rigid system into a diverse, integrated and flexible system that allowed for broadened access to learners and improved movement or articulation within and between institutions and programmes.

Three focus areas of the Act later proved to have major implications for HE Admissions. The first was that institutions were empowered to open up learning opportunities, especially for those who would have struggled to gain entry before (Section 37). Until that time access specifically to universities hinged on performance in the national exit level school exams. The minimum entry criteria to university was a pass at a specific level called matriculation endorsement. The Act specifically stated that HE institutions were allowed to implement additional admissions criteria, on condition that these criteria served to broaden, rather than restrict access (van der Merwe & de Beer, 2006). It opened the gateway for HE institutions to establish alternative pathways for entry to programmes.

The second area was that whilst HE needed to broaden access to previously disadvantaged students the Act made it the responsibility of institutions to adopt measures to redress past inequalities in terms of access (Section 37.3). This implied not only an opening up of doors, but an active pursuit to correct past inequalities and rebalance the distribution of the HE student population.

The third area was that of mergers between institutions. According to the Act, the Minister of Education could merge two or more public HE institutions into one (Section 23.1). No specifications were provided in the Act for basis on which these mergers would
occur, but it did empower the Minister of Education to reshape the landscape of HE institutions.

The National Plan for Higher Education (NPHE) of 2001 followed on the policy framework, goals, values and principles included in the White Paper on Higher Education (Department of Education, 1997). The intention of both documents was to develop a HE system that would amongst other things:

- promote equity of access and fair chances of success to all who are seeking to realise their potential through higher education, while eradicating all forms of unfair discrimination and advancing redress for past inequalities (Department of Education, 2001, p7).

The NPHE tied in with the HE Act and addressed three critical issues in HE. These were access, throughput and transformation.

Concerns over access and participation followed the second half of the 1990s, a time during which optimism reigned about the future of HE. After the 1994 elections participation rates in HE steadily climbed and appeared set to continue. The reason for the sharp increase in enrollment figures was however due to the number of learners resuming and completing their education that was disrupted by the unrest of the late 1980s. As the bottleneck eased and the applicant base for HE returned to its former market of school-leavers it became clear that participation rates would remain low. The NPHE responded by stressing the need to increase participation rates in HE from 15% to 20% and especially to strive towards the composition of learners being more representative of the ethnic makeup of our society. It also called for more learners to be enrolled in business and commerce, science, engineering, and technology programmes, as
the future labour demands in these fields were likely to be the highest (Department of Education, 2001).

Added to declining participation rates was the problem of throughput and graduation rates, which plummeted. This trend was evident towards the end of the 1990’s when, despite encouraging enrollment figures, 25 000 fewer graduates were produced in 1998 (Council on Higher Education, 2000). The decline continued and a study by the HSRC of seven universities found that 40% of students from a cohort of 2000-2002 dropped out in their first year (Letseka, 2007). Furthermore only 15% of students obtained their qualification in the minimum amount of time. The issues of participation and throughput placed an extra responsibility on HE Admissions to strike a balance between selection in line with equity targets, along with the outcome of boosting throughput, or access as participation together with success (Akoojee & Nkomo, 2007). Narismulu (1994) recognised the sensitivity of this issue by stating:

Neither the country, nor higher education institutions can afford irresponsible admissions which may actually undermine real equity. The principle of equity does presume high social quality particularly in terms of human resource development needs of this country. Furthermore, access and quality are not necessarily competing, antagonistic values. It has been argued that access without quality is a cruel deception, while quality without access is a betrayal of the ideal. (p.39)

Mindful of this the NPHE aimed for an increase in throughput rate from 15% to 30% (Department of Education, 2001) thereby placing the focus on meaningful access linked to student retention and graduation.
The South African government has remained committed to education through initiatives and the nature of the funding provided. For the 2008 national budget a sum of R123 billion rand was set aside for education, which constitutes 16.9% of government expenditure. Of this, R13.1 billion was earmarked for HE. This makes education the single largest expenditure within the social services domain of the budget. Such an expense must be justified through equal return, but the issue of participation and throughput remains a large stumbling block to success for the educational system.

Mindful of the need for throughput, the NPHE again stressed the need to identify more clearly those learners who are academically ready to successfully tackle HE programmes and those who could succeed but require development. The challenge facing HE institutions remains that of gathering appropriate information on applicants to identify future graduates, as well as to facilitate the development of needs-based programmes and academic support initiatives in line with student profiles.

The third critical issue addressed by the NPHE was the implementation of transformation in HE (Department of Education, 2007). The NPHE had at its core “the development of a single, national, coordinated higher education system, which is diverse in terms of the mix of institutional missions and programmes” (Department of Education, 2001, p.14). The task of creating this new system was addressed by proposed institutional mergers, which is covered in the next section.

2.3.2 Restructuring of the HE Landscape

With the changes taking place in education and the aims of changing the future outlook of disadvantaged citizens of South Africa, the face of tertiary education, as a
means of equipping the skilled workforce has changed. Tertiary institutions have had to evolve to face the challenges of a more diversified learner population (Gultig, 2000).

Traditionally, public HE in South Africa was conducted through 36 institutions which were divided into two main types. On the one hand, universities presented undergraduate degree-type qualifications that were generally more broad-based and took three to four years to complete. Undergraduate degrees lead to postgraduate honours, masters and eventually doctoral degrees. On the other hand, technikons provided diploma-type qualifications that were more occupationally focussed and prepared students for the world of work. These qualifications tended to be shorter in duration than degree courses. The two different types institutions tended to operate independently of each other, with an entire institutional network and culture of their own. In post-apartheid South Africa these institutions, which were created in the segregated apartheid system, were expected to function in a synchronised manner to meet the HE needs of the country. However, the Department of Education’s analysis was that the system had been unable to respond to the challenges of a changing South Africa. Despite the political and societal changes since the election of the Government of National Unity in 1994 HE had remained fragmented along racial lines. Radical changes were therefore proposed to achieve three fundamental objectives of a non-racial, non-sexist and democratic HE system (Department of Education, 2002a). The first objective was social justice through redress of the social and structural inequalities caused by apartheid. The second was to assist HE to produce graduates that would enable the economic and social growth envisioned for the country, especially in the face of globalization. The third objective was
to restructure HE to ensure that the limited resources were effectively and efficiently utilised (Department of Education, 2002a).

The most significant change within the HE landscape was that in November 2002 government approved a proposal to restructure HE through mergers between institutions (Department of Education, 2005). These mergers were initiated from 2004 to 2005 and resulted in the number of institutions being reduced from 36 to 23. As stated earlier, technikons and universities existed in different paradigms. It was therefore surprising when the mergers not only occurred between institutions of the same type, but across the divide. An example of this was that the University of Port Elizabeth and the Port Elizabeth branch of Vista University was merged with the Port Elizabeth Technikon. This was an example of a historically white university being merged with a historically black university, together with a technikon. During the mergers the term “technikon” was dropped and three types of universities were created, namely universities, comprehensive universities and universities of technology. After the restructuring some of the traditional universities continued to exist as before. The merged universities and technikons became known as comprehensive universities, that would offer both diploma and degree qualifications, with eventual articulation routes between different courses. Other former technikons were granted degree-granting status to provide continued education within its former programmes and became universities of technology (Department of Education, 2002a).

Comprehensive universities were a new addition to HE and little information was provided about the identity and characteristics of these institutions. Institutions were given autonomy to restructure the new institution, but the idea was clearly that the former
separate institutions should be replaced by a single, fully integrated institution. Substantial effort was required. The former Technikon and University structures were so different that integrating them would require a major rethink in many spheres (Blunt, 2005). Of major importance for incoming students would be the access routes to obtain admission into these institutions, as well as the philosophical approach to learning that would emerge from the new programme offerings. It was evident that such a hybrid institution needed to carefully formulate its programme mix to allow for maximum articulation between qualifications. Admissions requirements, criteria and processes would also require thoughtful planning to allow for a range of entry requirements, as well as facilitate the placement of applicants in appropriate programmes. Indeed, it was argued that in the start-up phase of merging institutions into a comprehensive institution, issues related to admissions practices would be a priority that needed to be urgently addressed (Department of Education, 2003).

As HE was coming to terms with its new structure, it also had to contend with the broader changes that were occurring within the educational landscape of South Africa. The school system and curriculum had been changing progressively from grade R through to grade 12 and it became clear that HE would soon be faced with a cohort of first-year applicants with an unfamiliar educational background. The changes in the schooling system would have a dramatic impact on HE, especially with regards to admissions practices. This will be discussed in the next sub-section.
2.3.3. Changes in the Educational Curriculum at School Level

In September 1997, the Heads of Education Departments proposed the Statement for the National Curriculum for Grades R to 9. This was accepted by the Council of Education Ministers on 29 September 1997 as three separate documents for Foundation Phase, Intermediate Phase, and Secondary Phase (Department of Education, 2002b). The curriculum was framed within outcomes-based education, which is a learner-centred model that values the process of learning as much as the content (Killen, 2000). Subjects (learning areas) were developed around specific outcomes and assessments were geared toward determining the extent to which competence had been attained with respect to these outcomes. The whole philosophy was first introduced by the South African Qualifications Authority (SAQA) in 1995, and was an attempt to provide quality assurance for an educational system that was marred by the different levels of performance and delivery inherited from the various Departments of Education (SAQA, 2000).

The new curriculum from Grades 1 to 9 replaced the old system up to Standard 7 and introduced a standardised curriculum for pre-school in the form of Grade R. Apart from unfamiliar new content and subject names the curriculum introduced the new subject of Life Orientation. The domain of life skills was not a formal fixture within the previous curriculum, but through Life Orientation with its focus on health promotion, social development, personal development, physical development and movement, and orientation to the world of work, this changed.

The curriculum was scheduled for gradual implementation from 1998. As implementation was gradual, it challenged teachers to function concurrently within the
new system in some grades and the old in others. In June 2000 the Council of Education Ministers agreed that some revision of the curriculum should be undertaken and upon completion of this process the Revised National Curriculum for Grades R to 9 was implemented in 2002 (Department of Education, 2002b).

In the old education system a pupil had the choice of completing senior secondary education through one of two avenues. The first was to continue in high school and complete Standard 8 to 10 (Grades 10 to 12). The learner had the choice of selecting at least six subjects on either a higher or standard grade. At the end of Standard 10 (Grade 12) learners completed the matriculation exams, which was set and controlled by the government, and which formed the sole basis for passing matric. A pass-level in the exams resulted in a matriculation certificate, called a Senior Certificate (SC), being awarded to the learner. If the learner sat for, and performed sufficiently well in, at least four higher grade subject exams and the subjects taken were from a predetermined mix, matriculation endorsement would be obtained. This was the minimum statutory requirement for entry into a university bachelor’s degree (Foxcroft & Davies, 2008).

The second avenue was to continue education at a Technical College, which offered a curriculum geared towards trades and occupations (du Toit & Craemer, 2000). The curriculum focussed on practical skills employed in the working sector, but a learner desiring further HE studies could proceed from levels N1 to 6 to develop skills which at a N3 level could lead to Technikon studies and at N6 level could articulate into a relevant University course.

On 20 July 2005, government gazetted a new system to replace the former Standard 8 to 10 and N-level systems. Instead of the former “Standards” learners would
be completing Grades 10 to 12, which would follow from their prior outcomes-based schooling. The new system was called the National Senior Certificate (NSC). Subjects would no longer be taken at higher or standard grade and learners would be taking seven subjects, instead of the former six. Six of the subjects would have a credit value attached of 20 points each, whilst Life Orientation was 10 credits. Four of the subjects were compulsory and three could be selected by the learner. The four compulsory subjects were two languages (home and first additional languages, one of which must be the language of instruction of the institution), life orientation, and either maths or maths literacy (Foxcroft & Davies, 2008). From the subjects offered in the NSC, a list of designated subjects was created. Pupils who were interested in pursuing degree studies were compelled to take at least one subject, but encouraged to take more, from this list, to make up the seven required subjects.

The old system of matriculation endorsement was replaced. Depending on performance the NSC could be passed at levels that allowed for entry into certificate, diploma, or degree programmes. Another difference between the NC and NSC is that in the NC the final exam formed 100% of the mark, but in the current NSC the internal assessment (or class mark) contributes 25% and the external exams 75% of the final mark per subject (Department of Education, 2005). The NSC was implemented in 2006, with the first batch of pupils completing the NSC at the end of 2008.

The Technical College system was also revamped. Institutions were renamed Further Education and Training (FET) Colleges and the N1 to 6 system was replaced by the National Certificate Vocational (NCV). The NCV followed a similar format to the NSC. The current NCV curriculum remains in line with the former N-system subjects. A
learner can follow a curriculum in one of twelve fields. They are Management, Marketing, Information Technology and Computer Science, Finance, Economics and Accounting, Office Administration, Electrical Infrastructure Construction, Civil Engineering and Building Construction, Engineering and Related Design, Primary Agriculture, Hospitality, Tourism, and Safety in Society (Department of Education, 2008). A major difference between the NSC and NCV is that only one language is taken as a compulsory subject, namely first additional language. This has implications for access to HE as the current NSC criteria for degree entrance requires two languages, one at home language level and another at first additional language level. How this matter can be overcome to provide improved articulation between FET Colleges and HE is still under investigation. The NCV was implemented in 2007 and as it takes three years to complete the first graduates are expected at the end of 2009.

The changes in the school curricula, especially the NSC and NCV impact greatly on HE. Institutions will have to reassess the school curriculum and assist incoming learners in those areas where their NSC education does not immediately prepare them for the current HE curriculum. Furthermore, new admissions criteria needed to be developed for the NSC and admissions departments will have to run two processes concurrently to accommodate applicants from both the old matric and new NSC system. By the end of 2009, admissions requirements for the NCV will also have to be developed and students from this curriculum accommodated into HE. In order to compare applicants from the various systems, especially in courses with very limited spaces, a common benchmark will be valuable. The next section deals with the developments around admissions and benchmark testing in South Africa.
2.4 Admissions Testing

With the booming applicant base and limited spaces available in programmes HE institutions have had the advantage of being able to be more selective. The HE Act of 1997 provided a legal framework that the basis on which admissions decisions are made should be defensible, so as to ensure that the goal of broadening access to previously underrepresented groups could be met.

South African HE institutions have tended to rely only on matric results for admissions purposes. The validity of this practice for all segments of South African population has been questioned however (Miller & Bradbury, 1999). The reason for this is that learners from disadvantaged backgrounds may have been inadequately prepared for the final exam because of the inferior education and lack of opportunities they have been exposed to. Even within the new integrated NSC system many schools still lack the resources to prepare pupils for the final exams. In the past the sole reliance on matric results were accused of contributing to the disproportional representation of learners at university level (Pavlich & Orkin, 1993). Despite the changes in both the school and HE systems, this disproportional representation could continue.

One argument against the use of matric results are that they are one-dimensional and are a greater reflection of educational background than the potential of the learner to succeed at HE studies. Further support of this argument is that a number of studies have cast doubt on the usefulness of matric results in predicting success at tertiary studies (Greyling & Calitz, 1997). This is especially true for black learners from disadvantaged schooling backgrounds (Badhsa, 1992; Sharwood & Rutherford, 1994). A critical interrogation has found that matriculation results alone did not suffice as a basis for
admissions decision (Koch & Foxcroft, 2003). It does not provide a clear enough picture of applicants and could contravene the HE Act by placing applicants from underprivileged backgrounds at a disadvantage for entry into HE (Gultig, 2000; Nel, 1997).

In an attempt to implement the HE Act and to add value to matric results institutions have adopted different mechanisms to distinguish between applicants. These have included weighted and unweighted matriculation point rating systems, admission tests, questionnaires, portfolios and interviews (du Toit & Craemer, 2000; Mochela, 2006). Of these, admission testing has grown in popularity because it provide an equal benchmark of applicants, is easily administered in a group setting, can be outsourced if needed, is quantifiable, and can be easily researched to prove its effectiveness (Murphy, 2004; Plug 1996; Rust & Golombok, 1999). The latter is important, as national and international research have underlined the importance of an institution’s internal research in guiding the development and refining of admissions and selection procedures (Foxcroft, et al., 2000; Potter & Van der Merwe, 1993; Smith 1992).

Despite the increased use of admissions testing most institutions have developed testing programmes independently with little interaction and cooperation. Some used existing measures whilst others preferred to develop their own (Koch, 2005; Seymour, 2002). When a decision is made to employ an admissions test the question naturally arises as to what should be assessed. In South Africa two broad domains have emerged, namely cognitive and non-cognitive factors (Seymour, 2002).

Testing of cognitive factors focuses on measuring skills or abilities related to academic success. South African HE institutions have developed or employed three
different sub-domains. The first is to measure those cognitive abilities that relate directly to the academic context of HE. These tests tend to centre on proficiency in language (usually the language of teaching and learning of the institution) and numeracy/mathematics, although some institutions also developed measures in specific subject fields, such as science and biology. Examples of competency-based tests include the Access Assessment Battery (AAB) at the Nelson Mandela Metropolitan University, the Placement Test in English for Educational Purposes (PTEEP) at the University of Cape Town, and the Access Tests (AT) Battery at the University of Stellenbosch (Seymour, 2002; University of Stellenbosch, 2008).

The second type is to measure general scholastic ability. The General Scholastic Aptitude Test (GSAT) has been widely used (Van der Walt, 2000). Another example is the Academic Aptitude Test, which was used by the University of the Free State as well as the former Port Elizabeth Technikon. The third type of cognitive test used is one that taps learning potential instead of developed competencies. Assessing potential is very difficult and therefore this sub-domain has been implemented to a lesser extent than the previous two (Koch, 2005). The Ability and Processing of Information and Learning Battery (APIL) at the former Randse Afrikaanse Universiteit (now the University of Johannesburg), and the Potential Index Battery (PIB) at the former Technikon Pretoria (now the Tswane University of Technology), serve as examples of this (Koch, 2005; Seymour, 2002).

Institutions have measured the non-cognitive factors related to academic studies through a variety of personality, career interest, and motivational questionnaires such as the Sixteen Personality Factors (16 PF) Questionnaire and the Self-Directed Search (SDS)
career interest measure (Seymour, 2002). Many institutions constructed a test battery consisting of both cognitive and non-cognitive tests to aid in the selection process of new applicants.

2.4.1 The National Benchmark Tests (NBTs)

The initial forays into the domain of access assessment eventually culminated in the construction of the National Benchmark Tests (NBTs), a project which was conceptualised by Higher Education South Africa (HESA) in 2004. This project was initiated to pool the experience and expertise of test developers from HE institutions across the country in order to develop a national access and admissions test.

The purpose of the NBTs, which are still being developed, is to assess components which do not overlap with the school-leaving examinations, but add an important dimension regarding the competency of school-leavers to successfully tackle HE studies. The tests’ results will be used to provide extra information about learners, which could be useful to HE institutions when making admissions and development recommendations for applicants (HESA, 2006).

There are three assumptions that underpin the necessity for the NBTs. These are:

1. The variability with which current school results reflect the true knowledge and competencies of learners.

2. The need for HE to have a greater understanding of the competencies that underpin academic success, and the need for developing these in learners entering the HE sector.
3. The need to establish stable benchmarks for the admission of learners, especially in the light of the introduction of the new National Senior Certificate (NSC) (HESA, 2006).

Most HE institutions in South Africa currently have a system of admissions testing in place that applicants have to complete as part of the admissions process. The drawback of each institution having its own assessment battery is that applicants who apply to more than one institution have to travel to write different tests for each institution applied to. This lengthens the admissions process and involves extra costs for applicants. This contributed towards the argument for the development of the NBTs which could be used by all HE institutions and only needs to be completed once by a test taker.

A further benefit of the NBTs is that it would serve as an across the board benchmark for all applicants. Admissions departments are facing an increasingly difficult task in fairly distinguishing between applicants from the ever-diversifying pool. Traditionally the majority of applications were received from recent school leavers. As from 2009 these applicants will present with either SC or NSC certificates, whilst as from 2010 the NCV will be added into the mix. On top of this the mature learner market is emerging as an ever-growing base, as more and more adults either change careers or decide to upgrade their qualifications. These applicants would present with qualifications and school-leaving certificates stretching across many years, and even educational departments. South Africa also has a large and burgeoning international student population that in 2006 already totaled 53 773 or over 7% of overall HE student numbers
International applicants can present admissions with a large variety of school leaving qualifications and may not have been exposed to English as a language of teaching and learning. Comparing applicants from these diverse backgrounds for courses with limited spaces is very difficult. Such a task can be aided by the introduction of a stable, centrally established benchmark. The NBTs are designed to fulfill this important task (HESA, 2006).

The NBTs are still in the item and testlet piloting phase of test construction, but from available examples it appears that the tests will employ a multiple choice item format. What is known about the tests is that they will include measures of academic literacy, quantitative literacy, and cognitive academic mathematical proficiency. The tests will initially be in English. Although less than 10% of the population speaks English as a first language it is the language of teaching and learning at most HE institutions. The test constructors do not view this as an unreserved support for English as the only viable medium for the tests. It merely serves as a starting point and keeps the door open for future parallel versions of the NBTs in other languages South African languages (HESA, 2006). As for the subtests, each is designed to address a specific question, which is linked to competencies that HE students need to be proficient in, in order to cope with the academic demands of first year programmes (HESA, 2006). More specifically the tests are:

a) **Academic Literacy:** This test is designed to assess the core competencies that first-year students should possess in order to cope with the demands of HE especially as it relates to the mediums of instruction and curriculum support within a HE institution
(HESA, 2006). The Academic Literacy subtest will measure different aspects of language ability under the following two domains:

1. Organisational Knowledge is divided into Grammatical and Textual domains:
   
   Grammatical knowledge will be assessed through vocabulary, spelling, and syntax, whilst Textual knowledge will be covered by testing the understanding of relations between parts of a text, skimming and scanning, and inference, extrapolation and application.

2. Pragmatic Knowledge is split into Functional knowledge and Sociolinguistic knowledge: Functional knowledge is measured by the ability to separate the essential from the non-essential, reading for meaning, understanding the communicative function of sentences with or without explicit indicators, the importance of self-ownership of creativity and expression, and the understanding of visually presented information (such as maps and graphs). Sociolinguistic knowledge covers understanding of basic numerical concepts, metaphorical expression, and text genre (including text register and tone).

b) Quantitative Literacy: For the purpose of the NBTs the domain of this subtest has been defined as:

   the ability manage situations or solve problems in practice, and involves responding to quantitative (mathematical and statistical) information that may be presented verbally, graphically, in tabular or symbolic form; it requires the activation of a range of enabling knowledge, behaviours and processes and it can be observed when it is expressed in the form of a communication, in written, oral or visual mode (HESA, 2006, p.30).
From the outset it was agreed that this subtest would be difficult to construct. Amongst the reasons for this were differences in understanding of what the domain entails and the degree to which it is embedded in everyday contexts. The guiding principles in constructing items for this subtest are that items should challenge test takers to solve problems in real contexts, to respond to or interact with a variety of information or data, and require the activation of a range of enabling knowledge, including quantitative, mathematical, and literacy skills (HESA, 2006).

c) Cognitive Academic Mathematical Proficiency (*CAMP*): Whereas the previous two subtests are intended to assess generic competencies this subtest looks specifically at the ability of the new NSC to prepare school leavers for the expectations placed on incoming HE students in terms of mathematical proficiency. This test draws the specifics of its domain from the learning outcomes of the Learning Programme Guidelines of the National Curriculum Statement for Mathematics for Grades 10-12, which centre around number and number relationships; functions and algebra; shape, space and measurement; and data handling and probability. From these guidelines the CAMP test has been designed to assess two domains through specific aspects. The specifics of this test are described in great detail in HESA (2006), but briefly, they encompass:

1. Problem solving and modelling within mathematical contexts: This domain is broad and includes amongst others four main areas of mathematics. They are algebra, geometry, trigonometry and statistics. Some of the key words that are used in the description of this domain include operation, interpretation, perception, representation, function, relationship, translation, conversion, and recognition. From this it is clear that a broad range of skills are covered. In the
assessment of this domain test takers are challenged to display evidence of number sense and the ability to make quantitative comparisons.

2. The manipulation of formal statements and interpretation of inferences: This domain covers a higher level of reasoning ability by challenging test takers to identify, evaluate and critique mathematical arguments; employ tentative or conclusive reasoning; see logical relationships; understand the concepts of chance and probability; and to manipulate algebraic expressions.

Despite the increasing popularity of benchmark testing the issue is still much debated with strong arguments for and against its use and usefulness in especially multilingual and multicultural environments (Foxcroft & Roodt, 2005; Koch, 2005). This is especially true when applying an assessment format that has been largely developed in a western framework in non-westernised settings. A major criticism is that many of the standardised tests used for admissions purposes in South Africa are in either English or Afrikaans. Although these languages may be familiar to most students, they are not necessarily the home languages for the majority and their instruction of these languages may not linguistically sound (Probyn, et al, 2002). If standardised testing is therefore employed, care should be taken that it is used within the context of broadening access and that those who are supposed to benefit from this initiative are not further disadvantaged by it (Koch, 2005).
2.5 Concluding Remarks

This chapter highlighted the difficult task experienced by HE institutions with respect to granting access to applicants. Only a few of the major developments were mentioned to demonstrate the challenging environment within which high-stakes admissions decisions have to be made. On the one hand access is required whilst on the other quality is demanded. In an attempt to aid comparison between applicants, admissions tests have served an ever-increasing role. Far from ending the debate, this practice has ignited new arguments about the fairness of this practice (Foxcroft & Roodt, 2005). Testing can add value to the admissions process, but it can also be abused to the detriment of applicants and institutions. The sensitivity of this issue increases the need for transparency, sound methodology and accountability in testing. It is important that test construction and implementation is conducted in line with best-practice guidelines that have been proven in multicultural settings. The domain of test construction is however well-researched and guidelines exist to further the cause of fairness in testing. Further aspects of this sub-discipline of psychometrics are explored in the next chapter.
Chapter 3
Issues in Admissions Testing in Higher Education

3.1 Introduction

In the previous chapter the growing phenomenon of admissions testing for Higher Education (HE) in South Africa was referred to. South Africa is not unique in this as testing is employed at many HE institutions across the world. Some countries have already had a fully developed equivalent of the fledgling National Benchmark Tests (NBTs) described in Chapter 2 for many years, such as the United States, Israel, Sweden, Singapore and Australia, and the list of countries considering national testing is continuously growing. As the demand for HE escalates, admissions testing offers an efficient solution to the dilemma of selecting students from the abundant applicant base for the limited spaces available.

Due to the high stakes involved in admissions testing it is important that tests of this nature are constructed and employed in an ethical manner. The main purpose of this chapter is twofold. Firstly, some of the issues that are pertinent to admissions testing are highlighted. This is done in order to flag some of the key aspects that should be included when conducting an evaluation of the effectiveness of admissions testing. Secondly, some of the guidelines that have been developed through international experience and research are mentioned. This is aimed at providing sound principles upon which to base admissions testing. The chapter commences with a brief overview of the history of admissions tests.
3.2 Admission Criteria and the Role of Testing in Admissions

Access to HE is a complicated issue that is informed by the dynamic educational, social, and political landscape within a country. The traditional view of access to HE was to limit access to those who demonstrated academic excellence. In more recent times the impact of social reform has resulted in a greater demand for HE to reflect the diversity of society. In this current environment it is the task of HE to balance the demands of merit, competence and social representativeness (Herman, 1995). There are a number of criteria that can be used to evaluate the effectiveness of admissions criteria. Some of these include whether the criteria serve to broaden access, achieve equity targets by increasing the participation of under-represented groups, facilitate the identification of learner development needs, increase retention and throughput rates, and are sensitive to labour market needs (Department of Education, 2001; Harman, 1994; Klitgraad, 1986). For HE admissions practices to remain defensible in modern times, all aspects of the admissions process must be proven to contribute to the effectiveness and fairness of selecting applicants (Zaaiman, van der Flier & Thijs, 1998).

Miller and Bradbury (1999) identified five indices to evaluate admissions criteria:

1. **Excellence** relates to the throughput and graduation rates achieved by students as a result of the implemented admissions criteria.

2. **Access** refers to the percentage of students who are denied admission through a specific criterion or component of the admission process.

3. **Validity** is explored by the subsequent failure rate of those who are admitted on the existing criteria.
4. **Equity** relates to the percentage of potential students who would succeed academically but are denied access because of the criteria.

5. **Accuracy** refers to the extent to which errors in decision-making are reduced by a specific criterion.

Admission testing aims to enhance the effectiveness of the above indices (Seymour, 2002). Attempts to improve performance in these indices have resulted in more complex admissions systems and the growth of admission testing. Admissions tests seek to measure the potential for academic success. Research has found that educationally disadvantaged learners succeeded in narrowing the initial gap in academic performance compared to non-disadvantaged learners through the duration of HE studies (Hagedorn & Nora, 1996; Huysamen, 2000; Seymour, 2002). This places an added expectation on admissions tests to identify applicants who have the ability to succeed in HE.

Apart from being able to identify learners with academic potential admissions tests correlate significantly with future academic success (Clarke, Madaus, Horn & Ramos, 2000). A study on a South African student sample found a 30% increase in predictive accuracy through the addition of test results to matric performance, thereby adding value to the admissions process (Greyling, 2000). Other South African studies have confirmed the positive impact that admissions tests can have on the validity of admissions decisions (Foxcroft, et al., 2000; Nock, 2001). International studies concur about the value and importance of admissions tests. In a recent study in the United States 60% of colleges reported that admissions tests added considerable value to admissions
decisions (National Association for College Admission Counseling, 2008). The National Admissions Test for Law (LNAT) in the United Kingdom has also been found to aid admissions decision-making by providing a more rounded picture of applicants’ abilities than A-level results alone (Phillipson, 2008).

Testing forms an integral part of admissions practices for an ever-increasing number of institutions, which means that testing forms part of the debates around access. Although it may add value to admission, it is also controversial and often viewed as an obstacle to access, especially for applicants from minority or underprivileged backgrounds (Linn, 1990). Research has indicated that groups which differ in terms of socio-economic, cultural, or education background perform differently on tests that measure cognitive ability (Miller & Bradbury, 1999). This tendency, as well as the high stakes impact of admissions decisions requires continued research evidence related to the effectiveness and fairness of admissions tests as part of the decision-making process. Such research findings on the efficacy of admissions tests should also be readily available to interested parties to further the causes of transparency, openness and accountability in testing.

When deciding on the inclusion of testing as part of the admission process it is important to be aware of its potential limitations, as well as best-practice models in terms of the implementation of such a programme. Learning from those who have treaded the course before is useful. Research based on the experiences of others could yield important information on a variety of testing-related topics, including what should be tested, what the practical applications of tests are, what the potential problem areas are together with possible solutions, and how a testing process should be implemented.
Testing has formed part of HE admissions internationally for many decades. Some of the guidelines that have evolved will be presented next, followed by issues that are salient for admission testing.

Testing for university entry can be employed for two purposes, namely admissions and placement. Although validity in all its forms are important for both, Anastasi and Urbina (1997) consider different types as essential for each. If the focus of testing is to provide entrance into a specific programme, the outcome is either to provide access or not. The finality of this process means that the predictive power of the process needs to be high and therefore predictive validity is very important. When access is not at stake, but the tests are used to place students in the most appropriate course, the aim is to measure competency as accurately as possible and to match skills to a specific programme that would not only suit the ability level of the student, but meet the student’s developmental needs. Tests with this focus need to have strong construct and concurrent validity (Anastasi & Urbina, 1997).

In the next section a brief historical perspective on testing and some of the resulting issues are presented.

3.3 The History of Admissions Testing

The term test was first used in an article in 1893 by J M Cattell to describe a series of tests that were designed and used to measure intelligence in university students (Foxcroft & Davies, 2008). In the US, the focus of testing in the HE setting soon shifted to the purpose of admission. Tests tended to be in essay format, tapping a variety of skills including arithmetic, language, spelling, handwriting, and subject-specific knowledge.
Essay tests were however lengthy to complete, labour intensive for administrators, and prone to subjective scoring practices. In 1900 the College Board was formed in the US following a need to develop a uniform benchmark for entrance to HE institutions (Clarke, et al., 2000).

The development of multiple choice item formats by Frederick Kelly in 1914 revolutionized testing by reducing marking time and improving standardization (Foxcroft & Roodt, 2005). In 1926 the Scholastic Aptitude Test (SAT) was developed as a multiple choice test by the College Board and first employed in the US for the purpose of admission to Colleges. Since then, millions of prospective students have taken the SAT, and it remains one of most widely used and well respected measures in the field. Through the decades the SAT has been widely researched, benchmarked and redeveloped. Following concern over the increasing importance of the SAT and the sensitivity of the word “Aptitude” the test was renamed to the Scholastic Assessment Test. The core subtests of the SAT tap the domains of English language ability and mathematics, whilst some subject-specific tests are available as extra components in the battery (McDonald, Newton, Whetton, & Benefield, 2001). The computer age introduced new heights in standardized assessment by reducing human influence in administration, scoring and interpretation. Test developers were quick to realize the value of computers and in 1992 the first computerized version of the SAT was constructed.

Many countries have developed national admissions tests for entry into HE. Two types of tests have been popularized namely entrance exams and standardized aptitude tests. Entrance exams are more subject specific, depending on the course of study the applicant wishes to pursue. Examples of countries that employ this type of tests include
China, Iran, Spain, Turkey, and Argentina. Standardized aptitude tests focus on generic competencies that are deemed as foundational for HE studies (Koch, Foxcroft & Watson, 2001). The main examples include the SAT in the US, the Swedish Scholastic Aptitude Test (SweSAT) in Sweden, and the Psychometric Entrance Test (PET) in Israel, although individual HE institutions in many countries employ aptitude tests for admissions purposes (Helms, 2008).

The validity of tests and the ethics of testing practices have been challenged through professional critiques and legal actions, which have shaped both tests and the testing profession. Major criticisms and concerns centred on the validity of assessment instruments for decision-making, the lack of guidelines for test development and use, the over-reliance on test results, and the protection of test takers’ rights. After the landmark Hobson v Hansen case of 1967, the court ruled that group-administered ability tests discriminated against minority groups and could thus not be used to assign test takers to different tracks or streams within programmes (Aiken, 2000). This decision was softened by the landmark case of Griggs v. Duke Power in 1971, where the US Federal Court ruled that tests used for personnel selection must be proven to be job relevant (Aiken, 2000). This placed the burden of proof on test developers and users to research and supply scientific evidence of the appropriateness of tests for all test taker groups before employing such tests for decision-making purposes.

According to Murphy (2004) tests are sometimes the best single method on which to base important decisions. Practitioners have however been criticized for over-reliance on test results to the detriment of other important factors, such as previous performance, personal history, motivation, and so forth. These cases challenged the practice of basing
important decisions solely on test results and highlighted the importance of using a multidimensional approach to enable decision-making.

In the United States the Family Educational Rights and Privacy Act of 1974 specified that test results may only be made available to other parties after informed consent has been obtained from the test taker or legal guardian. This served to protect the individual’s right to privacy and controlled the communication of sensitive information to outside entities (Aiken, 2000).

Critics of ability tests have deplored the potential for abuse and the lack of stricter regulations since the advent of modern psychometrics (Rust & Golombok, 1999). In 1984 three professional organizations, namely the American Educational Research Association (AERA), the American Psychological Association (APA), and the American Personnel and Guidance Association (APGA) formed the Joint Committee on Testing Practices (JCTP). The JCTP was tasked with working with test publishers for the improvement of the quality of tests and assessment practices, especially in the development of tests, selection of appropriate measures, interpretation of test scores, endeavours to promote fairness, and empowering test takers through increased access to information. In 1985 the AERA, APA and APGA also published the *Standards for Educational and Psychological Testing*. This document provided a set of ethical codes regarding test administration, standardization, and validity aimed at promoting the welfare of test takers against the abuse of assessment instruments (Aiken, 2000).

In the US the list of legal challenges against testing is long and varied, serving to inform test development and the field of psychometrics at different points in time. The end result is that test developers and users have to be more aware of the possible
ramifications of test results, the rights of test takers, and being able to provide research
evidence of the effectiveness and fairness of specific tests and the value that they add for
the benefit of both test users and test takers.

The development of testing in South Africa, as with so many other aspects within
the country, was greatly influenced by the political agenda of the day. Despite great
advances in testing between 1926 and 1985, most of it occurred specifically for the white
population (Foxcroft & Davies, 2008). It was only with the shift within the realm of
politics towards the end of the 1980s that tests designed for the general population started
to emerge. In 1991, the General Scholastic Aptitude Test (GSAT), a unified revision of
former tests for separate ethnic groups, was a landmark in test development in South
Africa (Claassen, et al., 1991). The increased demand for HE, especially from formerly
disadvantaged students sparked a growth in testing for HE admission. Applicants who did
not meet the direct entrance criteria for HE could be granted admission on special
grounds, such as performance on an admissions test. The GSAT has been used
extensively, but many others were developed to address the growing needs of different
institutions.

Increased concerns over the applicability of matric results for HE entry and the
flexibility afforded by admissions tests resulted in the growth of testing as part of the
standard admissions process at most HE institutions. The culmination of this expansion
resulted in the NBTs, a project commissioned by Higher Education South Africa (HESA)
in 2006. This project is aimed at pooling the expertise of different institutions to develop
a general test for entry to HE, and thereby optimize the quality and utility of the measure
for the benefit of the sector and test takers (HESA, 2006). The NBTs fall into the category of standardized aptitude tests mentioned earlier.

Apart from concerns about admissions tests as a barrier to access to HE, testing in South Africa has not been extensively challenged in court. Most court cases against universities have centred around the right to access where an application for admission was unsuccessful. In Motala and Another v. University of Natal (1995), for example, a female Indian student with good matric results was not accepted to medical school. The court ruled in favour of the University of Natal, citing that affirmative action policies could not be affected by this application. One possible reason for the lack of legal challenges to admissions tests in South Africa is that the Higher Education Act (1997) protects the rights of universities to set their own admissions criteria, and that these tests are currently generally institution-specific and on a relatively small scale. This could change with the NBTs as admission testing could become a national occurrence.

According to Taylor (1992) testing in South Africa has followed many of the practices that were developed in the US, thereby pre-empting legal issues, but as the NBTs become part of the national debate surrounding access to HE, the test developers will have to be able to supply research motivation of the value of the tests in the admissions process.

3.4 Guidelines for National Admissions Tests

National HE admissions tests have been around for decades in some countries, including the US and Israel, and much research has flowed from the construction and implementation of these measures (Mcdonald, et al., 2001). Through research several areas have been highlighted as being critical in the conceptualization and realization of a
national admissions test. The areas can broadly be classified into practical considerations around the design and implementation of the test, psychometric considerations such as reliability, validity, and bias, and issues related to the resilience of the tests and independence from the school curriculum.

The areas above have repeatedly been confirmed by expert opinion and international research. According to Linn (1990) the validity of tests, coaching to prepare for testing, and the appropriateness or fairness of tests, especially for minority or underrepresented groups has been a focus of controversy for some time. Almost two decades after Linn’s observations these issues remain important requiring continuous investigation. A study by McDonald et al. (2001) summarizing research on the SAT again raised the issues of validity, bias, and coaching in relation to admissions tests.

Fortunately, guiding principles have been developed that can steer new testing projects through the myriad of potential pitfalls. A case in point of this is by the International Test Commission (ITC), a global network of test developers, users, and publishers, that has formulated guidelines to promote quality and fairness in testing. In the following sections some of the salient aspects for admissions tests are discussed, together with possible solutions presented in the ITC guidelines. In some cases the SAT will be used as an example to highlight the relevant issues.

3.4.1 Considerations when Implementing a National Testing Programme

When designing a national testing programme it is important to consider the practical operational aspects of tests, the domains that should be assessed and the type of tests that should be considered.
There are numerous practical considerations that influence the choice of test and the development of assessment practices. A national administrative office must be established to oversee the development, implementation, research, and marketing of the test. According to the ITC, test users must ensure that all communication about a test is presented in an open and balanced manner to test takers, as well as through the media (2000, Guideline 1.1.6). Decisions need to be made about the efficacy of regional branches of this office, to enable interaction with the local customer base. Systems for making bookings and communicating testing-related information and test results to applicants need to be established. Test dates and venues have to be established.

The relationship between the national testing agency and separate institutions, as well with national government needs to be clarified and operationalized. The parameters of permitted test use must be agreed upon between the test developers and test users, in order to prevent potential misuse of test results. The ITC recommends that test users must avoid situations where they may have a vested interest in the outcome of the assessment, thereby damaging the relationship with test takers (2000, Guideline 1.1.7). In order to promote this, the activities of a national admissions test must be centrally controlled by an impartial organization, and the limitations of the relationship with stakeholders must be clearly spelt out.

Research has highlighted several areas that play an important part in student success. These include cognitive skills related to language, English proficiency, problem-solving skills, subject-specific knowledge, and study skills (Delvare, 1995). South African research has highlighted academic literacy, English proficiency, numerical skills, and non-cognitive factors (such as career preparedness, personal goals, and previous life
and work experience) as playing a vital role in undergraduate studies. Of these, academic literacy, English language proficiency, and numerical and mathematical proficiency were deemed as foundational for academic success (Koch, Foxcroft & Watson, 2001). Tests of language and numerical skills therefore often form vital components of an admissions battery.

Another aspect of tests relates to the purpose of the assessment. Test users can make use of either norm-referenced or criterion referenced tests. Norm-referenced tests are designed to measure level of attainment in a specific area in relation to other test takers. A benefit of this type of test is that norms of test performance can be created for specific population groups thereby enabling comparison of test takers’ performance with that of their peers (Anastasi & Urbina, 1997). Criterion-referenced tests aim to assess competence within a specific domain in order to measure the level of attainment. Cut-scores can be developed that describes the level of performance or the specific skills that have been mastered by a test taker. The benefit of this type of test is that it provides information on the extent to which a test taker is proficient within a domain, thereby highlighting those areas in which development may still be required (Anastasi & Urbina, 1997). This type of test adds value to the testing experience for both test users and test takers as it not only yields a test score, but provides a description of the level of skill.

For admissions tests the type of test selected will be guided by the focus of the process. If the purpose of admission is to purely compare the performance of applicants with others, then a norm-referenced test is sufficient. If admission is developmentally focused however, looking at the strengths and deficits of applicants, and the extent to
which development is required in order for applicants to succeed as in their studies, then a criterion-referenced test would be more helpful.

A final note on test type is that norm-referenced tests are less labour intensive, as the purpose is to only select those applicants who perform the best on the assessment. Criterion-referenced tests are more labour intensive as it not only places the emphasis on the performance of test takers, but also on the ability of the institution to provide the support that students require in order to advance in their studies.

Other aspects that require careful consideration are that of test format and mode. Multiple choice tests for example are easy to administer, score and interpret. On the other hand, this format has been criticized for favouring lower-level skills with not enough emphasis on the type of answer production that would be required in real-world tasks (Aiken, 2000; Linn, 1990). Essay-type questions may provide an indication of test takers’ ability to formulate arguments and write coherently, but this format requires more specialized scoring and may be hampered by subjective scorer influences. Computer-based testing (CBT) provides a quick alternative, with little administrator input, but this mode can be too technologically advanced for some test takers and require the availability of computer laboratories and technical support staff. The benefit of developing the NBTs as a computer-based test is that it would be in line with the international trend, and provide test results and individualized reports to institutions and test takers much faster than a manually scored and interpreted test. The drawback of such a system could be that many South Africans have not had sufficient exposure to computers, which could heighten test anxiety. However, Davies, Foxcroft, Griessel and Tredoux (2005) observe that with exposure to technology increasing in the South African
society, the adverse impact of lack of computer familiarity on CBT is decreasing. Another drawback is that computer laboratories may also not be readily available in rural areas, thereby adding travel and accommodation burdens to test takers from those areas.

 Designing and implementing a national admissions test is a costly exercise. It is therefore crucial that the aspects raised above are considered carefully before resources are committed.

3.4.2 Reliability

  Reliability refers to the extent to which a measure is consistent (Cozby, 2004). According to Hambleton it is important to establish both the validity and reliability of a test for each group before it is used in practice (2004). There are different methods for establishing reliability including test-retest, parallel forms, and split half reliability, as well as internal consistency (Murphy, 2004; Rust & Golombok, 1999). Of these, test-retest reliability is important as it indicates the consistency of measurement over time (Rust & Golombok, 1999). This is vital in an admissions setting where applicants may only get one chance per intake to perform on a test that is used to compare them to other applicants, often for limited positions available in a programme. Tests therefore need to be accurate and consistent to add value to the admissions process. For the purposes of reliability, the coefficient only needs to be above .60, but when a test is used to compare individuals, such as for admissions purposes, its reliability needs to be above .85 (Aiken, 2000).
3.4.3 Validity

At the heart of testing is the desire to understand and describe individual traits. Admissions tests aim to measure the skills that are fundamental for academic success. Validity is the extent to which a test measures the trait it is intended for (Rust & Golombok, 1999). This definition sounds simplistic, but is an overarching concept that includes a number of different forms including face, content, construct, concurrent, and predictive validity (Cozby, 2004; Rust & Golombok, 1999).

Face validity relates to the overall appearance of a test as being a sound measure. This is important for test takers as it creates the impression of the test as a valid instrument. This can be established by obtaining feedback from a sample of the intended test population, as well as testing experts on the clarity of test instructions, general layout, and format of test items (Aiken, 2000). Content validity refers to the extent to which test content reflects the purpose and topic of the test. This is ascertained through a review of the items by experts within the field, in terms of breadth and depth of test items.

Construct validity is concerned with whether the test accurately represents all the underlying aspects of the specific trait or skill it is intended to measure. Concurrent validity investigates the extent to which test results are in agreement with existing reputable measures of the same trait. These types of validity can again be verified by expert review or research into the similarity of performance between the measure being investigated and other established measures of the same construct (Cozby, 2004). In the case of admissions tests, new measures can be benchmarked against other established measures, in order to determine construct and concurrent validity.
Predictive validity is important for cognitive tests as it refers to the ability of the test to predict future outcomes (Rust & Golombok, 1999). This is especially important when assessing a candidate for the possibility of success in future studies. Test results are used to identify those who would be academically successful from those who would not. In HE admissions this information is used to provide access to applicants who may succeed in their studies, whilst not admitting those who are predicted to fail (Zaaiman, et al., 1998). Different theories exist as to how predictive validity must be established. A simplistic investigation of predictive validity simply examines the correlation between the criterion and predictive variables at face value (Seymour, 2002). In admissions testing this would be represented as a correlation of between zero (no correlation) and one (a perfect correlation) between the tests used in the process and the HE academic results obtained, with the possible inclusion of other predictive variables, such as school performance.

Other theorists view predictive validity as a more complex paradigm. According to Young and Kobrin (2001) several factors can influence the outcome of predictive validity studies. Some of these factors include the nature of the variables investigated, variability in performance in the criterion and predictor variables between different groups, range restrictions of samples, the researcher’s definition and understanding of predictive validity and the statistical techniques used to investigate predictive validity. The research strategy must be tailored to suit the variables being investigated and will be affected by the homogeneity or heterogeneity of the population or sample (Koch, 2005).

Although an admissions test should be carefully scrutinized for its validity, predicting academic success has proven to be difficult (Hughes, 1989). As academic
success can be affected by a myriad of personal, interpersonal, social, environmental, and political factors it is difficult to obtain validity coefficients of higher than .40 for academic success (Hughes, 1989). This becomes even lower when predicting success beyond the first year of study, with coefficients usually ranging between .25 and .30 (Cole, et al., 1998; Seymour, 2002). The predictive ability of tests may also vary depending on gender, age, or culture. According to the ITC, test users should ensure that there is a strong correlation between the skills that are assessed and the behavior about which inferences are made (2000, Guideline 2.1.4). The ITC furthermore recommends that studies are periodically conducted on the population group from whom test takers are sourced, as well as changes in the criterion about which inferences are made, to ensure the validity of the test and its relationship with the criterion variable (2000, Guideline 2.9.1).

When delving into research on predictive validity it is important to be aware of the aspects mentioned above, as well as the impact they can have on reported results, and to also take note of the approaches and statistical techniques that have proved useful for other studies on admissions tests.

In order to improve the predictive validity of the admissions decision the ITC also recommends that test scores are combined with collateral sources of information, such as previous academic performance. The aspect of the relevant weighting that should be given to each component should be carefully researched (Linn, 1990). For university admission in the United States, the SAT and school results are combined into an overall score. Both elements contribute equally (50:50) to the equation. If the composite score
method is employed in South Africa, research would have to be conducted to ascertain the optimal balance between test and school results.

The composite score method is however not the only approach to employing test information. In South Africa, admission testing at the Nelson Mandela Metropolitan University utilizes a profile approach. Test and school results are plotted on a graph, in combination with relevant course cut-scores, to provide an overview of the proficiency of applicants. This procedure relies less on final summative scores, but by providing an overview of each test taker enables decision-making based on proficiency level and need for development (Koch, 2005).

In closing, it is important to note that biographical variables such as gender, age, language, and cultural background can influence the validity of a test (Cole, Muenz & Bates, 1998). This can lead to different levels of validity for different groups, which should be reported in full to stakeholders (Huysamen & Raubenheimer, 1999).

3.4.4 Bias

In the case of programmes with limited numbers admissions tests are used as a comparative tool to distinguish between applicants in allocating spaces (Miller & Bradbury, 1999). It is therefore important that the tests used for this purpose do not discriminate against applicants on the basis of gender, language or cultural background (Aiken, 2000). This is referred to as bias, which in psychometric terms, occurs when there are systematic errors in measurement or prediction of performance on a test (Murphy, 2004).
The study of test bias is especially important in multicultural settings where applicants from different backgrounds may have a different understanding of the meaning of a question or specific words in a question. This could result in certain population groups underperforming on a test if it is biased towards them. It is also important that the information that is obtained from tests is used in a fair and equitable manner. If the results of a biased test are taken at face value and used for HE admissions without interpretation, it could lead to an unfair process and the effect would be an inequitable representation of certain population groups in HE (Seymour, 2002). The ITC guidelines state that if a test is to be used for different groups it must be proven to be valid, unbiased, appropriate, and meaningful to each group, and that evidence must be available of possible group differences (2000, Guideline 2.3).

Bias in a test can be minimized through continuous research in multicultural settings to determine bias in test scores as well as bias resulting from interpretation. Such research could lead to refinement of tests and application of test scores (Aiken, 2000). Although little may be done to obtain equal scores for test takers from different population groups on some tests, awareness of differences and specific norms, or benchmarks of performance for each group, could facilitate fair test interpretation and subsequent decisions. Another strategy to minimize bias in a selection procedure is to use more than one measure and method of assessment to safeguard against one test’s potential shortcomings (Foxcroft & Roodt, 2005; Murphy, 2004).
3.4.4.1 Differences in performance across groups

The issue of differential performance on admissions tests is not restricted to ethnic or cultural groups, but also extends to gender. An example of bias towards certain groups is on the SAT. As this test has been researched and sensitively redeveloped over decades it may be reasonably assumed to offer an accurate assessment for the majority of test takers. Research has however shown that African-Americans consistently score unjustifiably lower on the SAT than white test takers by about one standard deviation (Jencks, 1998). The verbal and math sections are also biased in favour of males. On top of this the SAT underpredicts the subsequent academic performance of female learners. White males tend to outperform other sections of the population on this well-known measure (McDonald, et al., 2001). The source of this discrepancy is difficult to pinpoint, but the effects thereof in terms of admissions decisions can be substantial. The SAT is employed by many of the top-rated colleges and universities in the US, as well as internationally, and the differences in performance across groups have been criticized for limiting study opportunities for especially minority groups within these institutions. Attempts to address these issues have lead to different norms being created for separate groups, in accordance with ITC guidelines (2000, Guideline 2.7.4), but even this is considered as an attempt to create artificial equality instead of solving the source of the problem. Even if admissions personnel are sensitive to test bias and ameliorate test scores, as is recommended by the ITC (2000, Guideline 2.7.12), applicants who perform lower tend to gravitate to programmes with lower entrance requirements, rather than face rejection from more selective programmes (Murphy, 2004). It has been argued that this has contributed to the under representativeness of black and female students in more
prestigious courses and key growth areas, such as science, medicine, and engineering (McDonald, et al., 2001).

3.4.4.2 Language issues in testing

Many countries have a rich diversity of cultural and language groups that share a national identity. In order for individuals from different languages to communicate many develop multilingual skills to enable successful integration into broader society. The level to which each person is competent in different languages naturally varies from person to person. South Africa is a country that recognizes eleven official languages. Despite the fact that English is not the home language of most citizens it is the main language of teaching and learning at school and HE level. Most textbooks are written in English and with its international appeal, proficiency in this language has become a valuable asset for students to communicate their skills to a broader market. It is therefore only natural that most admissions assessments for HE include competence in English as one of its key components. The preference for English in assessment has however not gone unopposed, with some questioning its predictive validity in terms of academic success (Koch, 2005). Even South Africa’s National Benchmark Tests (NBTs) project acknowledges that language is a contentious issue and admits that although the tests are constructed in English it is merely a starting point that leaves the possibility open for future versions in other languages (HESA, 2006). The SAT for example includes an English language test, despite the growing Hispanic population and test takers from international countries. The importance of ability to function in the language of teaching and learning of the institution cannot be denied, but many argue that demonstrating ability through tests in
the language most familiar to the test taker is a better predictor of future academic success (Koch, 2005).

The importance of equivalence across language groups in assessment has lead to much international research on test bias, adaptation, and translation. Numerous tests have been translated into multiple languages. Obtaining equivalence for different language groups on tests is complicated however. According to Koch (2005) there is a growing awareness that constructs differ between cultures, that norms that are developed in one population group may not be applicable to others, and that test performance may be affected by nuisance factors such as familiarity with the test mode and format. According to Van de Vijver and Tanzer (1997) unfamiliarity with the question format seriously disadvantages test takers who have not been exposed to it, but Hambleton (2004) indicates that with more experience this disadvantage is minimized. Poortinga (1995) states that knowledge of the language and the question format the test is presented in are the most important sources of cultural bias. Poortinga suggests that the best possible method to minimize bias is to avoid questions that contain gender- or culture-specific information, and to provide extensive examples of the question format as well as the method in which answers are to be provided. Other aspects include measures to promote equivalence of items by professional translation and back-translation, as well as the use of expert panels to judge the equivalence of translated tests in terms of language and content. When test results are available for different groups, analyses on separate items may also be performed.

Item response theory (IRT) and differential item functioning (DIF) can be used to identify undetected sources of bias in test items through statistical calculations, such as
item characteristic curve (ICC) and logistic regression. Hambleton (2004) argues that instead of translation the focus should be on adaptation of tests. Instead of developing perfectly translated versions of a test the ideal outcome should be culturally relevant and linguistically accurate versions of a measure. The different methods of establishing linguistic equivalence in practical terms include using fully bilingual groups, representative samples of monolingual groups, and matched monolingual group designs. All of these present with challenges however, such as the variations within bilingual ability, what selection criteria are used to construct a representative monolingual group, and what the criteria are for matching monolingual groups (Sireci, 1997). The ITC recommends that test developers must follow a rigorous methodology and be sensitive to issues of content, language and culture (2000, Guideline 2.3.9). In addition to this, adequate arrangements need to be in place for test takers with hearing, visual, or motor disabilities to minimize the impact on score validity (2000, Guideline 2.3.14).

3.4.5 Coaching effects and Overlap with the School Curriculum

In the United States, the SAT forms an important basis for admission to many HE institutions. A lucrative industry has emerged to prepare test takers for this test. Research on coaching for the SAT has yielded inconsistent results (McDonald, et al., 2001). In the instances where coaching had a positive effect on test scores studies found the increases to be negligible and insignificant (Aiken, 2000). Linn (1990) argues that even if coaching for the SAT only has a small effect on test scores, this may be enough to provide an applicant with an advantage over others, especially for courses with limited numbers and stringent selection criteria. Bond (1989) also questioned the negative impact that
coaching may have on the predictive validity of the admissions tests. The ITC recommends that test developers and users ensure that test results are not influenced through coaching or other measures (2000, Guideline 1.4.3). As a result of concerns about the integrity of the SAT the Educational Testing Service examines new pilot items not only for bias, but also for resilience to coaching, and revises questions accordingly before inclusion in official versions of the SAT (Aiken, 2000).

The coaching industry is not only prolific, but also costly. This means that it may place applicants who can afford the cost in a more favourable position for admission than those from disadvantaged socio-economic environments (Linn, 1990). This situation is one that needs careful attention when admissions tests are used for the sake of broadening access to underprivileged applicants, such as is the case in South Africa, requiring continuous research.

A further aspect of testing relates to its independence from the school curriculum. Admissions tests are used as stable benchmarks to compare applicants with different educational backgrounds. They form an integral part of the admissions process and in some cases carry as much weight in decision-making as school results (Aiken, 2000). The value of admissions tests hinges on their ability to measure competencies that are relevant for academic success without overlapping the information already obtained through school assessments. In the US the importance of the SAT has influenced the classroom curriculum however, to the extent that valuable contact time is spent on teaching and preparation for the SAT (Phelps, 1999). This occurrence has received widespread criticism and even had a negative impact on the image of the SAT as an independent benchmark (McDonald, et al., 2001). The South African NBTs are being developed as a
measure of competencies that underlie success in HE studies, and largely do not overlap with school assessments or the National Senior Certificate curriculum. As the profile of the NBTs escalates however the classroom situation will have to be carefully monitored to ensure that valuable teaching time is not diverted to preparation for this measure.

3.5 Concluding Remarks

In this chapter some important considerations for effective admission testing were discussed. The main indicator for a good access test is that it impacts positively on the fairness of the admissions process. This is achieved on various levels. All aspects related to the testing experience need to be easily accessible to potential applicants, regardless of financial, environmental, or personal constraints. The tests have to assess those skills for which they are designed accurately for all test taker groups. Admissions tests have to identify those who would be academically successful and promote entry to HE, regardless of background or access to specialized preparatory resources. They have to provide information that is useful for test users and test takers. Finally, they have to serve the needs of society. Some of these ideals may be difficult to achieve, but improvements can be made through ongoing research on reliability, validity, bias, and robustness with respect to specialized coaching. Due to the potential for abuse, admission tests are controversial, but their use can be defended by following rigorous methodology, being transparent about their inherent strengths and limitations, and the willingness to evolve and adapt with ongoing research findings.

In the next chapter the research question for the present study is discussed, followed by the methodological approach that was followed.
Chapter 4
Problem Formulation and Research Method

4.1 Introduction

This chapter sets forth the motivation for the present study and deals with the methodology employed. The chapter begins with the problem formulation that provides the stimulus for the study. This is followed by a short description of the methodology, which is expounded upon in the ensuing sections. The history and development of systematic reviews are discussed, as well as the steps that are followed in the process. Further details are provided on the method employed during each step of this study in order to facilitate replication. The advantages and disadvantages of this methodology are highlighted, as well as the ways in which it differs from traditional literature reviews. Issues around reliability and validity are explored, as well as the steps taken in this study to improve these components.

4.2 Problem Formulation

When Higher Education South Africa (HESA) embarked on the National Benchmark Tests (NBTs) Project, as was discussed in Chapter 2, the purpose of the project was indicated as being fourfold:

1. To assess entry-level academic and quantitative literacy and mathematics proficiency of students;
2. To assess the relationship between entry level proficiencies and school-level exit outcomes;
3. To provide a service to HE institutions requiring additional information in the admission and placement of students; and

4. To inform the nature of foundation courses and curriculum responsiveness.

(HESA, 2006, p.4)

This ambitious project was informed by the needs of the HE system, but what was absent was a widespread research foundation upon which to base the development of the NBTs and for the implementation of a national testing system. The research that was conducted consisted mainly of existing admissions testing practices at various institutions. While this is useful, such surveys have been done in the past and often provided little more than a descriptive narrative of the practices followed at the institutions surveyed and a thematic analysis so as to arrive at a summative overview (Harman, 1994).

Another useful approach would be to explore admissions testing practices followed internationally. Through such investigations one would not only be able to produce good practice guidelines that are based on rigorous research, but also benchmarks that are based on international and psychometric best-practice guidelines, against which national admissions testing practices can be compared. Some of the important national and international benchmarks are for example, whether the inclusion of admissions testing in the admissions process serves to broaden access, achieve equity targets by increasing the participation of under-represented groups, facilitate the identification of learner development needs, increase retention and throughput rates, and are sensitive to labour market needs (Department of Education, 2001; Harman, 1994;
Klitgraad, 1986). By focussing on the practical experience of others, the NBTs Project could be informed of potential issues, as well as practical implementation solutions.

The ITC recommends that before deciding on a course of action for testing a thorough investigation be performed that covers the range of tests available and potential uses and issues that may arise. The use of expert advice and independent reviews are encouraged (International Test Commission, 2000, Guideline 2.2). The International Test Commission furthermore mentions that research into validity, bias, and practical implementation should be performed before embarking on a testing programme (2000, Guideline 2.2.2). An example of this is the systematic review that was commissioned by the Sutton Trust and conducted by the National Foundation for Educational Research (NFER) in the United Kingdom. In 2000, HE institutions in the United Kingdom were debating the possibility of including aptitude tests in the admissions process in order to broaden access opportunities for applicants from underrepresented ethnic, cultural, economic and social groups. A systematic review was conducted of the Scholastic Assessment Test (SAT) in the United States as a possible model for implementation in the UK (McDonald, et al., 2001). This comprehensive assessment of the utility of admissions testing in the United States, a country with strong links and facing similar social issues to the United Kingdom, served to inform the debate regarding the addition of tests to the minimum A-level school requirements for the purpose of making admissions decisions.

South Africa’s fledgling NBTs Project could benefit from similar input in its development phase. This study is aimed at addressing this need. In order to present relevant options for South African HE admissions testing and the NBTs Project, and to
ensure that the research was sufficiently focussed, it was decided to perform a systematic review, similar to the study by the NFER, of the admissions testing practices in a country with a multicultural and multilingual population, similar to the diversity found in South Africa. The criteria for finding such a country was that it had to have a history of dealing with multicultural and multilingual HE admissions testing, be committed to improving access for students from a variety of backgrounds, and have an established and well-researched national admissions test. After investigation a decision was made to focus on admissions testing practices in Israel. The similarities between Israel and South Africa are explored in detail in Chapter 5, which will serve as the motivation for this decision.

4.3 Research Aims

A systematic review on Israeli HE admissions testing practices was undertaken to provide best practice guidelines for the development of the NBTs and to inform admissions testing practices in South Africa in general. This was achieved through specific objectives. The objectives were:

1. To identify salient themes regarding HE admissions testing practices in Israel, through a systematic review.
2. To relate the salient themes in HE admissions testing practices in Israel to the context of HE in South Africa.
3. To generate a set of recommendations that are grounded in evidence-based research and best practices, to guide admissions testing practices in the South African HE context.
4.4 Method

The methodology employed in this study was that of a retrospective qualitative systematic review. The primary goal of a systematic review is to systematically search for, assess, and summarise existing research studies by following rigorous methods (Glanville & Lefebvre, 2000). Although the systematic review is a recognised research methodology, it is sometimes confused with traditional literature reviews. This has resulted in some criticism about the scientific validity of the methodology. To recognize the value of the present study it is imperative that systematic reviews be clearly understood. The next sub-section will provide an overview of the systematic review by describing its purpose, history, methodology, advantages and disadvantages, the difference between traditional literature reviews and systematic reviews, how the issues of reliability and validity are relevant to, and approached by the methodology, and how statistical analysis can be used.

4.4.1 The Purpose of Systematic Reviews

A systematic review entails a review of a clearly formulated question that uses systematic methods to identify, select, critically appraise and synthesize relevant research literature so as to provide informative empirical answers to the research question (Centre for Reviews and Dissemination, 2001; Glasziou, Irwig, Bain & Colditz, 2001). An example of this is by McDonald et al. (2001), where a study was conducted to summarise research findings on national Higher Education (HE) admissions tests, especially focussing on the SAT in the United States. Another exemplar is the study by Kyriacou and Issitt (2008) on the most appropriate dialogue methods that teachers can employ to
teach mathematics at school. This study reviewed the findings of fifteen studies to highlight those educational methods that yielded the best results in the classroom setting.

In order to understand the use of the systematic review in the social sciences it is important to have an understanding of how this method has developed. The following sub-section provides a brief overview of the history of this methodology.

4.4.2 The History of Systematic Reviews

A formal method of systematic reviews first emerged in 1975 under the term “meta analysis” (Evidence for Policy and Practice Information Centre, n.d.). The term was coined by G V Glass who performed data syntheses in the areas of psychotherapy. Although the first forays into this methodology were in the social sciences, the methodology was quickly adopted by the medical and health sciences, where a call existed for policy-making that was grounded in evidence-based medical research (Torgerson, 2003). In the 1970s and early 80s researchers in Oxford conducted systematic reviews on the effectiveness of healthcare interventions. The methodology was expounded upon to meet the rigorous standards of medical research. In 1992 the Cochrane Collaboration opened in Oxford as an international network of researchers, academics and practitioners around the concept of evidence-based medicine (Cynthia & Oxman, 1997). Another important contributor to this methodology was the National Health Service (NHS) Centre for Reviews and Dissemination (CRD) at the University of York (CRD), 2001). The modern systematic review method therefore has its roots within these evidence-based policy-making initiatives of the early 1990’s in the United Kingdom.
The value of systematic reviews in synthesizing large volumes of data was eventually recognised by professionals in different fields. In 1995 the Evidence for Policy and Practice Information and Coordinating (EPPI) Centre was created in the Social Science Research Unit at the Institute of Education, University of London, to carry out research syntheses and develop review methods in the fields of social science and public policy (Evidence for Policy and Practice Information and Coordinating Centre, n.d.). The Campbell Collaboration was established in 1999 to adapt the Cochrane methodology for the behavioural, social sciences and education fields to provide a similar high level of quality to systematic reviews (Campbell Collaboration, n.d.). The work of organisations like these has led to a rebirth of the systematic review in fields outside medical research. The use of this methodology in the field of psychology could benefit from more development and implementation however. This served as one of the motivations for this present study.

4.4.3 The Methodology of Systematic Reviews

Systematic reviews can be divided into two main types, namely qualitative and quantitative. Both have emerged as equally valuable, depending on the phenomenon being researched and the specific focus of a study. Qualitative reviews are employed when results of studies are summarised and integrated to form a comprehensive whole. This can be of great value when research around a topic is diverse and covers many angles. Quantitative reviews (or meta-analyses) are used when the statistical results of two or more identical studies are combined (School of Health and Related Research, n.d.). These reviews are often conducted in the medical or scientific fields where the
rigorous control over research conditions, such as randomised control trials, enables replication of studies. By combining the statistical results of the various studies a meta-analysis can create a larger study, with greater statistical power, that yields a definitive single result to answer the research question. What both types of systematic reviews have in common is that they seek to follow a rigorous, transparent, and replicable method.

There are therefore discrete steps in the research process. These are:

1. A research question is formulated and a research protocol prepared to provide a theoretical, conceptual and motivational background for the study. The protocol contains inclusion and exclusion criteria for the research outputs to be used for the study.

2. A search is conducted for all research available on the topic. Amongst others, database searches, journal hand-searching, conference proceedings, and studying the reference section of publications are employed to find articles for the study.

3. Every research article is evaluated and selected for the study according to the inclusion and exclusion criteria set in the protocol.

4. Each relevant article is described and classified. A further refinement or selection may be done to limit the number of research articles included in the study.

5. The articles are processed using a data extraction sheet, in order to assess the quality of the articles.

6. The extracted data are summarised in order to form a map and synthesis of the results found. In a meta-analysis the data analysis component of the study is performed and refined.
7. A final report is prepared on the results and the applicability of results is interpreted. (Torgerson, 2003; Glasziou, et al., 2001).

4.4.4. Advantages and Disadvantages of Systematic Reviews

Globalisation and the internet have, amongst others, increased the amount of information available to readers. Through the electronic media information is not only relatively accessible, but an expectation is created for professionals to be familiar with the latest developments within their fields. According to the School of Health and Related Research (n.d.) there are currently over 40 000 journals worldwide. The increased pressure on academics and researchers to publish has resulted in an overload of information available on most topics, often with differing results and conclusions. Some of the advantages of performing systematic reviews are that:

1. they synthesize large quantities of information into a manageable format, by making efficient use of existing data.
2. they help professionals in a particular field stay abreast of trends by condensing the best relevant resources into a synthesized whole.
3. accuracy regarding debated topics is improved.
4. they speed up the process between scientific research and practical implementation of treatment strategies.
5. they provide an overview of the research landscape surrounding a topic and thereby identify and highlight specific gaps in the body of research.
6. by employing a transparent methodology, they improve the generalizability of findings to fields or applications outside the intended scope of the review.
7. inconsistencies in findings are represented and new hypotheses are formed.
8. concerted efforts are made to reduce bias.
9. meta-analyses may improve the precision of the general outcome by pooling the samples from included studies and thereby increase the statistical power of results (Greenhalgh, 1997).

With all the potential advantages of systematic reviews the methodology may be assumed to be relatively robust. It has however been shown that bias can enter the review at every step. This, amongst other criticisms, has emphasized the need for thoroughness in each step of the process. Some of the disadvantages of systematic reviews are that:

1. they use existing research outputs. Within research spheres a phenomenon called publication bias exists, which means that in some domains, studies with a certain outcome, may be favoured over others. If a researcher conducting a systematic review is not aware of this, and does not make a concerted effort to find all relevant research outputs on a topic, the review may suffer because of this bias, and not relate that a result contrary to that which is favoured in publication, is feasible. This can negatively impact on the representativeness and replicability of a systematic review and lead to reviews with differing results.

2. as studies in a review are simplified in order to aid comparison, important differences between the studies may be lost, such as the specific aims, methodologies and statistical techniques used.

3. despite the claims of objectivity in a systematic review it is in fact a form of observational study (School of Health and Related Research, n.d.). It is the
researcher who defines the perimeters of a study, who decides on what research is included and excluded, and who finally presents the findings in a way that is meaningful to the reader.

4. when conducting meta-analyses, studies of varying quality may be included. This could undermine the statistical relevance of a few high quality studies. In order to address this it is necessary to compare the outcome from the meta-analysis with the results of the original studies. If there is a significant difference, the meta-analysis may be unduly influenced by a dissimilar result in one or more of the original studies. As qualitative reviews do not conduct statistical analyses, they are less affected by this phenomenon. It remains important however to be selective about the quality of the studies that are included, and to report and compare all summative comments with the results of the source studies.

5. reviews in the West often only include research outputs in English. It is difficult and time-consuming to translate a study from the original language into English, which tends to be the language favoured by many international journals. There is therefore a tendency for non-English researchers to publish studies with less dramatic results in local journals whilst only translating studies with results that are more likely to appeal to a wider audience, or have a greater chance of being published internationally (School of Health and Related Research, n.d.). This could greatly influence the outcome of a review.

Despite the disadvantages, much of the concerns around systematic reviews can be minimised or overcome through concerted effort by the researcher to obtain a
reflective cross-section of research, and to report results with as much objectivity as possible.

4.4.5 Traditional Literature Reviews and Systematic Reviews

Although both traditional reviews and systematic reviews use literature and research outputs to summarise the knowledge base in a particular field, they are fundamentally different in their approaches. Whilst the systematic review is a pre-constructed method with distinctive, predetermined steps, the traditional method has been criticized for a lack of guidelines and methodological accountability (Dickersin, 2002). The systematic review aims to address this criticism by following the pattern of rigorous scientific research. Emphasis is placed on objectivity and transparency in the process, in order to yield reliable and valid results that would be replicated by any similar study. The flaw of the traditional review is the risk that it may be haphazard or overemphasize one side of a debate, whilst a systematic review is intended to prevent chance effects and bias, by presenting all researched angles (Clarke & Horton, 2001; Mulrow, 1994). The key to the objectivity of the systematic review is the effort that is made to obtain all relevant research outputs, whether formal or informal, and the documentation of each output. With a traditional review the reader cannot be sure that all relevant sources have been consulted and considered. With a systematic review however the search for information is the second step in the process and as the success of the study depends on the quality and exhaustiveness of the search, it is well-documented.
4.4.6 Reliability

Reliability deals with the soundness of the methodology employed in a study (Rust & Golombok, 1999). The importance for systematic reviews lies in the repeatability of results, which is a main aim of this type of research. According to the School of Health and Related Research (n.d.), reliability can be enhanced by obtaining formal approval of a study through a research proposal. Furthermore, all efforts must be made to stay within the approved boundaries of the project and to adhere to the methodological guidelines for systematic reviews. In order to aid the replicability of results, a list of all research outputs included in the review should also be supplied.

4.4.7 Validity

In systematic reviews validity refers to the efforts that are made within the design to prevent systematic errors and bias (School of Health and Related Research (n.d.). Validity consists of internal and external validity.

The internal validity refers to the extent to which the differences observed between groups can be attributed to the topic being researched. Issues that are important for internal validity relate to both the selection of studies and methodology of the studies used. It is important when selecting studies to be aware of researcher bias and publication bias. Researcher bias refers to preferences that a researcher may have towards a specific outcome. Publication bias is the potential within published research to reflect one finding within research above another (Torgerson, 2003).
External validity refers to the extent to which the results of the study are
generalizable to other populations or settings, thereby informing a broader population
(Roodt, 2003).

A final note on reliability and validity is that the systematic review process has
been constructed following the rigorous scientific methods of medical research. The
benefit of following this methodological approach is that explicit and transparent
methods are used. Each piece of collected evidence is reported and evaluated for
inclusion, using a standard method. This is performed to reduce inaccurate
representations and distortions of results. The research also has a standard set of stages
which are mapped out during the initial proposal before the research is undertaken. This
ensures that the review process is not influenced midway by findings. If any changes in
direction or methodology are needed during the process, this is noted in the final report of
the study. Finally, results are intended to be replicable and accountable. The quality of
the findings is dependent on the rigorousness of the process followed. Therefore, effort
must be made to construct and follow an accountable and methodologically sound
research method (Evidence for Policy and Practice Information and Coordinating Centre,
n.d.).

4.4.8 Data Analysis

As was stated earlier, the present study is a qualitative systematic review. Data
analyses are reserved for quantitative reviews or meta-analyses, which aim to synthesize
studies with randomized controlled trials and employ the research methods of the hard
sciences. For those studies the statistics involve a two-stage process. The first stage
involves the calculation of a complex summary statistic for each study included in the review, whilst in the second stage the individual statistics are combined into a weighted average, to obtain a final result (Deeks, Altman, & Bradburn, 2001).

4.4.9 Concluding Remarks

This section presented information on the methodology of the systematic review method in order to provide more insight into the steps that were followed in the present study. One of the objectives of this method is to facilitate replication by other studies. To facilitate this, the next section details the implementation of each step of the methodology, as well as how validity and reliability was addressed.

4.5 Objectives of This Study

The aim and objectives of this study were listed earlier. The research outputs obtained for the systematic review were used in two ways. Firstly, the articles and reports found for this study were used to construct a descriptive overview of the educational landscape in Israel, the factors leading to the development of the Psychometric Entrance Test (PET), and the practices surrounding HE admissions testing. This was done to locate the similarities between the Israeli and South African educational contexts and to provide justification for relating the experience of the Israeli admissions testing practices to South Africa. Secondly, the articles and reports were appraised and synthesized through a traditional systematic review on the efficacy of HE admissions testing in Israel to inform the South African context. The specific steps that were followed are outlined in the next section.
4.6 Steps and Procedure

The first step of this study consisted of a research proposal, which outlined the background and purpose of the study. The proposal also included the inclusion and exclusion criteria for research studies to be reviewed in this study. The specific criteria employed in the selection of research articles in the review were as follows:

- **Type of admission** – the focus was only on undergraduate admissions testing practices. The reasons for this are twofold. Firstly, the present study is focused on HE admission at an undergraduate level. Secondly, the National Benchmark Tests in South Africa, and the Psychometric Entrance Test in Israel are designed for entrance to undergraduate programmes. Studies related to postgraduate studies were thus excluded from the review.

- **Type of HE institution** – any HE institution (e.g., universities, polytechnics, technical institutes, teacher training institutes, colleges).

- **Type of study** – any research output (published or unpublished) that scientifically investigated or reported on the effectiveness of admissions tests or testing practices was included into the formal review. All resources found were assessed for inclusion in the descriptive overview.

- **Language** – Only research outputs written in English were considered for the review. The reason for this is that the researchers of this study are only proficient in English. Furthermore, due to the technical methodological and statistical components of the required research outputs, specialized translation services were not readily available.
In the second step, searches were conducted for relevant research outputs. For internet and database searches a list of keywords was generated related to the study. The keywords used by the NFER to conduct searches for the UK study were consulted, expanded and utilised. This list can be found in Appendix A. All relevant articles were investigated and their references scanned for other potential research studies. Journals about psychological assessment, education, HE, and test construction were hand-searched and conference proceedings consulted for other research articles. Articles found on the website of the National Institute of Test and Evaluation (NITE), the organization that is responsible for developing, researching, and administering the national admissions test in Israel, were considered for inclusion in the review. A total of 43 research outputs were found through these searches. These outputs were consulted and duplicate versions eliminated. The remaining outputs were scrutinized for relevance to this study and all that met the inclusion criteria were included in the systematic review. Those that did not meet the stringent requirements were assessed for inclusion in the narrative overview component of the results (Section 5.2). A further breakdown of the research outputs that were included in both the narrative overview and systematic review aspects of this study can be found in Chapter 5 (Section 5.1).

The main obstacle that was encountered in the research outputs that were found was that many were in Hebrew, which discounted them from the review. When translated abstracts of these outputs were available in English efforts were made to find references to these outputs within the English research outputs found for the review. This was done to create awareness of major findings that may not be directly reflected in the results sections of the English publications. Through the activities of NITE however, research on
admissions testing in Israel was found to be thorough, ongoing, critically focussed and well-documented. This has resulted in many of the Hebrew outputs being updated or reported on in later research published in English.

In the third step each output was appraised according to the inclusion criteria listed above. Those outputs that met the inclusion criteria were evaluated in terms of methodological soundness, clarity of research method and results, as well as overlap of data sets between different studies. Potential research outputs were submitted to an independent consultant for approval before a final list was constructed for the systematic review.

Steps four and five were consolidated into one activity. Originally, it was expected that individual research outputs would be focussed in terms of areas of investigation, which would be analysed through the separate steps of article classification in step four and data extraction in step five. During the analysis process it was however found that many outputs investigated broad areas, containing a variety of findings. This necessitated a redesign of the article classification sheets to include a data extraction-type section that documented each area of investigation. The motivation for this was to ensure that small but relevant findings would not be missed in subsequent steps of the review process.

The sixth step involved the construction of a summarising map of all the main themes that emerged in the research (see Appendix C). The classification sheets and summarising map were submitted to the independent consultant, a senior researcher at a HE admissions testing unit, for verification of the themes. The consultant played an
important role in identifying changes to the original classification sheets and original 
extraction sheet and was instrumental in the redesign of the classification sheet.

This treatise document served to finalise the seventh and final step in the 
systematic review.

Apart from the review, a further aspect involved the evaluation of the research 
findings against literature and best practices, and the development of recommendations to 
guide aspects of the implementation of the National Benchmark Tests (NBTs) in South 
Africa.

4.7 Reliability and Validity

The methods that form part of a systematic review aim to minimise both bias 
through systematic errors and chance caused by random errors (Oliver & Peersman, 
2001). In order for the systematic review of this study to maintain research integrity 
however, the steps that were taken to maximise reliability and validity are discussed in 
further detail.

4.7.1 Reliability

The first step in promoting the reliability of this study was the drafting of a 
research proposal that outlined the aims, inclusion criteria and, methodological steps. The 
research was subsequently conducted in terms of the criteria set forth in the proposal. To 
enhance the reliability of the study the services of an independent consultant was 
employed to verify the results. Furthermore, apart from the article classification sheets
and summarising map (see Appendices B and C), the steps used in this research were
detailed earlier (Section 4.6), to enable comparison with the methodology guidelines.

4.7.2 Validity

For this research, efforts were made to maximise validity. This was attained by
operating from awareness of the potential pitfalls at every stage, by making a concerted
effort to collect the maximum number of published and unpublished research, by
employing an independent consultant, and by documenting the implementation of each
step.

4.8 Data Analysis

The process of obtaining and analysing the data was discussed earlier in steps four
to six of the Steps and Procedure section (Section 4.6). Each research output was
summarised to highlight important themes (see Appendix B). These themes were collated
in a summarising map (see Appendix C). Each theme was reported, discussed, and
summarised (Section 5.3) in Chapter 5, Results and Discussion, along with the finding of
every relevant research output. No statistical analyses were performed, as this study was
of a qualitative nature.

The results of the systematic review are presented and discussed in the next
chapter, Chapter 5.
Chapter 5

Results and Discussion

5.1 Introduction

In this chapter the results that were obtained from analysis of the research outputs found for this study are presented. From the initial sample of 43 research outputs, 20 were selected for the systematic review. Of these, 14 were unpublished research reports, four published research articles, and two conference papers. Four of these outputs also contributed to the narrative overview, along with a further seven sources.

Of the seven sources employed specifically in the narrative overview two were research articles, two conference presentations, one research report, one book, and the website of the National Institute for Testing and Evaluation (NITE) in Israel. The specific outputs are listed in Table 1, along with the aspect each output contributed to.
Table 1

Research outputs included in the systematic review and narrative overview

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Narrative Overview</th>
<th>Systematic Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beller, M. (2001). Admission to higher education in Israel and the role of the Psychometric Entrance Test: Educational and political dilemmas.</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
This chapter consists of two sections. The first is a narrative overview of the educational system in Israel, with emphasis on the development, implementation and use of the national admissions test. The second section comprises a systematic review of the Higher Education (HE) admissions test and testing practices employed in Israel. It is important to note that the narrative overview is an essential part of the results of this
study as many findings that are relevant to the study emerged from it. Due to the variety of subjects covered in this chapter each sub-section is followed by a brief discussion of the findings.

5.2 Narrative Overview

As stated earlier the narrative overview was compiled through sources found for the study, some of which were included in the systematic review, and some of which were not. The overview focused on the educational system in Israel, the expansion of Higher Education (HE) in the country, as well as the development and implementation of the national admissions test, called the Psychometric Entrance Test (PET). The main themes that emerged related to the characteristics of the PET as the test used as part of the admissions process, the process and utility of this national admission test, and the HE admissions policy that is used. In this sub-section the Israeli educational system is outlined briefly, after which the themes that were extracted is highlighted. Each sub-section is followed by a discussion and where possible, comparisons are made with South Africa, or the viability of employing a similar system in South Africa discussed.

5.2.1 The Educational System in Israel

In the following sections the Israeli school and Higher Education (HE) systems are described. This is followed by a discussion to highlight the similarities with the South African school and HE systems.
5.2.1.1 The school system in Israel

The Israeli school-system uses a three tiered approach of primary school, middle school, and upper secondary school. It is compulsory for Israeli children between the ages of 5 and 16 (kindergarten to tenth grade) to attend school. The medium of instruction at school is Hebrew for all learners except Arabic speakers, who account for about 20% of the population and are schooled in their mother-tongue at this level (Beller, Gafni, & Hanani, 1999). Israel also has a large, Russian-speaking, immigrant population that accounts for about 15% of the population. These learners are however instructed in Hebrew at school (Beller, 2001).

Following pre-school and six years of primary schooling, scholars enter middle schools at the age of 12 to complete grades seven to nine. This is followed by upper secondary schools for grades ten, eleven and twelve. The upper secondary system consists of two streams, one academic and the other technical (Beller, 2001).

One stream has an academic focus. The aim of this stream is to prepare students for the external matriculation exams (called “bagrut”), that are controlled by the Ministry of Education (Ayalon & Addi-Raccah, 2004). The desired outcome of this stream is a matriculation diploma, the prerequisite for entry into most HE institutions. The results for the matriculation diploma are calculated by adding the internal school marks and the external exams. Both of these components carry an equal (50:50) weighting.

The second stream has a technical focus, which aims to educate scholars for the labour market in specific vocations (Ayalon & Addi-Raccah, 2004). The technical stream is not designed to prepare students for HE, but in recent years there have been an increase in the number of technical students sitting for the bagrut. Currently about 20 percent of
technical students obtain a matriculation diploma. The educational system has experienced a phenomenal growth-rate, with the number of learners in each cohort obtaining a matriculation diploma rising from 29% in 1980/81 to 48% in 1997/98 (Beller, 2001).

5.2.1.2 Higher Education in Israel

The first universities were opened in Israel in the 1920s with the Technion of Haifa being founded in 1924 and the Hebrew University of Jerusalem in 1925 (Beller, 2001). Due to the increase in demand for HE after the State of Israel was established in 1948, a further five universities were established in the 1950s and 60s. These were the University of Haifa, Tel-Aviv University, Ben-Gurion University of the Negev, Bar-Ilan University, and the Weizmann Institute of Science in Rehovot. The Open University was established in the 1970s (NITE, 2007).

In the 1980’s and 90’s demand for HE in Israel increased, which led to the transformation of this sector (Shavit, Ayalon, Bolotin-Chachashvili, Menahem, Shwed, 2003). The number of HE institutions granting degrees rose from around 10 to over 80. Before the transformation HE students usually attended one of six research universities. The increase in the number of institutions were due to the creation of private universities and colleges, many which specialize in training in specific fields, such as teaching, law, and business. A number of international universities also entered the market by establishing branches in Israel. The number of students increased from 50 000 to over 120 000 during that time (Beller, 2001).
The increased number of learners obtaining a matriculation diploma, has created a situation in which the demand for HE outstrips supply. The number of HE learners escalated from approximately 76000 in 1990 to over 180 000 in 2002 (Soen & Davidovitch, 2004). The growing number of HE institutions has alleviated the problem of supply to some extent. The surge in applicants means that institutions are able to be more selective of the students that are admitted to programmes. This is particularly the case for the highly regarded traditional universities and HE institutions.

Traditionally universities made admissions decisions based on the matriculation diploma, but several factors necessitated a change in this practice:

1. The ever-growing number of applicants resulted in increased competition for the limited number of places in some programmes (usually in the economic and hard sciences). The matriculation scores, which formed the basis for HE admissions decisions, are calculated by computing an average of the external bagrut exams and the teachers’ internal assessments of the learner at school. As half of this mark is based on an element which varies between schools a more standard benchmark was required to place all applicants on an equal footing. (Ayalon & Addi-Raccah, 2004).

2. There was a desire on the part of institutions to broaden access to those who may not have applied themselves at school, or come from a disadvantaged economic or education background, and may therefore be further disenfranchised when selection is based solely on the matriculation diploma (Beller, 2001).

3. There was a need for a selection tool for international applicants who did not complete their school studies in Israel (Beller, 2001).
During the 1970’s several universities expanded their admissions criteria. For most, the changes favoured the inclusion of a standardized non-curricular test, aimed at predicting academic achievement, in addition to the matriculation diploma (Beller, 1995). The new individualized systems increased the burden on applicants, as those who had applied to more than one university had to complete the tests for each institution. In 1981 the seven public universities mentioned earlier (with the exception of the Open University, which has an open-access policy), reached a decision to establish the Inter-University National Institute for Testing and Evaluation (NITE), a public, registered, non-profit association (NITE, 2007). The purpose of NITE was to gather the expertise of the various institutions in terms of test construction, research and previous test item databases to develop a standard admissions instrument for all the institutions. It was intended that this measure would be administered to all applicants and that test results would be made available to any of the universities. However, institutions were still free to develop individual cut scores or criteria for their own admissions purposes (Beller, 2001).

The expertise of the members of NITE resulted in a national standardized scholastic aptitude test battery called the Psychometric Entrance Test (PET), which will be described in Section 5.2.2.

5.2.1.3 Discussion

The Israeli education system is complex and attempts to balance political and sociocultural demands from opposing sides. Israel is located in one of the most war-torn regions on the planet with internal and political struggles that have existed for thousands of years. Although the differences between Israel and South Africa are great, there are
also marked similarities. After millennia of marginalization, black South Africans have emerged as a powerful force in both the dismantling of the oppressive Apartheid system and the naissance of a democratic political system. This rebalancing is similar to the Israeli peace process that has witnessed greater calls for integration and participation from both the Jewish and Arab sides of the political divide.

Similarities also exist between the educational systems of the two countries. At school level both countries employ a three-tiered system. In Israel the tiers are called primary school, middle school, and upper secondary school, whilst in South Africa they are called foundation phase, intermediate phase and further education and training phase. In both countries one language is preferred as the medium of instruction, despite it not being the home language for a multitude of its population. In Israel, Russian-speaking students are educated in Hebrew, and in South Africa the language of instruction tends to be either English or Afrikaans, despite the fact that most students have a black African language as their mother tongue.

In both countries the final school performance is a combination of both an external government controlled exam and an internal assessment mark, which can fluctuate between schools. Two types of secondary schooling exist for both; one that has traditionally been linked to HE and another with a more vocational focus, and in both countries these old systems are in transition.

Higher Education (HE) institutions in both countries have realized the value of the addition of an external benchmark on which to base admissions decisions, and in each case this measure has taken the form of standardized aptitude or achievement testing in general verbal and quantitative reasoning domains. For both countries the growing
number of mature applicants as well as increasing interest from international applicants has necessitated an assessment of skills that are relevant to the HE environment. The desire to increase access to those from disadvantaged educational and economic backgrounds has also spurred the development of reliable and fair evaluative measurements. In Israel this desire was fuelled by the increased numbers of applicants to HE and the desire for greater representativeness of the student population (Beller, 2001). In South Africa this was achieved through social and political reform, and the protection of the individual’s right to education (Higher Education Act, 1997).

The resultant growth of testing for admissions purposes have underpinned the need for a national admissions test, those being the Psychometric Entrance Test (PET) in Israel and the National Benchmark Tests (NBTs) in South Africa. The only major difference between the two is that Israel has been involved in this process for decades longer than South Africa, which places the latter in the position to learn from the experience of the former.

The influence of sociopolitical factors on education and the drive for reform in HE is strong in both countries. This places a demand for maximum accountability and effectiveness on the HE admissions process and basis on which educational opportunities are granted.

In the next sub-section the Israeli PET is described and comparisons made with the South African NBTs.
5.2.2 The Psychometric Entrance Test (PET)

The PET is described by NITE as a tool for predicting academic performance (NITE, 2007). The PET is a collective name for an assessment that contains the subtests of Verbal Reasoning, Quantitative Reasoning, English, and the Hebrew Proficiency Test. The subtests are continuously researched for validity and updated accordingly (Kennet-Cohen, Bronner, & Oren, 1999). The aim of the PET is to measure aspects of developed ability. It makes use of the kind of basic verbal and mathematical skills that develop over the years, both in and out of school. The content of the test does not reflect specific curriculums, although it is designed to be consistent with school-based learning. (Donlon, 1984, p.58)

In 1990, research prompted a change in the structure of the tests. The traditional psychometric sub-tests of General Knowledge and Figural Reasoning were removed whilst the English, Verbal Reasoning and Quantitative Reasoning sections were expanded into three subtests (NITE, 2007). The reason for this was that General Knowledge and Figural Reasoning tests were considered to be classical components of ability tests, whilst there was a desire for the Psychometric Entrance Test to be more curricular-focussed (Beller, 1995). The subtests that form part of the PET at present are discussed in the next sub-section.

5.2.2.1 Specifications of the Psychometric Entrance Test (PET)

The current Psychometric Entrance Test (PET) consists of eight sections, which cover the core sub-domains of the assessment. The Verbal Reasoning, Quantitative
Reasoning, English subtests are each split into two sections with a time limit of 25
minutes for every section. These six sections contribute to the total score obtained on the
PET. The score obtained on each subtest is added through a weighted formula to obtain
the total PET score. The weighting is 40% for both Quantitative Reasoning and Verbal
Reasoning, and 20% for English. The final score is a scaled score ranging from 200 to
800 with a mean score of 500 and standard deviation of 100 (Beller, 2001). The
remaining two sections consist of pilot items for each subtest and do not contribute to the
PET scores. The purpose of these items are discussed next.

To avoid over-familiarity with the test there are several versions of the PET that
are used in different test sessions. As NITE is continuously developing new items in
order to increase the number of potential questions in its item bank each test has two
sections of new questions, which do not contribute to the final test score, but is piloted to
gather statistical data for potential inclusion as a test-item in future versions of the PET
(Beller, 1995).

The Verbal Reasoning, Quantitative Reasoning and English subtests consist only
of multiple-choice items (Beller, 1995). Each question is followed by four possible
answers from which the correct one must be selected. The Hebrew Proficiency Test
(HPT) consists of two sections, the first of which is multiple-choice, and the second
requiring the test taker to write a composition. The test has a time limit of one and a half
hours. (NITE, 2007).

A short description of each of the core PET subscales follows:

- English: The test consists of 54 items and covers proficiency in areas of the
  English language which are considered to be critical for the reading of academic
texts. More specifically the test covers reading comprehension, sentence completion and restatements (Beller, 2001).

- Verbal Reasoning: The 60 items of this test assess linguistic skills that are necessary for academic studies. The source version of this subtest is constructed in Hebrew. Two thirds of the test is then translated into the other target languages and the final third of items are constructed separately for each target language, to avoid items that do not translate well, as well as promote cultural and linguistic relevance for each group (Rapp & Allalouf, 2003). Questions cover analogies, sentence completion, reading comprehension, logic and antonyms.

- Quantitative Reasoning: There are 50 items that assess number use and mathematics. The questions include algebra, geometry and equations. The test taps those skills that are foundational for mathematical reasoning. Only a basic level of mathematical knowledge is needed to understand the test and all formulae that may be required are provided in the test booklet (McDonald, et al., 2001).

The Verbal Reasoning and Quantitative Reasoning subtests are constructed in Hebrew and translated into English, Arabic, Russian, French, and Spanish (NITE, 2007). The reason for translating the tests into various languages is to encourage assessment in a language with which the test taker is comfortable, which is believed will promote an opportunity for optimal performance, and enable meaningful comparison between applicants from different language backgrounds (Beller, Gafni, & Hanani, 1999).

Over a quarter of test takers prefer to complete a translated version of the PET. Those test takers who prefer must also complete the Hebrew Proficiency Test (HPT), as
this is the major language of instruction in HE institutions. The Hebrew Proficiency Test (HPT) does not contribute to the composite score. The reasons for this are that only a minority of students complete the HPT, that the test is only used as an indication of proficiency in Hebrew instead of academic success, and that the ability to communicate in Hebrew is assessed in the HPT through a composition, which is not scored, but used in a qualitative manner by institutions (NITE, 2007).

5.2.2.2 Discussion

Some similarities exist between the Israeli Psychometric Entrance Test (PET) and South African National Benchmark Test (NBTs). Firstly, both tests employ a multiple-choice answer method that enables the group-assessment test takers, as well as the ability to score and process tests quickly. Secondly, both test batteries assess language and numerical / mathematical skills for admission to Higher Education (HE). The National Institute for Testing and Evaluation (NITE) has been proactive in conducting research on the subtests of the PET and has not shied away from acknowledging deficits within the battery and to redevelop the battery in accordance with research evidence, as was evident when the General Knowledge and Figural Reasoning subtests were abandoned in favour of more comprehensive Verbal Reasoning, Quantitative Reasoning, and English subtests.

In Chapter 3 (Section 3.4.4.2) the opposing arguments for assessing language skills in the home language versus the language of teaching of the institution was mentioned. The test developers of the PET have approached this from a perspective that attempts to regard both arguments. The PET Verbal Reasoning and Quantitative Reasoning subtests are translated into different languages to enable assessment in a
language with which the test taker will be comfortable. This separates verbal reasoning ability from competence in the language of instruction, whilst still providing a relevant assessment of both skills. Despite the translation of these subtests, proficiency in the language of instruction is still recognized and assessed through the Hebrew Proficiency Test. Therefore, although verbal ability in a test taker’s mother tongue is included in the calculation of the total PET score, the ability to function at an acceptable level in the language of instruction cannot be disregarded.

Apart from the basic language skills that are assessed by the Hebrew Proficiency Test (HPT), the assessment also includes a composition, which emphasizes the need for students to not only be able to read and understand the language of instruction, but also to communicate their own thoughts in that language. It may be useful to add this task to the NBTs in South Africa, in order to obtain a sample of writing from test takers.

5.2.3. Assessment Procedure for the Psychometric Entrance Test (PET)

Currently the National Institute for Testing and Evaluation (NITE) tests around 65 000 HE applicants every year in group-administered sessions (NITE, 2007). NITE assumes responsibility for scheduling appointments with test takers (during five main testing periods throughout the year), administering and scoring tests, converting results to a scaled score, and relaying test results to institutions and test takers (NITE, 2007). It is also involved in areas of research such as the systems of selection employed by HE institutions, student drop-out variables and preferred fields of study. In order to write the PET a test taker must register with NITE. NITE is a separate entity that works independently from HE institutions and therefore registering with NITE does not
represent an application for HE studies. Any application to study at a HE institution must be made with the institution concerned. After registration the test taker can select a date and venue from the available options, and prepay a nominal fee for the testing (NITE, 2007).

Test results are treated as valid for HE admissions for seven years. Test takers may however retake the test as many times as they desire, but only at intervals of between four and five months. NITE publishes the dates on which a test taker may retake the test and no exceptions are allowed. The PET is administered in locations around Israel and abroad. Test takers are not allowed to use calculators or dictionaries during the test (NITE, 2007).

5.2.3.1 Discussion

The experience of implementing a national admissions test in Israel has impacted on the evolution of the testing system. Firstly, all aspects related to the test are conducted by a central organization, the National Institute for Testing and Evaluation (NITE), which is accountable to all the major universities. This curbs competition between institutions and over-involvement by one interested party, had the institute been located at one HE institution.

Secondly, the agency takes responsibility for all test-related communication with applicants, from information sharing, test booking, communicating test results, and appeals for retesting. As the retest policy of the agency is centrally implemented, whilst admissions decisions are made by the HE institutions themselves, admissions departments cannot be influenced to make exceptions for individual applicants, by for
instance allowing a test taker to have another opportunity to improve on test results. Although HE institutions have autonomy with regards to admissions decisions, another advantage of a central agency is that it can provide a vital, yet non-competitive research function to individual institutions, whilst developing a global picture of the effectiveness of both the admissions test as well as the national admissions situation. Thirdly, NITE can easily be contacted through multiple means, including phone, mail, cellphone sms-service and online.

Fourthly, testing occurs at regular intervals throughout the year and not just during the peak admissions season. This allows enough time for applicants to attend appropriate career counselling and apply for suitable courses. It resolves a problem that is experienced by individual South African HE institutions where test takers are only assessed after an application form is received. If the application is unsuccessful test takers often have little time to consider other career options, and many end up spending a year in a course they settled for, but are incompatible with.

A fifth development is that testing is conducted in many centres across the country and abroad, thereby cutting down on the travel and other expenses that may discourage disadvantaged students from applying. Another benefit of centralized testing is that all test takers are charged a standard fee to cover costs. In a decentralized system external service providers can be contracted in to provide a localized testing service and may charge different fees depending on costs incurred, such as labour, travel expenses and venue hire.

In the next sub-section the application of test results in HE admissions in Israel is reported and discussed.
5.2.4 Higher Education (HE) Admissions and the Psychometric Entrance Test (PET)

In order for Israeli students to gain entry into HE, both PET and matriculation results are considered. Both of these components (total PET score and average matriculation mark) are added in an equally-weighted formula to form a composite score. Each institution develops cut scores for every programme. Applicants are rank-ordered on the composite score and the best are selected for each programme in accordance with the cut scores (Kennet-Cohen, Bronner, & Oren, 1999). It should be noted that the use of more than one assessment result (for example school and test results) is seen as being a good practice in admissions decision-making (Foxcroft & Roodt, 2005; Murphy, 2004).

For students applying from outside of Israel only the PET total score is however used. The reason for this is that applications are received from many different countries. The difficulties of determining the comparability of school results from all these countries, and the fact that the PET is an internally controlled assessment and results are more familiar to institutions, means that the PET results are relied on to make admissions decisions (Beller, 2001). In addition to this, as was pointed out previously, the PET has also been translated into several languages.

The importance of the PET in making Higher Education admissions decisions for Israeli and international applicants and fact that the Verbal Reasoning and Quantitative Reasoning subtests are translated into various languages places added pressure on the psychometric soundness of the test battery itself, as well as the comparability of the different language versions (Beller, 1995).
5.2.4.1 Discussion

A composite score method that combines school and test results into a single result has been favoured for HE admissions purposes in Israel. This approach may work well when, as is the case in Israel, applicants are rank-ordered and the best taken, but if employed in South Africa, it may disadvantage learners from disenfranchised backgrounds. The Higher Education Act of 1997 empowers South African institutions to set their own admissions criteria, on condition that the criteria broaden access. An admissions practice that does not take the educational background of the applicant and other socio-economic factors into account would probably not be well-received. Nonetheless, admissions criteria that consist of both school and admission test performance are being perceived to be less discriminatory than just relying on one them (Murphy, 2004).

5.2.5 Summative Comment

A final comment on the narrative overview is that although the PET has developed according to the requirements and influence of the Israeli context, many points have emerged through the experience of NITE that deserve consideration during the development of the South African NBTs. These include the national admission test that is employed, the implementation of the testing programme, and the method by which the test results are employed for HE admissions purposes. In the next section specific research on the PET and the admissions criteria of Israeli HE institutions are presented.
5.3 Systematic Review

The articles found for this study were systematically assessed and grouped into the main areas of investigation. The areas that emerged were reliability, validity, bias, specifications related to general psychometric issues, and coaching. This is similar to the areas that were identified by Linn (1990) and Mcdonald, et al.(2001). The results and discussion of the review are therefore structured around each of these topics. Due to the broadness of the areas of investigation, the results and discussion components have been combined for each theme, in order to promote cohesion. For ease of reference, the broad themes that emerged in the narrative overview and systematic review are presented in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Theme</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components of the national admission test</td>
<td>5.2.2</td>
</tr>
<tr>
<td>Implementation of the national admission test</td>
<td>5.2.3</td>
</tr>
<tr>
<td>Use of the test results for Higher Education admission</td>
<td>5.2.4</td>
</tr>
<tr>
<td>Reliability of the admission test</td>
<td>5.3.1</td>
</tr>
<tr>
<td>Validity of the admission test</td>
<td>5.3.2</td>
</tr>
<tr>
<td>Bias in the admission test</td>
<td>5.3.3</td>
</tr>
<tr>
<td>Psychometric aspects of the admission test</td>
<td>5.3.4</td>
</tr>
<tr>
<td>Coaching for the admission test</td>
<td>5.3.5</td>
</tr>
</tbody>
</table>

5.3.1 Reliability

The 3 studies on the reliability of the Psychometric Entrance Test (PET), that were found for this study, primarily compared the reliability of the test for different groups, in order to identify potential non-equivalence in the measure across groups.

High reliability was found for Hebrew and Russian students on the PET Verbal Reasoning, Quantitative Reasoning and English subtests as well as the PET total score
Reliability on the total score was very high with coefficients ranging between 0.91 and 0.95 (Beller, 2005). A study on Arabic students found lower reliability for the Verbal Reasoning subtest (0.68). It was felt that some of the items in the Arabic Verbal Reasoning test were too difficult. The test was subsequently tailored with the addition of easier items which elevated the reliability to 0.84 (Beller, et al., 1999). The split-half reliability of the Hebrew and redeveloped Arabic Verbal Reasoning test was found to be equally high for both languages (Rapp & Allalouf, 2003).

Generally, reliability was found to be high for translated versions of the PET, which indicates consistency of measurement across groups. Reliability scores generally met the .85 requirement suggested by Aiken (2000). The fact that reliability had to be increased through the redevelopment of the Arabic version again points to the importance for each group to be treated as a separate entity, especially when problems with equivalence are known to exist. By conducting appropriate, group-specific research, problems can be identified and corrective measures taken to minimize inequality.

5.3.1.1 Statistical methods used to investigate reliability

Only two of the above studies reported the statistical method that was used to determine reliability (Beller, 2005; Rapp & Allalouf, 2003). Both used correlation coefficients. This method was referred to in Chapter 3 (Section 3.4.2), and provides a simple indication of the reliability of a measure. Reliability studies form an important part of psychometric investigations and the use of correlations are justified for different samples (Anastasi & Urbina, 1997).
5.3.2 Validity

The validity studies centered mostly on predictive validity, with only one study (Sireci, 2003) exploring the construct validity of the PET. Predictive validity studies were conducted for general samples of test takers, for different cultural, gender, and age groups, for students with disabilities, and for students writing the PET in their home language versus a second language. The results of the studies are reported and discussed in the next sub-sections followed by an overview of the statistical techniques that were used to determine predictive validity.

5.3.2.1 Construct validity of the Psychometric Entrance Test (PET)

Research was conducted on the construct validity of the Verbal Reasoning subtest for Hebrew and Russian students using weighted dimensional scaling (Sireci, 2003). The two tests were found to tap similar constructs, meaning that the tests measured the same underlying skills (Sireci, 2003). This type of research is important when tests are translated and would be crucial when the South African National Benchmark Tests (NBTs) are translated in future.

5.3.2.2 Predictive validity for general samples

A study of the predictive validity of the PET and Bagrut of first-year grade point average (FGPA) for 6,359 students found that both components contributed positively to the predictive validity of admissions decisions (Kennet-Cohen & Bronner, 1998). The study also concluded that traditional corrections that are applied to compensate for range restriction resulted in underestimation of FGPA.
A further study of 24,969 students by Kennet-Cohen et al. (1999) found a moderate correlation of 0.55 between the PET and Bagrut Composite score and FGPA. The Composite score also predicted College GPA (i.e., the average obtained through all years of study) more accurately than FGPA. A slightly stronger correlation existed between the PET and FGPA (0.45) than Bagrut and FGPA (0.41). This finding was thought to be due to the fact that Israeli students only enter Higher Education (HE) after two to five years compulsory military service. Another possible explanation is that the Bagrut, which is the average of the standard government controlled exam and the internal school mark, could be affected by variability in the latter. The study also found that predictive validity could be improved by breaking the overall sample down into the various faculties that students registered in, thereby grouping students based on general area of study (Kennet-Cohen, et al., 1999). Kennet-Cohen, Bronner, and Cohen (2003) also found the Verbal Reasoning and Quantitative Reasoning subtests to contribute more to predictive validity than the English subtest.

In general the findings for overall predictive validity were similar to the international trend of 0.40 mentioned by Hughes (1989) in Chapter 3 (Section 3.4.3). The finding that predictive validity improved when samples were divided into faculties should be investigated in the South African setting, as it may prove useful for HE admissions.

5.3.2.3 Predictive validity for cross-cultural samples

Cross-cultural studies on predictive validity resulted in varied findings. A large study (N=41,314) of data from six years found differences in the predictive validity of the Composite score, PET, and Bagrut for Hebrew and Arabic students (Turvall, Bronner,
Kennet-Cohen & Oren, 2008). Again, correlations with first-year grade point average FGPA were strongest for the Composite score, but definitely more accurate for Hebrew (0.49) than Arabic (0.38) students (Turvall, et al., 2008). A study comparing the Hebrew and Russian version of the PET found a similar predictive validity (0.46) for the two groups (Beller, 2005).

A 12-year study that included 134,513 students was conducted in Israel between 1985 and 1996 to investigate the predictive validity of the PET and Bagrut across gender for Hebrew, Russian, and Arabic students Gafni, Beller, & Bronner, 2000). The study found that all predictive variables correlated positively with FGPA, and predictive validity was always higher for females than males. The greatest predictive validity existed between the Composite score and FGPA for Hebrew and Arabic females (0.48 and 0.46) and males (0.43 and 0.34). The Composite score had the highest predictive validity followed by the Bagrut and finally the PET. School results were not available for the Russian students, but prediction on the PET was similar for both sexes (0.34 to 0.35) (Gafni, et al., 2000).

The research found differences in predictive validity for separate cultural groups. The Composite score, for instance, always had higher predictive validity for Hebrew students than Arabic students, who are the two largest population groups in the Israeli education system. These differences found between ethnic groups as well as gender again underlined the importance of researching these variables in predicting academic success (Cole, Muenz & Bates, 1998; Huysamen & Raubenheimer, 1999; Seymour, 2002).
5.3.2.4 Predictive validity for different age groups

Israeli students tend to enter HE after years of military service, resulting in students from a variety of age brackets. In research comparing 795 students by age group the predictive validity of the PET total score was found to be the lowest for applicants aged 30 and above (Zeidner, 1987). No significant differences were found for the other age brackets (18-21, 22-25, 26-29), meaning that predictive validity on the PET remained consistent for first-year students under the age of thirty. Despite the differences observed between groups both school results and performance on admissions tests were proven to add unique value that contributed positively to the predictive validity of the admissions decision-making process.

Validity studies have been shown to be important as they flag potential issues for admissions authorities, such as the influence of cultural group, gender, or age on predictive validity (Beller, 2001). An example of this is the study above about the predictive validity of the Psychometric Entrance Test (PET) for different age groups (Zeidner, 1987). In the face of the growing mature learner market internationally and in South Africa, this finding is important and demands further exploration on the extent of its impact in decision-making.

5.3.2.5 Predictive validity for students with disabilities

A study on the predictive validity of the selection process for students with disabilities found that although all variables correlated positively with first-year grade point average (FGPA), the Composite score tended to be highest, although the strength varied for students depending on the specific disability (Oren & Even, 2005). Prediction
was most accurate for blind students \((r=0.5)\) and least valid for students with learning disabilities \((r=0.22)\). The PET total score underpredicted FGPA for students with hearing and learning disabilities. The source of this was specifically the English subtest, whilst the quantitative subtest overpredicted performance. The Bagrut overpredicted performance for blind and hearing impaired students. Generally however, the PET and Bagrut had similar predictive profiles although each component has differential additive value for the Composite score. The predictive validity of the PET was also higher for disabled students who were given special testing conditions, such as extended time to complete the test, than those who did not. The authors furthermore suggested that the effect of variability in the FGPA of students with disabilities on predictive validity should be investigated (Oren & Even, 2005).

5.3.2.6. Predictive validity for tests of ability in first language and second language

In Chapter 3 (Section 3.4.4.2), the debate regarding the predictive validity of home language versus second language was alluded to. Some of the research outputs found for the systematic review included data on the predictive validity of the PET subtests and academic performance. A study of Hebrew and Arabic students found higher correlations for verbal reasoning in home language and FGPA (0.32 and 0.26) than English proficiency and FGPA (0.21 and 0.19) (Turval, et al., 2008). This finding was confirmed for Hebrew and Arabic students, but the reverse was true for Russian students (Gafni, et al., 2000).

The studies on the predictive validity of home language versus second language are relevant for South Africa, as they suggest that admission tests should not only be
developed in the language of instruction at HE institutions, but should be translated into the languages with which test takers are most comfortable.

5.3.2.7 Statistical techniques used to investigate predictive validity

The statistical techniques that were used to investigate predictive validity were not always reported by the researchers (Beller, 2001). Of the studies that reported on statistical techniques, most used correlations (Beller, 2005; Gafni, et al., 2000; Kennet-Cohen, Bronner, & Oren, 1999; Oren & Even, 2005; Turval, et al., 2008; Zeidner, 1987). According to the different approaches discussed in Chapter 3 (Section 3.4.3) these correlation methods investigate predictive validity in a simplistic, but effective way (Seymour, 2002), by purely looking at the relationship between the predictor and criterion variables. Although correlations are not complex, they appear to have been effective in establishing the relationship between the predictor variables of the PET, Bagrut, and Composite score on the one hand, and the predictor variable of GPA on the other.

Only one study used regression coefficients (Kennet-Cohen & Bronner, 1998). Regression coefficients are the results of multiple regressions, which investigate how well a linear combination of predictor variables (like test results, school results, cultural group, etc.) predict a criterion variable (Howell, 1989). This method is more complex than correlations, but is useful to establish the contribution made by each predictor variable towards the criterion variable (Anastasi & Urbina, 1997). Both correlations and regression coefficients play important roles in determining the predictive validity of all components, including admission tests, involved in the making of admissions decisions.
Investigations in South Africa on the predictive validity National Benchmark Tests (NBTs) should therefore include both of the methods. In the next section the results for studies on bias are detailed and discussed.

5.3.3 Bias

Of the 10 studies that investigated bias six focussed on differences in performance between cultural groups on the Psychometric Entrance Test (PET) and Bagrut, two focused on gender differences and one study each on students with disabilities and second-language test performance. The findings of these studies, as well as the statistical techniques that were used to investigate bias, are reported and discussed in the following sub-sections.

5.3.3.1 Cross-cultural bias

A cross-cultural study between Hebrew and Arabic students found that Hebrew students performed better than Arabic students on the PET, whilst the reverse was true for the Bagrut (Turvall, et al., 2008). Hebrew students also had a higher first-year grade point average (FGPA).

A number of cross-cultural studies investigated the comparability of the Verbal Reasoning subtest for different language versions. The Hebrew and Arabic versions of this subtest were found to be unequal by Rapp and Allalouf (2003), based on comparisons of test performance between the two language groups. Through statistical modeling the possibility was investigated that the anchor items, which are constructed in Hebrew and then translated into the other languages, were the source of the inequality
between the Hebrew and Arabic versions. Rapp and Allalouf (2003) found that equality did not improve through similarity or dissimilarity of anchor items. Another approach was that the Arabic version of the Verbal Reasoning subtest was changed through the inclusion of easier items (Beller, et al., 1999). This however appears not to have had the desired impact on test-performance as subsequent research still found that Arabic test takers performed one fifth of a standard deviation below their Hebrew counterparts (Rapp & Allalouf, 2003). Inequality was therefore not the result of the anchor items.

Differential Item Functioning (DIF) studies on the Hebrew and Russian version of PET Verbal Reasoning found that 34% of items functioned differently across languages. These items had not been previously identified by panels of language experts as being biased. Analogy and sentence completion items exhibited the most DIF, whilst logic items had the lowest DIF (Allalouf & Sireci, 1998; Sireci, 2003).

The test construction and translation methods employed for the PET are stringent. Once the tests have been translated, back-translated and checked by expert panels for potential bias they are administered to separate monolingual groups for equating (Rapp & Allalouf, 2003). The main problem with such a practice is that it is based on the assumption that the source language and target language versions retain the same meaning and psychometric properties (Rapp & Allalouf, 2003). Amongst the problems that have been encountered are that item length is affected by translation, that the translated anchor items are not culturally representative or relevant for all the languages, that differences in ability levels of groups may vary, and that differential performance by groups may be due to cultural factors, motivation levels, familiarity with the format, and so forth (Rapp & Allalouf, 2003). A number of items that functioned differentially were
not identified by experts and the usefulness of DIF as a statistical checking device for subtle instances of bias was evident.

Despite the inequalities in translated versions Beller (2005) motivated that testing in home language and applying equating methods to enable comparison between test takers of different language group was still preferable to testing in a second language. Beller (2005) did not elaborate on the specific reasons for this argument, but it is rooted in good assessment practices (International Test Commission, 2000).

5.3.3.2 Gender bias

The investigation of bias for gender groups found that males generally performed better on the Psychometric Entrance Test (PET) than females, especially on the Quantitative Reasoning subtest, whilst females tended to outperform males on the Bagrut (Gafni, et al., 2000). Using the Mantel-Haenszel DIF detection method one third of items on the Verbal Reasoning and Quantitative Reasoning subtests were found to display moderate to high levels of DIF. These items tended to tap fields that are stereotypically viewed as domains of interest for either males or females, and it was suggested that items of this nature be removed from the PET. Almost none of the items in the English subtest had unacceptable DIF levels (Gafni, 1990).

5.3.3.3 Language bias

A study on second language proficiency was conducted by Allalouf (2004) to compare performance of Arabic and Russian students on the Hebrew Proficiency Test, a compulsory component of the PET for test takers who opt not to write the PET in
Hebrew. Arabic students performed better than Russian students, and 42% of items were found to function differently between the groups through Mantel-Haenszel analyses, mostly in favour of Arabic test takers. Sentence completion items had the highest DIF, and reading comprehension the lowest. Vocabulary, verb, and preposition items tended to favour Arabic students, and text comprehension and restatement items Russian students. Proficiency in Hebrew correlated higher with verbal reasoning ability in Arabic (0.54) than Russian (0.4) which could be because Arab students had studied Hebrew for longer than Russian immigrant students. The study found that removing high DIF items positively affected validity, but the authors proposed a more radical solution of constructing separate versions of the Hebrew Proficiency Test for the different language groups based on DIF studies (Allalouf, 2004). This would eliminate the types of test items that are problematic for each test taker group, whilst only including those items that were the best indicators of proficiency in Hebrew for each group. Similar research and tailoring of the National Benchmark Tests (NBTs) in South Africa could lead to fairer test of language ability for different population groups.

5.3.3.4 Bias against students with disabilities

No bias against disabled students was found on the PET as they tended to perform similarly to non-disabled students on computerized versions of the test (Moshinsky & Kazin, 2002). The International Test Commission ((2000, 2.3.14) places the responsibility on test developers and users to minimize the impact that disabilities may have on test performance. The finding of the above study is encouraging for students with
disabilities, who often have to overcome many other obstacles en route to further their education and career aspirations.

5.3.3.5 Statistical methods used to investigate bias

The most commonly used method for determining bias in the above studies was the Mantel-Haenszel, which was employed by 4 studies (Allalouf, 2004; Allalouf & Sireci, 1998; Gafni, 1990; Sireci, 2003). This technique investigates differential item function whilst making adjustments for confounding variables (such as cultural group, language group, gender, and age) when analyzing test performance (Allalouf, 2004).

Other techniques were also employed to investigate bias. Turval et al (2008) and Gafni et al (2000) used descriptive statistics. Other studies utilized correlations (Rapp & Allalouf, 2003), item analysis (Beller, 2005), and t-tests and Analysis of Variance (ANOVA) (Moshinsky and Kazin, 2002). All of these techniques can be used in South Africa to add value to research into bias on the National Benchmark Tests (NBTs).

5.3.4 Psychometric Issues

Six studies were found that researched a variety of properties related to testing. The major areas included research on test items, equating different versions of the tests, the equivalence of different test modes, the influence of test time limit on validity, and the effect of guessing behaviour on test performance. The results will be presented and discussed in the following sub-sections.
5.3.4.1 Types of test items included in the Psychometric Entrance Test (PET)

A study conducted by Kennet-Cohen et al. (2003) researched items that were the most time efficient whilst adding most to the validity of test results. In the PET Quantitative Reasoning subtest algebra, geometry and quantitative comparisons were found to add more value than graphs and tables. For Verbal Reasoning analogies were better than vocabulary questions. Reading comprehension and restatements also contributed more than sentence completion questions. By increasing the number of better performing items, the time taken to complete the test would be lowered whilst increasing validity. This finding is important for the South African National Benchmark Tests (NBTs) as similar research could inform the development of the battery and potential tailoring of tests for different groups of test takers.

5.3.4.2 Methods for equating different versions of the Psychometric Entrance Test (PET)

To avoid over-familiarity with the test the National Institute for Testing and Evaluation (NITE) maintains a database of items from which new test versions are constructed. It is important for test developers to be able to compare different versions of a test to maintain the level of test difficulty, as well as validity and reliability. Rapp (1999) investigated 19 different versions of the PET and found that linear equating produced similar results to equipercentile equating. Both methods investigate the general performance of test takers on each version of a test.

In the equipercentile method, test scores are translated into percentiles for each version of a test. The end result is that an equal percentile score on both test versions are considered as indicating the same level of achievement in both (Anastasi & Urbina,
1997). In linear equating test scores for each test version is converted into a standardized score. Means and standard deviations are then computed. The score obtained on the test is then dependent on the difficulty level of the test, which allows for test results from different of varying difficulty levels to be equated (Aiken, 2000). Linear equating is easier to perform, as the method for converting test scores to standardized scores are succinct. Linear equating also requires lower sample sizes than equipercentile methods, and by establishing its validity, much time can be saved in equating new versions of the PET with established versions.

Three aspects of the above research are important for South Africa. Firstly, in order to avoid over-familiarity with the National Benchmark Tests (NBTs) new items and test versions will have to be constructed at regular intervals. A database of the difficulty level of test items would assist in the construction of new test versions. Secondly, research will be required to determine the best method of equating different versions of the NBTs. Thirdly, another database of the comparability of different test versions will need to be maintained to assist Higher Education (HE) institutions in setting cut scores for each version of the test.

5.3.4.3 Equivalence of the paper-and-pencil and computerized versions of the Psychometric Entrance Test (PET)

The National Institute for Testing and Evaluation (NITE) constructed a computerized version of the PET for administration to disabled students (NITE, 2007). The value of such a test is that font sizes can be changed to accommodate visually challenged individuals, and special answering mechanisms can be implemented
according to physical disability. Research on the computerized PET found no difference between this mode and the paper and pencil version for non-disabled students. Furthermore, no difference was found in performance on the computerized PET between disabled and non-disabled students, thereby confirming it as an equally valid instrument (Moshinsky & Kazin, 2002).

In the modern age computers can enable test developers to create and tailor tests to maximize both the assessment experience and information obtained from test takers. A computer can act as an ever-present and -observant test administrator that records test information and tailors the assessment for each test taker, whilst providing valuable data for test developers and researchers. According to the International Test Commission (ITC) special arrangements should be made for disabled test takers or those who special testing arrangements (2000, Guideline 2.3.14). Computerized testing can offer solutions in accommodating these needs and improve the validity of test scores.

5.3.4.4 Time limits on the Psychometric Entrance Test (PET)

The PET is a timed test, with time limits for each subtest. Kennet-Cohen et al. (2003) investigated the best method of setting time limits and found an overall limit to be more effective than individually timed subtests. By changing to a general time limit and allowing test takers to complete subtests at their own pace, the predictive validity of the PET was found to increase by 20% (Kennet-Cohen, et al., 2003).

At present, it is unclear what the time limits for the National Benchmark Tests (NBTs) will be, but this factor will have to be researched carefully by the test developers to maximize the performance of test takers. At the Nelson Mandela Metropolitan
University (NMMU) in South Africa, for instance, the Access Assessment Battery (AAB) is untimed. The benefit of this practice is that test takers do not feel rushed to complete the tests, allowing them to proceed at a pace with which they feel comfortable (Koch, 2005; Seymour, 2002).

5.3.4.5 Differences in guessing behaviours on the Psychometric Entrance Test (PET)

The PET mostly employs a multiple choice format. Test takers are encouraged not to leave out any questions and to rather guess an answer than omit it (NITE, 2007). Students are not penalized for incorrect answers, which means that those that take a chance by guessing may obtain higher scores than those who do not. Despite these instructions Arabic, Russian and French test takers were found to leave out substantially more items than Hebrew, English and Spanish students (Gafni & Estela, 1988). This finding indicates that cultural differences still affect test-taking methods at the most basic level, even when clear instructions are provided. The authors suggest that test takers should be alerted to this fact and greater emphasis should be placed on guessing in the test instructions (Gafni & Estela, 1988).

5.3.5 Coaching

Coaching for the PET has become an important part of the application process for many applicants to Israeli HE institutions. Allalouf and Ben-Shakhar (1998) reported that the number of applicants that participated in coaching courses for the PET escalated from one percent in 1984 to 77% in 1996. This has sparked growing concern over the potential influence of such courses on the stability of test scores.
Allalouf and Ben-Shakhar (1998) investigated the impact of coaching on the PET by randomly assigning 274 applicants into an experimental group and control group. The experimental group received 40 hours of specialized coaching on the PET, whilst the control group did not. Dependent sample t-tests on the pretest-posttest data indicated that the experimental group showed significant improvement on the PET Verbal Reasoning and Quantitative Reasoning subtests, with scores of about 25% of a standard deviation higher. The largest gains were obtained in the Quantitative Reasoning subtest. Despite the increases observed there was no significant improvement in predictive validity and it did not result in prediction bias against those who had not received coaching. The study also found a measurable improvement in the posttest results for the control group, and concluded that simply taking the test a second time resulted in the greatest improvement in test performance (Allalouf & Ben-Shakhar, 1998). According to Beller (2001) this finding is supported by other research. A study by Oren in 1992 found that the greatest improvement in PET scores resulted from retaking the test. This was followed by the development of test-preparation booklets, and then by coaching schools (Beller, 2001).

The research above appears to indicate that specialized coaching results in only minor gains for test takers. Greater gains are obtained by retaking the test, which may be due to increased familiarity with the test-format and a decrease in the anxiety that may be experienced in the first test session. Test developers could improve the testing experience as well as the predictive validity for test takers by providing in-depth test information, which includes examples of test questions or mock-tests.
5.4 Summative Comments

In this chapter the results of the narrative overview and systematic review were presented and discussed. The research indicated the similarities that can exist in different countries and that although the educational context evolves according to socio-economic needs and political influences, the outcomes can have considerable similarities. The findings highlighted the importance of the test that is used for Higher Education (HE) admission, the implementation of a national admission testing programme, and how the test results are used to make admissions decisions. The findings furthermore concur with international research that issues around reliability, validity, bias, psychometric matters, coaching, and aspects of implementation are important for the effectiveness of admissions tests.

The ITC (2000) recommends that information on differences in performance across groups should be available, and the National Institute for Testing and Evaluation (NITE) has succeeded in portraying the difficulties that have been experienced in obtaining equivalence in test performance and predictive validity across groups on the Psychometric Entrance Test (PET). The reality for admissions testing is that true equivalence may only remain an ideal, but a commitment to research and report on the reality of test results can serve to sensitize test users of the potential pitfalls of over-reliance on a single measure (Foxcroft & Roodt, 2005).

Furthermore, the ITC recommendation, mentioned in Chapter 3 (Section 3.4.3), to combine test results with other sources of information to improve the effectiveness of decisions, was supported by the Israeli studies (Kennet-Cohen, et al., 1999; Turvall, et al.,
2008), which is a positive indicator for the National Benchmark Tests (NBTs) to build on.

In the next chapter some of the recommendations that have flowed from the findings of this study, and which may guide aspects of the implementation of the NBTs in South Africa, are highlighted.
Chapter 6
Conclusions and Recommendations

6.1 Introduction

This chapter serves to conclude the objectives of this study. The study set out to achieve three objectives. The first was to conduct a systematic review of admissions testing practices in Israel, whilst the second was to relate those themes to the South African context. These tasks were performed in Chapter 5. Chapter 6 deals with the third objective, which is to construct a set of admissions testing recommendations to guide the development of a national admission test and admissions practices, that are grounded in the research findings.

Through this study the value of the systematic review method in the area of psychological assessment has been demonstrated. The methodology, which has been developed and refined in the stringent fields of medical and scientific research can be adapted for use in the social sciences. Research that involves psychological attributes of people often yields conflicting results, thereby reflecting the complexity of humanity. In this study for example the research outputs that were analyzed contained contradictory findings for the same phenomenon, especially when cultural differences were investigated. Despite these incongruencies, relevant information that has practical applicability can be obtained through this systematic review.

In the following section the major themes that emerged from the systematic review are recapped and linked to recommendations for the implementation of the NBTs. Finally, the limitations of the research are considered and recommendations for future
6.2 Conclusions and Recommendations

6.2.1 Theme one: The admission test (domains, test format and language)

In the Israeli setting the Psychometric Entrance Test (PET) assesses verbal and quantitative reasoning abilities, language skills, and for those who prefer not to write the test in the language of instruction in Higher Education (HE), proficiency in that language is also assessed. This adds support for the domains that the National Benchmark Tests (NBTs) will tap in South Africa, which include academic literacy (including verbal reasoning), quantitative literacy, and mathematical proficiency.

Both the PET and the NBTs follow a multiple-choice format. The fact that the PET has retained this format for decades, point to its utility in assessing, scoring and processing the results of large groups of test takers. This format should therefore be retained for the NBTs, until further research indicates differently.

One area that requires further research in South Africa is the issue of testing in home language versus the language of instruction in HE. If the Israeli example is followed the NBTs will have to be translated into different languages, and a further measure of competence in the language of instruction at HE level will have to be developed for those test takers who do not have English or Afrikaans at a Home Language level. This route would take several years to research and implement. However, it could be argued that not following this route is not an option if fair assessment practices are to be employed when using the NBTs (Koch, 2005).

There has been an increase in multi-language tests, especially in the admissions
and credentialling fields (Wang, Wang, & Hoadley, 2007). The increase is due to factors such as globalization and increased multiculturalism within countries. To redress construct-irrelevant variance in test scores in multilingual and multicultural contexts, due to proficiency in a specific language, the test should be translated into different languages. Furthermore, the equivalence of the translated versions of the test need to be empirically proven (Hambleton, 2004). The the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), the International Test Commission’s test use guidelines (2000) as well as the guidelines for adapting and translating tests (2001), and the guidelines issued by the Psychometrics Committee of the Professional Board for Psychology (2002), all stress the importance of assessing test-takers in the language with which they are most proficient, if fair assessment practices are to be followed.

Consequently, in view of the fact that:

- the availability of multiple language versions of tests are becoming the norm rather than the exception internationally,
- a rigorous methodology exists to develop and validate multiple language versions of a test, and
- good, ethical assessment practice demands that a test taker be assessed in the language in which he/she is most proficient in,

the NBTs cannot only be available in the medium of instruction used in HE. There will have to be multiple language versions. The manner in which this has been achieved in the PET can serve as an example.
6.2.2 Theme two: The implementation of a national admission test

In Israel the PET is administered by the National Institute for Testing and Evaluation (NITE), which is accountable to Higher Education (HE) institutions. Higher Education South Africa (HESA) should establish a similar institute to administer the NBTs. This institute should be a separate entity, with dedicated staff, and should be accountable to HESA, possibly even as a substructure of HESA. To ensure that HESA remains the primary stakeholder of the NBTs and the testing programme, neither the test nor the assessment function should be outsourced to a private concern. To avoid over-representativeness of one HE institution, this institute should also not be housed within any institution.

The purpose of this institute should be to implement and continuously refine, update and research the NBTs, to develop materials to enable test takers to prepare for the tests (see also 6.2.8), and to provide test results to HE institutions and test takers. The institute would be responsible for marketing the NBTs and for communicating with test takers. It would be tasked with arranging test venues, setting dates for assessment, and receiving fees payable. It should also be responsible for implementing the retest and appeals policies, which will have to be finalized by HESA in consultation with HE institutions and community stakeholders, before the first assessments are conducted. A website must also be constructed for the NBTs as a marketing tool and for communication with test takers.

Apart from administering the NBTs, this institute should be engaged in ongoing research on the effectiveness of the NBTs in adding value to HE admissions. This institute must also provide relevant information to HE institutions, to inform student
academic development initiatives, and curriculum responsiveness and refinement.

6.2.3 Theme three: The fair use of test and school results for Higher Education (HE) admission

Research should be conducted on the most effective method in which test results should be used for HE admission. Although Higher Education (HE) institutions should be responsible for developing and researching their own admissions criteria, Higher Education South Africa (HESA) should protect the rights of test takers, by ensuring that the NBTs test results are used in the fairest, and most ethical manner. In Israel the Psychometric Entrance Test (PET) and school results (Bagrut) are combined into a composite score. Test takers are rank-ordered on this score and the best are admitted to an HE institution. If a similar system is employed in South Africa research will first have to be conducted to ensure that it will not discriminate against test-takers from educationally disadvantaged backgrounds. Furthermore, the benefits of a more holistic, developmentally-focused approach to HE admission using a profile approach for combining test and school results should be explored (Koch, Foxcroft & Watson, 2001).

6.2.4. Theme four: Reliability of the national admission test must be established

The reliability of the NBTs, as an accurate and effective assessment of academic and quantitative literacy, and mathematical proficiency should be researched. Research findings should be available to stakeholders and the research findings must be utilized to inform the adaptation of the NBTs. All the standards set for psychometric tests indicate that the reliability of a test needs to be proven to be satisfactory for the context in which it
6.2.5 Theme five: Establish the validity of the national admission test

The validity of the NBTs should be confirmed through research. Using the validity studies conducted on the PET as an example, specific areas of focus should be on the construct and predictive validity of the measure for different faculties, different cultural and age groups, for both genders, for home and second language, and for test-takers with disabilities. To determine predictive validity, studies should use both correlational and regression techniques. Research findings should be duly considered and relevant changes to the NBTs should be implemented, as well as the method in which the test results are used in Higher Education admission.

6.2.6 Theme six: Bias within the national admission test must be researched

Studies should be performed on the potential bias that may exist in the NBTs, as well as bias that may result as a consequence of how the test results are applied in decision-making (Jencks, 1998; Murphy, 2004). Bias studies should include different cultural and language groups, both genders, and test-takers with disabilities. As is the case with the PET research, a range of statistical techniques should be employed to investigate bias, including correlations, item-analysis, t-tests, analysis of variance (ANOVA), and differential item functioning, especially by employing the Mantel-Haenszel technique.

Steps should be taken to minimize bias and to promote fairness in the NBTs in terms of test results, as well as the use of test results. Research reports should be placed
on the website dedicated to the NBTs, to promote openness about the limitations of the measure.

6.2.7 Theme seven: Psychometric aspects related to the national admission test should be explored

Research should be conducted to determine the types of items that contribute significantly to the validity of the National Benchmark Tests (NBTs). Such research should be conducted on for separate cultural and language groups on the subtests of the NBTs. If differences arise, and a harmonious solution cannot be found for all groups, separate versions of this subtest will have to be developed for each group.

The most optimal method of setting time-limits on the test should be investigated, and the possibility of not having any time-limits should be entertained. Furthermore, the differences in guessing behaviours between groups need to be studied, and test takers should be alerted to the influence of guessing behaviour on test results.

Finally, the NBTs should be computerized as the benefits of such a system in gathering information, standardizing administration, and the ability to tailor the assessment for disabled test takers in particular will add value to the testing programme. As demanded by good assessment practices when using computer-based tests, the equivalence of the paper and pencil and computerized versions should be explored for all test taker groups, but especially those who have had little exposure to computers.

6.2.8 Theme eight: Preparation for the national admission test

The impact of specialized coaching on the scores obtained on the NBTs should be
researched. Efforts should be made to develop test items that are resistant to such coaching. Preparatory booklets must be developed and made available to test takers. These booklets must contain general information on the test, the time-limits that may apply, the equipment that test-takers must bring to the test session, and examples of the types of test questions and expected format of answers. Practice-tests should be developed and placed on the website, to provide test-takers with an opportunity to obtain test experience and to minimize test anxiety. Both the lack of test-taking skills and the concomitant test anxiety have been found to be important variables that impact on test performance in South Africa (Griessel, 2005). Consequently, it is important to minimize these factors by providing practice and preparatory opportunities.

6.3 Limitations of the Research and suggestions for Future Research

Despite rigorous implementation of the methodology by this study the findings were limited by some factors.

The research only employed research outputs in English for a country where HE instruction mostly occurs in Hebrew. Although attempts were made to uncover findings that were originally reported in Hebrew, no guarantee can be provided that all possible findings emerged through this research. It is recommended that a future replication of this study involve researchers skilled in English and Hebrew, and include research outputs in both languages.

Another limitation of the study is that testing practices from only one country was researched. A larger study of this nature, that includes testing practices in a greater number of countries, may yield interesting results.
This study focussed on the effectiveness of admissions testing in terms of test performance and ability to add value to the admissions process. A benefit of testing, which was not explored in this study is the opportunity to identify developmental needs of students, in order to tailor programmes that suit the individual and to inform curriculum development. The research outputs found for this study also did not cover this topic. Further research focussing on these aspects is recommended as it could add valuable information on the practical application of test results, to inform the utility of South Africa’s National Benchmark Tests.

6.4. Concluding Remarks

This study has contributed to the understanding of national admission tests and testing practices. The potential value of admission tests to Higher Education admissions decisions were highlighted, together with the need to follow the best-practice guidelines that have been developed from international research and experience. Finally, important recommendations flowed from the research, which offers guidance for the development and implementation of the National Benchmark Tests in South Africa.
References


Conference on Adapting Tests for use in Multiple Languages and Cultures, Washington, USA.


Cooper, C., Hamilton, R., Mashabela, H., Mackay, S., Kelly, J., Sidiropolous, E.,
Johannesburg: South African Institute of Race Relations.


Council on Higher Education. (2000). Towards a new higher education landscape:
Meeting the equity, quality and social development imperatives of South Africa in
the 21st century. Pretoria, South Africa: Authors.


The Cochrane Library.

introduction to psychological assessment in the South African context (pp. 176-
191). Cape Town, South Africa: Oxford University Press.

heterogeneity and combining results from several studies in meta-analysis. In M.
Egger, G. D. Smith & D. G. Altman (Eds.). Systematic reviews in health care:
Meta-analysis in context. London: BMJ.

Institute for Race Relations.

Department of Education (1997). White paper 3: A programme for the transformation of
education. Pretoria, South Africa: Authors.
Higher_Education/National Plan.-Final Draft.htm


Professional Board for Psychology. (2002). *Policy on the classification of psychometric measuring devices, instruments, methods and techniques*. Health Professions Council of South Africa: Authors


Appendix A

Search Keywords for the Systematic Review
Search Keywords for the Systematic Review

Database searches contained different combinations of keywords from the following four categories:

1. Israel OR Israeli
2. Psychometric Entrance Test OR PET
3. Higher Education OR University OR College
4. Entrance OR Entry OR Admission OR Admissions

Searched databases included:

1. EBSCOHost
2. EBSCOHost Electronic Journals Service
3. Gale INFOTRAC
4. JSTOR
5. PubMed Central
6. Sabinet
7. ScienceDirect
8. SwetsWise
9. Google
10. Scholar Google
Appendix B

Article Classification Sheets
Article Classification Sheet 1

Article Number: 1
Title: The predictive validity of the components of the process of selection of candidates for Higher Education in Israel


Aim of Study:
1. Ability to predict FGPA.
2. Ability to predict CGPA

Sample description: All 1st year students at six Israeli universities between 1991-1993. 24 969 students with only freshman grade point average (GPA). 16731 students with freshman GPA and full College GPA. No gender or culture group breakdowns.

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Predictive validity</td>
<td>PET, HSM and FGPA</td>
<td>Correlations</td>
<td>Composite predicted better than PET and HSM separately</td>
</tr>
<tr>
<td>2. Predictive validity</td>
<td>PET, HSM and CGPA</td>
<td>Correlations</td>
<td>Predictive validity stronger for CGPA</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: Predictive validity for FGPA was stronger for those who graduated.
Article Classification Sheet 2

Article Number: 2
Title: Fairness in higher education admissions procedure: The psychometric entrance test in Arabic.


Aim of Study:
1. Comparison of performance on PET and HSM, and
2. Ability to predict FGPA for Hebrew and Arabic students. Students tested in home language.


Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Performance</td>
<td>PET and HSM.</td>
<td>Descriptive</td>
<td>Hebrew students performed better on PET and lower on HSM than Arabic</td>
</tr>
<tr>
<td>(Bias)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Predictive</td>
<td>PET and HSM on FGPA</td>
<td>Correlations</td>
<td>Predictive validity lower for Arabic than Hebrew.</td>
</tr>
<tr>
<td>validity</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: Differences in performance on PET and predictive validity for Hebrew and Arabic students.
Article Classification Sheet 3

Article Number: 3
Title: The effect of coaching on the predictive validity of scholastic aptitude tests.


Aim of Study:
1. The effect of coaching on PET results (comparing pretest and posttest results).
2. Comparing posttest performance of those who had coaching vs. those who did not.

Sample description: 1st year applicants. Total N= 274. 207 Experimental, 67 Control. No gender or ethnic group breakdown.

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Compare improvement</td>
<td>PET</td>
<td>Dependent t-tests</td>
<td>Significant improvement</td>
</tr>
<tr>
<td>2. Compare experimental vs. control</td>
<td>PET</td>
<td>Independent t-tests</td>
<td>Significant difference, especially on quantitative subtest.</td>
</tr>
<tr>
<td>3. Impact on predictive validity.</td>
<td>PET and FGPA</td>
<td>Correlations</td>
<td>Predictive validity not significantly higher for experimental than control group.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: Taking the test a second time showed greater improvement than coaching.
Article Classification Sheet 4

Article Number: 4
Title: Age bias in the predictive validity of scholastic aptitude tests: Some Israeli data

Author(s): Zeidner, M. (1987)

Aim of Study:
1. Predictive validity of PET for groups from different ages.

Sample description: Test data from 795 undergraduate applicants to an Israeli university were divide into 4 groups, 18-21, 22-25, 26-29, and 30+

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Predictive validity</td>
<td>PET composite mark and FGPA</td>
<td>Correlations</td>
<td>Poorer predictive validity of FGPA for students in the 30+ group than the other groups.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research:
Article Classification Sheet 5

Article Number: 5
Title: A cross-cultural perspective on gender differences in higher education: Admissions and scholastic achievement.


Aim of Study: Comparison of performance on PET, HSM, and prediction of FGPA for Hebrew, Arabic, and Russian male and female students between 1985-1996 at six Israeli universities. All tested in home language.
   1. Comparison
   2. Predictive validity

Sample description: N=134513

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Comparison</td>
<td>PET, HSM, FGPA</td>
<td>Descriptive stats</td>
<td>Differences between groups. Males higher on PET and females higher on HSM.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research:
Article Classification Sheet 6

Article Number: 6
Title: Evaluating Cross-lingual equating.


Aim of Study: Developing techniques for equating different language versions of the PET.
1. Comparing 2 sections of each test to its same language counterpart.
2. Comparing each section with its counterpart in the other language.

Sample description: 9500 applicants to Israeli universities. 7000 Hebrew, 2500 Arabic. Test taken in home language.

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Same language comparison</td>
<td>PET verbal</td>
<td>Correlations</td>
<td>Sections perform equally</td>
</tr>
<tr>
<td>2. cross-lingual comparison</td>
<td>PET verbal</td>
<td>Difference functions</td>
<td>Cross-lingual equating was unstable. Translated forms not equal.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research:
Possible reasons for unequal performance in translated versions could be a) that translating changes the statistical and psychometric properties of individual items and sets, b) changes in these properties could be due to cultural and educational differences, c) item length is affected by translation, d) the anchor item set may not be representative of both language groups, e) differences in ability levels of the groups, and f) differential performance in the tests due to cultural factors, motivation levels, etc.
Article Classification Sheet 7

Article Number: 7
Title: Constructing, adapting and validating admissions tests in multiple languages.


Aim of Study: Reporting on previous studies (that are only available in Hebrew). Areas include:
1. Equality of translated items of PET verbal test.
2. DIF across languages.
3. Reliability for different language groups.

Sample description: NA
Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Equality of translations</td>
<td>PET verbal</td>
<td>Not reported</td>
<td>Translated versions not equal.</td>
</tr>
<tr>
<td>2. DIF across translations</td>
<td>PET verbal</td>
<td>Not reported</td>
<td>1/3 of items still functioned differently across groups.</td>
</tr>
<tr>
<td>3. Reliability across groups</td>
<td>PET verbal</td>
<td>Not reported</td>
<td>Reliability was good for all but Arabic.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: Arabic test was changed subsequently to include easier items, which increased reliability.
Article Classification Sheet 8

Article Number: 8
Title: Translating, equating and validating scholastic aptitude tests: The Israeli case.


Aim of Study: Reporting on other studies conducted by NITE on the Hebrew and Russian versions of the PET:
1. Equality of different language versions.
2. DIF across different language versions.
3. Reliability
4. Predictive validity.

Sample description: varied. Test takers were applicants to university and were tested in home language.

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Equality</td>
<td>PET</td>
<td>Correlations</td>
<td>Performance was equal</td>
</tr>
<tr>
<td>2. DIF</td>
<td>PET</td>
<td>Item analysis</td>
<td>No serious DIF issues</td>
</tr>
<tr>
<td>3. Reliability</td>
<td>PET</td>
<td>Coefficients</td>
<td>Reliability was high for both.</td>
</tr>
<tr>
<td>4. Predictive validity</td>
<td>PET and FGPA</td>
<td>Correlations</td>
<td>Similar for both (0.46)</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research:
Article Classification Sheet 9

Article Number: 9
Title: Constructing a computerized psychometric adaptive test for university applicants with disabilities.


Aim of Study: Hebrew applicants to HE who wanted an early estimation of PET scores.
1. Comparison of computer and pencil and paper versions.
2. Comparison of disabled and non-disabled students.

Sample description:  667 (338 Computerized) (329 Paper and Pencil)

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Computer vs. P&amp;P</td>
<td>PET</td>
<td>t-tests and ANOVAS</td>
<td>No difference for mode (for gender or between those who used computers regularly vs those that did not).</td>
</tr>
<tr>
<td>2. Disabled vs. Non-disabled</td>
<td>PET</td>
<td>t-tests and ANOVAS</td>
<td>No difference between disabled and non-disabled on computerised version.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research:
Article Classification Sheet 10

**Article Number:** 10  
**Title:** Evaluating the effect of ability differences between groups and the use of a non-representative anchor on equating in cross-lingual circumstances.

**Author(s):** Rapp, J., & Allalouf, A. (2002).

**Aim of Study:** 1. Looking at the role of anchor items in contributing to DIF across language versions (Hebrew and Arabic).

**Sample description:** Only statistical modelling is reported. No sample has been specified.

**Area(s) of Investigation:**

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Similar vs. Dissimilar anchor items.</td>
<td>PET verbal</td>
<td>Statistical modelling</td>
<td>Equality did not improve through similar or dissimilar anchor items.</td>
</tr>
</tbody>
</table>

**Conclusions / recommendations / suggestions for future research:** Inequality was not due to non-representativeness of anchor items.
Article Classification Sheet 11

**Article Number:** 11

**Title:** The fairness and validity of the higher education selection system for students with disabilities.

**Author(s):** Oren, C., & Even, A. (2005).

**Aim of Study:** Comparing predictive validity of disabled students, using data from non-disabled as a benchmark.
1. Predictive validity of PET.
2. Comparing predictive validity of PET with high school mark (HSM).

**Sample description:** Students with disabilities and students without (1992-1997). No gender or ethnic group details provided. 821 learning disabled, 159 physically, 90 visually, 19 blind, 28 hearing impaired. All were given special testing conditions.

**Area(s) of Investigation:**

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Predictive validity</td>
<td>PET and FGPA</td>
<td>Correlations, group differences (D)</td>
<td>Total PET underpredicted for learning and hearing disabled.</td>
</tr>
<tr>
<td>2. Predictive validity</td>
<td>PET vs. HSM</td>
<td>Correlations, group differences (D)</td>
<td>PET and HSM has equal validity</td>
</tr>
</tbody>
</table>

**Conclusions / recommendations / suggestions for future research:** The possibility of FGPA contributing to differences should be investigated.
Article Classification Sheet 12

Article Number: 12
Title: Improving the predictive validity of a test: A time-efficient perspective.


Aim of Study: Determining types of items that were most time-efficient whilst adding to predictive validity.
1. PET quantitative.
2. PET verbal.
3. PET English.
4. Time limits of test
Sample description: 4543 first-year students from 1997. No gender or ethnic groups mentioned.

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Best item types</td>
<td>PET Quantitative</td>
<td>Correlations</td>
<td>Algebra, geometry and quant comparisons better than graphs and tables.</td>
</tr>
<tr>
<td>2. Best item types</td>
<td>PET Verbal</td>
<td>Correlations</td>
<td>Analogies better than vocabulary</td>
</tr>
<tr>
<td>3. Best item types</td>
<td>PET English</td>
<td>Correlations</td>
<td>Reading comprehension and sentence restatements better than sentence completion.</td>
</tr>
<tr>
<td>4. Time limits</td>
<td>PET</td>
<td>Correlations</td>
<td>Setting a time-limit on whole test instead of subtests raised predictive validity by 20%.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research:
Article Classification Sheet 13

Article Number: 13
Title: Differential Item functioning: Performance by sex on reading comprehension tests.


Aim of Study: 1. DIF across gender.


Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. DIF across gender</td>
<td>PET</td>
<td>Mantel-Haenszel</td>
<td>Almost 1/3 items for Verbal and Quant had dif. Almost none for English.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: The sources of DIF were specifically items that dealt with gender-stereotypical domains.
Article Classification Sheet 14

**Article Number:** 14  
**Title:** Differential tendencies to guess as a function of gender and lingual-cultural reference group.

**Author(s):** Gafni, N., Estela, M. (1988).

**Aim of Study:** Tendency to guess (which can positively affect total score)  
1. By Gender  
2. By language group.

**Sample description:** Test takers in 1984 and 1987. N=22972, Males= 12440, Females= 10532. Hebrew, English, Spanish, French, Russian and Arabic applicants tested in home language.

**Area(s) of Investigation:**

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Guessing by Gender</td>
<td>PET</td>
<td>ANCOVA</td>
<td>No difference by gender</td>
</tr>
<tr>
<td>2. Guessing by language</td>
<td>PET</td>
<td>ANCOVA</td>
<td>Hebrew, English and Spanish tended to guess more than Arabic, Russian and French.</td>
</tr>
</tbody>
</table>

**Conclusions / recommendations / suggestions for future research:** Emphasize the importance of guessing, as it can raise score.
Article Classification Sheet 15

Article Number: 15
Title: Appraising item equivalence across multiple languages and cultures.


Aim of Study:
1. Construct equivalence of translated tests (Russian and Hebrew), and
2. DIF on individual items.

Sample description: Different samples of about 1500 Russian and a maximum of 7000 Hebrew HE applicants who wrote three different versions of the PET. Applicants tested in their home language

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Construct Equivalence</td>
<td>PET verbal</td>
<td>Weighted dimensional scaling</td>
<td>Constructs were equivalent</td>
</tr>
<tr>
<td>2. DIF</td>
<td>PET verbal</td>
<td>Mantel-Haenszel</td>
<td>Over 1/3 of items had high or moderate DIF.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: Language differences remain one of the greatest problems in obtaining equivalence in translation. Qualitative analyses by expert panels can still miss great numbers of biased items. Mantel-Haenszel is a useful DIF detection method.
Article Classification Sheet 16

Article Number: 16
Title: Detecting sources of DIF in translated verbal items.


Aim of Study: 1. To report on the sources of DIF found for the PET verbal Hebrew and Russian versions. (see Article 15)

Sample description: about 1700 Russian and 6400 university applicants. Test takers were tested in home language.

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sources of DIF</td>
<td>PET verbal</td>
<td>Mantel-Haenszel</td>
<td>Sentence completion had high DIF. Reading comprehension moderate. Logic low.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: The suggested sources were changes in word difficulty, item format, translation errors, and differences in cultural relevance.
Article Classification Sheet 17

Article Number: 17
Title: The predictive validity of the components of the process of selection for Higher Education in Israel: A correction for sample-selection bias using Heckman’s model.


Aim of Study: 1. Predictive validity of PET and HSM.

Sample description: Applicants to Tel-Aviv University in 1994-1995 (no ethnic groups specified). N=6359

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Predictive validity</td>
<td>PET, HSM, FGPA</td>
<td>Regression coefficients</td>
<td>PET and HSM contributed positively to predicting FGPA.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: Traditional corrections applied to compensate for restricted sample ranges serve to underestimate the predictive validity of these components.
Article Classification Sheet 18

Article Number: 18
Title: Improving second language proficiency assessment: a differential item functioning study.


Aim of Study:
1. Using DIF to refine a test of ability in a second language.
2. Comparing proficiency in home language and proficiency in a second language.

Sample description: Arabic and Russian speakers tested on the PET (in home language) and Hebrew Proficiency Test, which must be completed for selection to HE institutions. 2200 Arabic and 1500 Russian.

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. DIF</td>
<td>Hebrew Proficiency Test</td>
<td>Mantel-Haenszel</td>
<td>42% of items functioned differently for different language groups. Sentence completion had highest DIF and Reading Comprehension lowest.</td>
</tr>
<tr>
<td>2. Correlation between home language and second language</td>
<td>PET verbal and Hebrew Proficiency Test</td>
<td>Correlation</td>
<td>Moderate correlation. Correlation depended on number of years learning Hebrew.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: Removing high DIF items improved validity, but a better solution would be to construct different versions for each language group.
Article Classification Sheet 19

Article Number: 19
Title: Linear and equipercentile methods for equating PET.

Author(s): Rapp, J. (1999).

Aim of Study: NITE has an item pool from which different versions of the PET are constructed. The study looked at two methods of equating versions.
1. Linear vs. Equipercentile method

Sample description: 19 versions of Hebrew PET administered between 1997 and 1999 with about 1000 Hebrew test takers per version.

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Comparing equating methods</td>
<td>PET (V, Q, and E)</td>
<td>Linear equating and equipercentile equating</td>
<td>Almost no difference was found in the results of the two methods.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research: Linear equating can be used as it is easier, requires smaller samples and is just as accurate.
Article Classification Sheet 20

Article Number: 20
Title: Admissions to higher education in Israel and the role of the psychometric entrance test: Educational and political dilemmas.


Aim of Study: The article reports on previous studies on the PET.
1. Predictive validity.
2. Bias of using either PET or HSM
3. Coaching

Sample description: Various. No ethnic or gender groups specified

Area(s) of Investigation:

<table>
<thead>
<tr>
<th>Area</th>
<th>Variable</th>
<th>Techniques</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Predictive validity</td>
<td>PET, HSM, FGPA</td>
<td>Not provided</td>
<td>Composite of PET and HSM had highest predictive validity.</td>
</tr>
<tr>
<td>2. Bias</td>
<td>PET, HSM, FGPA</td>
<td>Not provided</td>
<td>Using PET or HSM alone results in greater bias.</td>
</tr>
<tr>
<td>3. Coaching</td>
<td>PET</td>
<td>Not provided</td>
<td>Limited increase through coaching. Taking the test a second time resulted in greatest gain.</td>
</tr>
</tbody>
</table>

Conclusions / recommendations / suggestions for future research:
Appendix C

Summarising Map
# Summarising Map

<table>
<thead>
<tr>
<th>Theme</th>
<th>Article Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
</tr>
<tr>
<td><strong>Construct Validity</strong></td>
<td>X</td>
</tr>
<tr>
<td><strong>Predictive Validity</strong></td>
<td></td>
</tr>
<tr>
<td>- General</td>
<td>X</td>
</tr>
<tr>
<td>- Cross-Cultural/Lingual</td>
<td>X</td>
</tr>
<tr>
<td>- Gender</td>
<td>X</td>
</tr>
<tr>
<td>- Age</td>
<td>X</td>
</tr>
<tr>
<td>- Disabled</td>
<td>X</td>
</tr>
<tr>
<td>- Home lang vs 2\textsuperscript{nd} lang</td>
<td>X</td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>X</td>
</tr>
<tr>
<td>- Cross-Cultural/Lingual</td>
<td>X</td>
</tr>
<tr>
<td>- Gender</td>
<td>X</td>
</tr>
<tr>
<td>- Disabled</td>
<td>X</td>
</tr>
<tr>
<td>- Test in 2\textsuperscript{nd} language</td>
<td>X</td>
</tr>
<tr>
<td><strong>Coaching</strong></td>
<td>X</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>X X X</td>
</tr>
<tr>
<td><strong>Psychometric Issues</strong></td>
<td></td>
</tr>
<tr>
<td>- Equating versions</td>
<td>X</td>
</tr>
<tr>
<td>- Test items</td>
<td>X X X X</td>
</tr>
<tr>
<td>- Time limits</td>
<td>X</td>
</tr>
<tr>
<td>- Computerized vs P&amp;P</td>
<td>X</td>
</tr>
<tr>
<td>- Guessing</td>
<td>X</td>
</tr>
</tbody>
</table>