ACKNOWLEDGEMENTS

I wish to express my gratitude for the assistance of the following people in completing this work:

♦ My supervisor, Mr Z.C. Bergh for his guidance.

♦ Prof Frans Cilliers for teaching me that content without structure is chaos.

♦ Colleagues for their advice and willingness to help, especially Jannie Hartzenberg for the data analysis and interpretation.

♦ Participants in this study.

♦ LEANNE.
THE VALIDATION OF A SELECTION BATTERY FOR SCREENING UNIVERSITY BRIDGING – COURSE STUDENTS

by

DAVIDA – SUZANNE NGOZWANA

Submitted in partial fulfilment of the requirements for the degree of

MASTER OF ADMINISTRATION

in the subject

INDUSTRIAL AND ORGANISATIONAL PSYCHOLOGY

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR : MR Z.C. BERGH

NOVEMBER 2002
I declare that **The validation of a selection battery for screening university bridging-course students** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

(MS D. NGOZWANA)

06/05/04
(DATE)

150.287 NGOZ

UNISA
BIBLIOTEK/LIBRARY
04 OCT 2004

Class
Klas
Access
Aanwin

0001855862
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>xii</td>
</tr>
<tr>
<td>CHAPTER 1 BACKGROUND TO AND MOTIVATION FOR THE RESEARCH</td>
<td></td>
</tr>
<tr>
<td>1.1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.2 LEGAL AND SCIENTIFIC IMPERATIVES IMPACTING ON PSYCHOMETRIC ASSESSMENT</td>
<td>3</td>
</tr>
<tr>
<td>1.3 PROBLEM STATEMENT</td>
<td>4</td>
</tr>
<tr>
<td>1.4 RESEARCH QUESTIONS</td>
<td>4</td>
</tr>
<tr>
<td>1.4.1 General research question</td>
<td>4</td>
</tr>
<tr>
<td>1.4.2 Specific research questions for the literature review</td>
<td>4</td>
</tr>
<tr>
<td>1.4.3 Specific research questions for the empirical investigation</td>
<td>5</td>
</tr>
<tr>
<td>1.5 RESEARCH AIMS</td>
<td>6</td>
</tr>
<tr>
<td>1.5.1 The general aim</td>
<td>6</td>
</tr>
<tr>
<td>1.5.2 Theoretical aims</td>
<td>6</td>
</tr>
<tr>
<td>1.5.3 Empirical aims</td>
<td>7</td>
</tr>
</tbody>
</table>
1.6 PARADIGM PERSPECTIVE 8
1.7 RESEARCH DESIGN 10
1.8 RESEARCH METHODOLOGY 11
1.9 SUMMARY 13
1.10 CHAPTER DIVISION 14

CHAPTER 2 THE ROLE OF PSYCHOMETRIC ASSESSMENT IN SELECTION 16

2.1 INTRODUCTION 16

2.2 SCIENTIFIC SELECTION 16
2.2.1 The nature of scientific selection 17
2.2.2 Steps in compiling a scientific selection battery 17
2.2.3 The utility of scientific selection 20

2.3 PREREQUISITES FOR PSYCHOMETRIC ASSESSMENT 21
2.3.1 Defining psychometric assessment 21
2.3.2 Reliability 23
2.3.2.1 Procedures for the determination of reliability 23
2.3.2.2 Practical implications 25
2.3.3 Validity 26
2.3.3.1 Content validity 26
2.3.3.2 Construct validity 27
2.3.3.3 Criterion-related validity 29
2.3.4 Standardisation of testing procedures 32
2.3.5 The conversion and interpretation of psychological test results 33
2.3.6 Test bias 34

2.4 LEGAL IMPERATIVES IMPACTING ON SELECTION 36
2.4.1 The Health Professions Act 37
2.4.2 The Employment Equity Act 38
2.4.3 The Labour Relations Act 38
2.4.4 Implications for psychometric assessment 39

2.5 SUMMARY 40

CHAPTER 3 THE NATURE AND ASSESSMENT OF INTELLIGENCE 42

3.1 INTRODUCTION 42

3.2 DEFINITION OF INTELLIGENCE 43

3.3 THE CONCEPTUALISATION AND THEORIES OF INTELLIGENCE 45
3.3.1 The Galton – Cattell approach 45
3.3.2 Binet – Simon’s conception of intelligence 46
3.3.3 Spearman’s two – factor theory 46
3.3.4 Sternberg’s triarchic theory 47
3.3.5 Multiple factor theories 48
3.3.6 The hierarchical model: an integration 49

3.4 THE MEASUREMENT OF INTELLIGENCE 50
3.4.1 Individual intelligence tests 51
3.4.1.1 Binet – Simon's Intelligence Quotient
3.4.1.2 The Wechsler Scales
3.4.2 Group intelligence tests
3.4.2.1 Army Alpha and Army Beta
3.4.2.2 Raven's Advanced Progressive Matrices
3.4.3 South African - Based Intelligence Tests
3.4.3.1 Individual intelligence tests
3.4.3.2 Group intelligence tests

3.5 THE RELATIONSHIP BETWEEN INTELLIGENCE AND ACADEMIC ACHIEVEMENT

3.6 FACTORS MODERATING THE RELATIONSHIP BETWEEN INTELLIGENCE AND ACADEMIC ACHIEVEMENT
3.6.1 Bias in the assessment of intelligence
3.6.2 Optimalising the culture fair assessment of intelligence
3.6.3 Emotional intelligence

3.7 SUMMARY

CHAPTER 4 RESEARCH METHODOLOGY
4.1 INTRODUCTION
4.2 PURPOSE OF THE STUDY
4.3 DESCRIPTION OF THE POPULATION AND SAMPLE
4.4 MEASURING INSTRUMENTS
4.4.1 The Abilities, Processing of Information and Learning (APIL) Battery

4.4.1.1 The aim of the test

4.4.1.2 Description of the constructs measured by the APIL Battery

4.4.1.3 Administration

4.4.2 Raven's Advanced Progressive Matrices (APM) Test

4.4.2.1 The aim of the test

4.4.2.2 Description and administration

4.5 CRITERION DESCRIPTION

4.6 DATA COLLECTION REGARDING INDEPENDENT VARIABLES

4.7 DATA COLLECTION REGARDING THE DEPENDENT VARIABLE

4.8 RESEARCH HYPOTHESES

4.9 STATISTICAL ANALYSIS

4.9.1 Correlation coefficient

4.9.2 Regression analysis

4.9.3 Comparisons of independent groups

4.10 SUMMARY

CHAPTER 5 RESEARCH RESULTS, DISCUSSION, LIMITATIONS AND RECOMMENDATIONS
5.1 INTRODUCTION

5.2 RESEARCH RESULTS

5.2.1 Results and discussion regarding the predictive power of the APIL Battery

5.2.2 Results and discussion regarding the predictive power of Raven's APM

5.2.3 Results and discussion regarding the predictive power of the total selection battery

5.2.4 Results and discussion regarding the effect of gender on academic performance

5.2.5 Results and discussion regarding the effect of ethnicity on academic performance

5.2.6 Results and discussion regarding the effect of home language on academic performance

5.3 CONCLUSIONS

5.3.1 General conclusion

5.3.2 Specific conclusions in terms of the literature review

5.3.2.1 Conclusions in terms of the role of psychometric assessment in selection

5.3.2.2 Conclusions in terms of the nature and assessment of intelligence

5.3.3 Specific conclusions in terms of the empirical study

5.4 LIMITATIONS

5.4.1 Limitations in terms of the literature review

5.4.2 Limitations in terms of the empirical study

5.5 RECOMMENDATIONS FOR FURTHER RESEARCH

5.5.1 Recommendations in terms of the literature review
5.5.2 Recommendations in terms of the empirical study 98
5.5.3 Recommendations regarding future recruitment and selection 99
5.5.4 Recommendations in terms of training 100

5.6 SUMMARY 100

REFERENCES 101
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>SUBTESTS OF THE WAIS -- R</td>
</tr>
<tr>
<td>4.1</td>
<td>GEOGRAPHIC DISTRIBUTION OF THE SAMPLE</td>
</tr>
<tr>
<td>4.2</td>
<td>GENDER DISTRIBUTION OF THE SAMPLE</td>
</tr>
<tr>
<td>4.3</td>
<td>ETHNIC DISTRIBUTION OF THE SAMPLE</td>
</tr>
<tr>
<td>4.4</td>
<td>LANGUAGE DISTRIBUTION OF THE SAMPLE</td>
</tr>
<tr>
<td>5.1</td>
<td>PREDICTIVE POWER OF THE APIL BATTERY</td>
</tr>
<tr>
<td>5.2</td>
<td>PREDICTIVE POWER OF RAVEN'S ADVANCED PROGRESSIVE MATRICES</td>
</tr>
<tr>
<td>5.3</td>
<td>PREDICTIVE POWER OF THE TOTAL SELECTION BATTERY</td>
</tr>
</tbody>
</table>
Legal and scientific imperatives necessitate the validation of a psychometric battery before using it for the purposes of personnel selection and decision-making. The aim of this investigation is to validate a selection battery, i.e. the Ability, Processing of Information and Learning (APIL) Battery and Raven’s Advanced Progressive Matrices (APM), used in selecting university bridging-course students. The empirical study is informed by a literature review focussing on the legal and scientific parameters of psychometrics within selection, the conceptualisation of intelligence and its relationship to academic performance. Hypotheses are posed regarding the predictive power of the selection battery and the effect of biographical factors on academic performance.

Results indicate that the APIL Battery and Raven’s APM are both valid predictors of academic performance, although the former appears more effective. This investigation emphasises the influence of moderating factors, i.e. factors other than cognitive ability, on academic performance.

Key terms:

Validation; psychometric testing, predictive validity, pilot profile, labour legislation; job analysis; job description; scientific selection; APIL Battery; Raven’s APM
CHAPTER I  BACKGROUND TO AND MOTIVATION FOR THE INVESTIGATION

1.1 INTRODUCTION

Legal as well as scientific and ethical arguments point to the need to first determine the validity of a specific psychometric battery before any further decision – making can proceed. Therefore, any person or institution tasked with ensuring the fairness as well as the legal and scientific defensibility of selection procedures, cannot afford to neglect this subject (Cascio, 1998). The purpose of this investigation is the validation of a selection battery containing the Ability, Processing of Information and Learning (APIL) Battery and Raven’s Advanced Progressive Matrices (APM).

On a macro - level, the world is rapidly moving away from uni - culturalism and has started to embrace diversity in all spheres of life. In South Africa, the ushering in of democracy has created a number of changes in the community at large which also impacted on the way in which organisations are managed and business is conducted. South Africa was relieved of its pariah status and once again became part of a broader community of nations. Thus, it is also increasingly being affected by current globalisation trends which escalatorily impacts on the use and type of technology, culture and economic survival. The logical extension of the above, is that these external factors will and do impact on organisations as microcosms of its external reality (Boon, 1998; Foxcroft, 1997; Gibson, Ivancevich & Donnelly, 1994).

Changes in South Africa have also highlighted enormous anomalies between the various racial components of society. Approximately a quarter of the South African population is Westernised, educated and relatively comfortable economically, whereas the overwhelming majority of South Africans are still relatively traditionally orientated and economically disadvantaged. Organisations are invariably affected by this apparent contradiction and attempts are made to balance the complexities inherent in this disjointed state of affairs (Foxcroft, 1997; Mjoli, 1987). Organisational attempts at redressing these as well as other societal imbalances are often
contained in their social responsibility programmes (Beaty, 2000; Van Rooyen, 2001). The beneficiaries of such a social responsibility programme, specifically a university bridging – course for learners from disadvantaged communities, serve as the population on which this research is based.

A university bridging – course is a structured educational endeavour that normally extends over a year, in which students are given practical training before entering tertiary training. It contains both developmental as well as academic components, aimed at overcoming various educational and socio – economic deficiencies (Hunter, 1989). In terms of the effectiveness of university bridging – courses in meeting these requirements, the jury is still out. Troskie – De Bruin (1999) found that the long – term effects of a bridging programme at the University of Stellenbosch, could not be established with certainty. Furthermore, Van Rooyen and Huysamen (2000) found that a bridging programme under investigation apparently did not contribute to tertiary academic performance. In contrast, Kieswetter (1996) found that a preparatory programme for engineering studies at the University of the Western Cape was reasonably successful. The success of this particular programme is ascribed to the quality of the learning environment, social support, as well as the fact that it focussed on both cognitive and affective development. Curtis and De Villiers (1992) also found that a university bridging – course contributed positively to academic achievement.

A validation study is conducted on the selection battery used for the selection of these university bridging – course learners. This particular university bridging – course is Government sponsored, but a private institution provides the actual training. Learners from previously disadvantaged backgrounds are selected for training based on the marks obtained during their final matriculation examination (Kentron News, 2001). Due to the fact that this selection battery has not been validated yet, performance in the psychometric tests was not taken into account during the selection process.

Acting as impetus for this project, is the assumption that many promising black learners are denied entry into university based on poor matriculation results. Considering the history of South
African education, poor matriculation performance quite often occur due to systemic inefficiencies such as a lack of qualified teachers and educational facilities rather than an inherent incapacity to perform (Foxcroft, 1997; Van Rooyen, 2001). This bridging programme is therefore aimed at addressing this state of affairs by providing selected learners with an opportunity to improve their matriculation marks whilst undergoing teaching in a more conducive learning environment. A further motivating factor is to affect social justice by redressing the imbalances that have systematically been created within the South African education system. The main focus area is the facilitation of improved performance in Mathematics, Science and where applicable Biology at matriculation level. Participants are then expected to complete a supplementary matriculation examination in these subjects in order to obtain university entrance. The organisation then assists them in entering South African universities for further education, based on their performance during the final year supplementary matriculation examination. As the aims of this programme are predominantly altruistic, participants are not contractually obligated to the Government upon completion of this programme (Curtis & De Villiers, 1992; Kentron News, 2001; Webb & Erwee, 1990).

1.2 LEGAL AND SCIENTIFIC IMPERATIVES IMPACTING ON PSYCHOMETRIC ASSESSMENT

A number of legal sources place limitations upon the manner in which psychometric assessment is conducted in South Africa. These are, *inter alia*, the Constitution of the Republic of South Africa (Act No 108 of 1996), The Health Professions Act, No 56 (1974), the Labour Relations Act, No 66 (1995) and the Employment Equity Act, No 56 (1998). The Employment Equity Act (1998, p. 16) states that: “psychological testing and other similar assessments of any employee are prohibited unless the test or assessment being used has been scientifically shown to be valid and reliable; can be applied fairly to all employees; and is not biased against any employee or group”. Good scientific practice would require that similar considerations prevail when conducting psychometric assessment. Therefore, both science and the law emphasise the importance of reliability and validity in terms of psychometric assessment. Chapter 2 entails a discussion of the meaning of these two concepts within the psychometric assessment context.
1.3 PROBLEM STATEMENT

The general research problem can be stated as follows: Does the existing psychometric selection battery predict academic performance of a group of university bridging - course students? This research problem is relevant from both a legal and scientific perspective. As was stated previously, legal parameters demand that psychometric tests be valid and reliable within a selection context. Good scientific practice also has the same requirements in order to ensure that sound selection decisions are made (Cascio, 1998). Considering the fact that a substantial financial investment is made in order to train participants, it is imperative that legal and scientific requirements are adhered to.

1.4 RESEARCH QUESTIONS

Based on the stated research problem, the following research questions have been formulated. A general research question is first formulated, followed by specific research questions pertaining to the literature review and the empirical investigation.

1.4.1 General research question

The general research question may be stated as follows: To what extent does a selection battery predict academic performance in a university bridging - course?

1.4.2 Specific research questions for the literature review

The following specific research questions pertaining to the literature review are addressed within the two literature chapters (chapters 2 and 3).

1. What is the role of psychometric assessment within the broader selection context?
2. How is the construct “intelligence” conceptualised within the context of a university bridging course?

3. What is the relationship between intelligence and academic achievement within the context of a university bridging course?

4. What are the factors moderating the relationship between intelligence and academic achievement within the context of a university bridging course?

1.4.3 Specific research questions for the empirical investigation

The following specific research questions pertain to the empirical investigation and correspond to the number of research hypotheses stated in chapter 4.

1. What is the relationship between scores obtained in the APIL Battery and the variance observed in academic performance in a final year supplementary matriculation examination?

2. What is the relationship between scores obtained in Raven’s APM and the variance observed in academic performance in a final year supplementary matriculation examination?

3. What is the relationship between scores obtained in the total selection battery and the variance observed in academic performance in a final year supplementary matriculation examination?

4. To what extent do significant differences between male and female participants exist regarding academic performance in a final year supplementary matriculation examination?

5. To what extent do significant differences between different ethnic groups exist regarding academic performance in a final year supplementary matriculation examination?
6. To what extent do significant differences between participants from different language groups exist regarding academic performance in a final year supplementary matriculation examination?

1.5 RESEARCH AIMS

Emanating from the above-mentioned problem statement, the following research aims are formulated. These are subdivided into the general aim, theoretical aims and research aims. The research aims logically inform the formulation of research hypotheses as well as the statistical analysis performed on the data. These and other methodological aspects are discussed in chapter 4.

1.5.1 The general aim

The general aim of this study is to determine the predictive validity of a psychometric test battery used in the selection of university bridging - course learners.

1.5.2 Theoretical aims

A literature review is conducted in order to fulfill the theoretical aims of this investigation. These aims correspond with the stated specific research questions for the literature review.

1. To investigate the role of psychometric assessment within the broader selection context.

2. To investigate how the construct “intelligence” is conceptualised within the context of a university bridging - course.

3. To investigate the relationship between intelligence and academic achievement within the context of a university bridging - course.
4. To investigate factors that moderate the relationship between intelligence and academic achievement within the context of a university bridging – course.

1.5.3 Empirical aims

The following empirical aims are stated in order to investigate the empirical component of this study. In the following statements, *performance* denotes the average academic performance achieved by individual participants. This score is obtained by calculating the average mark obtained for Mathematics, Science and Biology in a final year matriculation supplementary examination. The relationship between the various psychometric tests and the various subjects respectively are also explored. This is however, not included in either the empirical aims nor the stated hypotheses (chapter 4) as such exploration only serves to highlight the predictive impact of various psychometric tests on different school subjects. Also, the exploration of individual school subjects would result in unnecessary “clumsiness” regarding the statement of hypotheses. The research hypotheses stated in chapter 4 are therefore solely based on the following six research aims. These are:

1. To determine the relationship between academic performance in a final year supplementary matriculation examination and the scores obtained in the APIL Battery.

2. To determine the relationship between academic performance in a final year supplementary matriculation examination and the scores obtained in Raven’s APM.

3. To determine the extent to which the total selection battery declares the variance observed in academic performance in a final year supplementary matriculation examination.

4. To determine the extent to which differences exist between males and females regarding academic performance in a final year supplementary matriculation examination.

5. To determine the extent to which differences between different ethnic groups exist regarding
6. To determine the extent to which differences between different language groups exist regarding academic performance in a final year supplementary matriculation examination.

1.6 PARADIGM PERSPECTIVE

Babbie (1992) as well as Mouton and Marais (1994) state that no research project could ever proceed from a clean slate. The existence of a specific research or disciplinary paradigm mitigates against an assumed *tabula rasa* approach to research. A paradigm is defined as "a fundamental model or scheme that organizes our view of something" (Babbie, 1992, p. 56). A paradigm can be likened to a yardstick against which a research problem is measured in order to determine its feasibility, researchability as well as its importance to the scientific community. All research proceeds from a given scientific paradigm that presupposes certain laws of and assumptions about behaviour. It also informs the research techniques that will be followed by the researcher (Mouton & Marais, 1994). Therefore, although it may not necessarily answer important questions, it guides the researcher into the direction where the solutions may eventually be found (Babbie, 1992).

Regarding the psychological paradigm considered, the trait – factor perspective, appears to be the most appropriate. The underlying assumption of this perspective is that human behaviour may be ordered and defined along dimensions of specific traits. Guilford (as cited in Gregory, 2000, p. 490), defines a trait as any "relatively enduring way in which one individual differs from another". Traits are viewed as the primary determinants of behaviour and represents consistent tendencies to behave similarly regardless of the environmental context. Although the impact of environmental factors are recognised, trait theorists suggest "that traits are inner dispositions that tend to be associated with relatively consistent behavior across a variety of situations" (Walsh & Betz, 2001, p. 86). The adoption of a trait perspective is extremely popular in everyday human interactions as well as in the formal field of Psychology (Gregory, 2000). Many psychologists have contributed to this perspective, but only Cattell’s position will be highlighted here.
The work of Cattell is most pertinent here considering the fact that the two psychometric tests that serve as independent variables in this study, are both direct derivatives of this theoretical stance. Cattell distinguished between surface and source traits. Source traits are those that are less visible than surface traits, but are more important in explaining human behaviour. Cattell's factor analytic trait theory incorporated refined factor - analytical methods to uncover the basic traits underlying personality. Eventually, 16 source traits emerged which were subsequently incorporated into the well-known Sixteen Personality Factor Questionnaire (Gregory, 2000; Weiten, 1994). Traits may be juxtaposed to, for instance, unconscious thought processes and/or experience. Trait theorists tend to prefer a positivistic position and quantitative research methodology in the measurement of these constructs (Meyer, Moore & Viljoen, 1997). In terms of the paradigm perspective pertaining to the empirical aspects of this research, positivism would be the most appropriate paradigm to follow (Mouton & Marais, 1994). A positivistic stance implies that only "knowable” matters may be subjected to scientific inquiry. This means that only that which can be observed with the senses should be the focus of scientific investigation. That which cannot be observed, are considered to fall outside the scope of scientific study (Meyer et al, 1997).

The contribution of these theorists, are especially evident in the development of psychometric theory and assessment (Walsh & Betz, 2001). Most psychological constructs are not directly measurable. In fact, a construct is defined as “theoretical concepts formulated to serve as causal or descriptive explanations ... a kind of 'theoretical scaffolding' between variables” (Rosnow & Rosenthal, 1996, pp. 43 - 44). Constructs are therefore “theoretical creations based on observations but which cannot be observed directly or indirectly” (Babbie, 1992, p. 109).

*Intelligence* is an abstract concept and not an entity in its own right. It is therefore a construct and as such cannot be directly observed. Babbie (1992, p. 109) states that intelligence “is constructed mathematically from observations of the answers given to a large number of questions on an IQ test”. The psychometric tests under investigation are therefore in fact "translations” of this construct into observable and measurable behaviour (Babbie, 1992; Huysamen, 1998; Smit,
1996). The fact that the APIL Battery and Raven's APM are purported to measure "intelligent" behaviour accurately and objectively, is in line with the central tenets of the trait perspective (Raven & Court, 1998; Rosnow & Rosenthal, 1996; Smit 1996; Taylor, 1997). These and other attributes of psychometric assessment and the measurement of intelligence are discussed at length in chapters 2 and 3.

1.7 RESEARCH DESIGN

This study incorporates a literature review and a quantitative empirical investigation into the predictive validity of this test battery. The unit of analysis is the individual participant in the university bridging course. The sample size is 167, consisting of 127 males and 40 females. The sample used is equivalent to all the individuals who have applied to be admitted to the programme. Furthermore, the entire population consists of Black individuals of various ethnic origin and home languages. According to Babbie (1992), the psychometric instrument/s used in a validation study must be administered to a relatively large sample (± 100 depending on the number of tests or instruments included in the battery) which is representative of the population in which the instruments are going to be used. This sample is therefore deemed adequate for use in a validation study (Babbie, 1992; Huysamen, 1998).

In keeping with the goal of research, i.e. to determine whether an independent variable impacts on a dependent variable, two psychometric tests are administered to participants to determine whether they have an effect on academic performance (Huysamen, 1998). A variable may be defined as a "logical set of attributes" (Babbie, 1992, p. 27). Rosnow and Rosenthal (1996, p. 44) define a variable as "an event or condition that the researcher observes or measures or plans to investigate and that is liable to variation". A dependent variable is the "effect" or outcome of an investigation. These are variables that are assumed to depend on or be caused by another (the independent variable) (Babbie, 1992; Rosnow & Rosenthal, 1996).

The APIL Battery as well as Raven's APM were administered to participants one month before the start of the education programme thereby acting as independent variables. As will be
discussed in chapters 3 and 4, both these tests are considered to be cognitively loaded measures of intelligence. Their validity and reliability in different contexts have also been established (Bors & Stokes, 1998; Taylor, 1997). In order to determine the predictive validity of a psychometric test, a valid criterion (e.g. measurement of academic performance) must be identified and be available in numeric format (Babbie, 1992; Huysamen, 1994). Huysamen (1998) cautions that it is imperative that criterion measures are valid measures of the criteria themselves. In this investigation, participants' average academic performance in a final year supplementary examination marks are used as the criterion measure or dependent variable. Marks obtained in the final year examination are considered by the researcher as the most reliable, valid and objective indicator of academic performance because it is a national examination that is externally compiled and marked. Furthermore, most South African Universities consider matriculation marks as the most important entry requirement (Curtis & De Villiers, 1992; Webb & Erwee, 1990). These aspects are discussed at length in chapter 4.

1.8 RESEARCH METHODOLOGY

Following, are the research methodological steps that are followed in this investigation.

Phase One Literature Review

In order to fulfil the above-mentioned theoretical aims, a literature review is conducted. The most recent and available sources pertaining to selection and the measurement of intelligence are consulted and integrated into the literature review. Also, previous research relating to this research topic is integrated into the literature review. The following steps are followed in the literature review:

Step1: Chapter 2 entails a discussion of the various legal as well as scientific parameters within which psychometric assessment in selection proceeds. Specific emphasis is placed on the Labour Relations Act and the Employment Equity Act. Also the characteristics and role of psychometric assessment within selection are discussed at length.
Step 2: In chapter 3, the nature and assessment of the construct intelligence, specifically *fluid intelligence*, are discussed and integrated into the research project. The latter construct is particularly important, as it is an underlying construct in terms of both the psychometric instruments under investigation.

**Phase Two  Empirical Study**

The following steps are followed in the empirical investigation.

**Step 1:** The sample and population are described in terms of gender, home language and ethnicity.

**Step 2:** The measuring instruments, i.e. the APIL Battery and Raven's APM, are described in terms of their aims, administration and psychometric properties.

**Step 3:** The criterion, academic performance at the end of the university bridging course, is discussed.

**Step 4:** Data Collection. The aforementioned psychometric instruments were administered to participants before the start of the programme. Subsequently, data regarding their academic performance in various subjects in the final year supplementary examination was collected upon completion of the programme.

**Step 5:** Formulation of the empirical hypotheses covering the objectives of the research.

**Step 6:** The SPSS computer package was used to perform the data analysis. For the purposes of this research project, determining the predictive validity of the above-mentioned selection battery is crucial. Therefore, data was statistically manipulated by means of, amongst other techniques, correlation and regression analysis as well as discriminant analysis. Regression
analysis was performed to determine the extent to which the psychometric battery under investigation declared the observed variance in academic performance. Combined, these procedures are useful in the determination of the effectiveness of the psychometric battery, in the prediction of success in academic performance. Further statistical analysis was performed to determine the impact of biographical variables on academic performance.

Step 7: Reporting on and interpretation of research results.

Step 8: Conclusions are drawn based on the research results and the interpretation thereof.

Step 9: Research limitations are discussed.

Step 10: Lastly, recommendations are suggested based on the research results and its interpretation. This may be used to inform the future recruiting and selection procedures that are followed by the organisation.

1.9 SUMMARY

In this chapter the background to and motivation for this study are discussed. The socio-economic necessity for the establishment of this university bridging-course was highlighted. Furthermore, the role of psychometric assessment and various legal requirements within the selection context were introduced. A more extensive discussion on these issues is reserved for chapter 2. Also general and specific research questions are stated. Emanating from these research questions are the research aims. The general research aim acted as the basis for the formulation of specific theoretical and empirical aims.

The trait perspective is discussed as a psychological paradigm of this study. It is mentioned that this particular paradigm limits the empirical paradigm to positivism, which grounds the empirical component of this investigation. The limiting effects of the paradigmatic stance assumed are further highlighted in chapter 5.
This chapter also serves as an introduction of the research design and methodology that will be adhered to in this investigation (see chapter 4). Finally, the steps followed in the empirical component of this investigation are presented.

1.10 CHAPTER DIVISION

The following chapter division will be adhered to:

Chapter 1: The background to and motivation for the study are discussed. This chapter entails a discussion on the nature of and need for university bridging – courses. Also, the problem statement, research questions and aims, the paradigm perspective adhered to as well as the research design and methodology are mentioned.

Chapter 2: The role of psychometric assessment in selection is discussed. It entails a discussion on the legal and scientific parameters within which psychometric assessment for selection purposes proceeds.

Chapter 3: The nature and assessment of intelligence are discussed within the context of various conceptual frameworks. Also, the relationship between intelligence and academic achievement, as the crux of this research effort, is discussed. Finally, factors that may moderate this relationship are highlighted.

Chapter 4: The chapter on the research methodology adhered to entails a description of the population and sample, the measuring instruments and criterion as well as a discussion of the data collection and analysis procedures that are followed.

Chapter 5: The main focus here is on the research results and a discussion thereof. Based on this, general as well as specific conclusions are drawn in terms of the literature review and empirical study. Finally theoretical and empirical limitations are mentioned and recommendations for
further research are generated.
CHAPTER 2 THE ROLE OF PSYCHOMETRIC ASSESSMENT IN SELECTION

2.1 INTRODUCTION

South African organisations are faced with the need to ensure diversity and representivity while simultaneously ensuring that capability is maintained and even improved upon. The need to select individuals with the "right" combination of skills and characteristics needs to be balanced with those of practices like affirmative action. As organisations are first and foremost concerned with their economic survival, these issues do indeed become rather contentious. The development of political and economic disparities between the various race and gender groups in South Africa, are well - documented (Curtis & De Villiers, 1992; Foxcroft, 1997; Mjoli, 1997; Taylor, 1997). Organisations are caught up in the apparent contradiction that the skills required and the "type" of people needed to remove this socio - political and economic disparities do not necessarily match. It therefore seems that there is a great need to balance the complexities inherent in trying to "right" past wrongs, on the one hand, with the need to remain competitively viable on the other (Mjoli, 1997; Nainaar, 1999; Taylor, 1994).

What concerns the Industrial/Organisational Psychologist is how these socio - political imperatives can be merged with the recruitment and selection of competent individuals. The question is whether these two considerations are necessarily incompatible, as it would seem at first glance. Within the South African context, psychometric assessment should not only focus on ensuring competent and efficient human resources, but the methods and techniques used should also be legally and ethically defensible (Mjoli, 1997; Nainaar, 1999; Taylor, 1994).

2.2 SCIENTIFIC SELECTION

In this section the nature of scientific selection will receive attention, specifically, the scientific as well as the legal dictates within the South African context, will be explored.
2.2.1 The nature of scientific selection

Psychometric assessment cannot be isolated from the greater scientific selection process. In the simplest terms, assessment is about the generation and collection of evidence. All forms of assessment can be included in this description – including the interview. In a scientific selection process, the purpose is to collect sufficient evidence or information that individuals can perform to the specified standards in a specific role or post (Cascio, 1998). The crux of scientific selection is found in its use of scientific methodology in order to collect the above mentioned evidence. In essence then, scientific selection is based on the systematic process of collecting and analysing information regarding candidates. This would entail that specific steps and/or “rules” are followed before making a selection decision. These steps and “rules” are discussed in later sections (Babbie, 1992; Huysamen, 1994; Leedy, 1997; Smit, 1996).

2.2.2 Steps in compiling a scientific selection battery

According to Smit (1996), one of the most important preconditions for the growth of any science is the existence of measuring instruments in order to gather reliable, valid and accurate information. Within the field of psychometrics for selection purposes, data is continually collected and used to make inferences and evaluations regarding the suitability of a candidate for utilisation. In addition to this information being reliable, valid and accurate, it should also be directly relevant to the requirements of the job in question.

The conventional selection process, consistent with the paradigm perspective chosen at the outset, may be viewed as systemic in nature. The systemic view of the selection process may be described as “a network of sequential, interdependent decisions” (Cascio, 1998, p. 34). It is cyclical, starting at a specific point, moving through distinct phases and ultimately ending at the re-evaluation and changing if necessary, of the whole process. It may furthermore be described as sequential to the extent that information gathered at one phase directly feeds into and determines subsequent phases (Cascio, 1998; Smit, 1996). The compilation of a test battery, is an important subsystem within the selection process.
Cascio (1998) emphasises that ideally, different selection strategies should be used for different jobs and secondly, that the phases in this process are highly interdependent. Typically, the personnel selection cycle contains the following steps:

**Phase 1:** Firstly, the job in question should be analysed. Cascio (1998) and Smit (1996), agree that job analysis is the foundation of the entire selection process as it is crucial in the recruitment, utilisation and finally performance appraisal of personnel. It is critical to first determine the what of a particular job before any decision regarding who can be made. Furthermore, the utilisation and evaluation of an employee should rest squarely on how her/his job has been defined at the outset. Therefore, any person tasked with ensuring human resource capability can ill afford to neglect this subject (Robbins, 1998).

According to Cascio (1998), job analysis aims at defining each job in terms of the behaviours necessary to achieve it; it can be subdivided into two sub processes; the determination of job descriptions and job specifications. Job descriptions is described as a written statement about what the employee does, as well as how and why it is done (Robbins, 1998).

Cascio (1998) states that firstly the purpose of the analysis should be clearly stated. Furthermore, the job should be systematically analysed as well as its specific context. Also, care should be taken that the individual job incumbent provides a true view of the job and not just her/his interpretation of how it should be done. Therefore, as many people as possible in the same job should be utilised to limit the individual’s subjectivity. A further requirement is that the process should act as the basis for all major personnel decisions. Furthermore, the identified skills, knowledge and abilities should be operationally defined or anchored in behaviour (Clifford, 1994).

**Phase 2:** The above information is then converted into job specifications; this means that the critical elements of the job are translated into person characteristics and attributes. Job specifications entail the determination of “those knowledge, skills, abilities, and other
personal characteristics deemed necessary to perform a job” (Cascio, 1998, p. 130). It entails the minimally acceptable standards for selection and later evaluation purposes. Job specifications are valid only to the extent that those possessing the identified traits do indeed perform better at the job than those who do not possess those traits (Cascio, 1998; Clifford, 1994; Robbins, 1998).

**Phase 3:** The design of a selection strategy is a critical phase in matching the individual with the needs of the organisation. Contained within this strategy, would be an attempt to match potential candidates with the requirements of the job. This decision making process is guided by, amongst others, utility and financial considerations. Cascio (1998) states that a selection strategy may be done either mechanically (eg. psychometric tests and biographical questionnaires), judgmentally (eg. interviews) or by using a mixture of both.

For the purposes of this research project, the compilation of a selection test battery is pivotal. In this regard, Nainaar (1999, p. 2) stresses the “need to establish methods and instruments that are valid and reliable for organisational use, with emphasis on cross – cultural and predictive validity”. A selection battery may be either designed or selected from pre – existing batteries. As this aspect serves as the central focus of this investigation, it will be dealt with in more detail in the discussion of the empirical study, i.e. chapter 4.

**Phase 4:** Scores in terms of various measuring instruments are obtained and integrated into an overall “picture” of the candidate. This would then be used to determine the candidates’s suitability for the job in question. Whereas the previous phases entailed measurement, phase 4 entails prediction, i.e. combining scores in a manner that would reduce predictive error (Cascio, 1998).

**Phase 5:** A critical last step should be the evaluation of the success or predictive validity of the selection process measured against actual performance on the job. Due to the impact of the selection process on organisational effectiveness as well as the wellbeing of the individual
employee, this phase may not be overlooked (Cascio, 1998). As the process of validation is the crux of this investigation, more attention will be afforded to it later in this chapter.

2.2.3 The utility of scientific selection

It should be clear from the above discussion that scientific selection is a process consisting of a number of decisions. Approaching it from a scientific perspective, serves to enhance the quality of decisions being made. Therefore, the utility of scientific selection and the predictions that emanate from it lies in its outcomes. The practice of psychology within an organisational context, necessitates that its outcomes should enhance organisational efficiency. In the final analysis, considering the substantial costs involved in recruitment and selection, it should be considered to be justifiable and contributing to the organisational bottom line.

Numerous strategies for determining the utility of selection practices are available. The I/O Psychologist may choose, amongst others, between the Taylor - Russel; Naylor - Shine and Brogden - Cronbach - Gieser models for determining the utility of a particular selection strategy (Babbie, 1992; Cascio, 1998; Kerlinger, 1986). In the interest of brevity and considering the aims of this investigation, these models are not discussed fully.

According to Cascio (1998) and Taylor (1994) some of the potential benefits that may emanate from a process of scientific selection are:

* It minimises the erroneous acceptance of low performing candidates.

* It enhances the human resource management process, eg. succession planning, placement and performance appraisals.

* More specifically, organisations could potentially ensure that they get a "return on investment" in terms of training if the "right" people are recruited, placed and trained.
As a result, it may lead to the appropriate utilisation of personnel and greater organisational efficiency.

2.3 PREREQUISITES FOR PSYCHOMETRIC ASSESSMENT

I/O Psychologists are continually faced with the challenge of providing professional and scientific psychometric services to clients that will meet with the approval of all stakeholders (organisations, unions and individual employees). As was mentioned earlier, psychometric tests have for a long time been scrutinised due to perceptions that they are subjective and biased. Labour organisations, especially view psychometrics as deliberate stumbling blocks to black advancement (Nainaar, 1999; Potgieter & Van der Merwe, 2001). Traditional, norm-referenced psychometric tests do indeed have certain limitations. Criticism is leveled at the fact that they are mainly clinical in nature and do not take the specific job context into consideration. Furthermore, psychometric tests cannot always be directly linked to job performance (Potgieter & Van der Merwe, 2001; Smit, 1996).

Despite these and other criticisms, psychometric instruments are still considered valuable to the extent that they can determine individual potential, can be used for large group testing and are time efficient. Therefore, psychometrics may enhance selection by instilling some degree of objectivity into the process (Potgieter & Van der Merwe, 2001; Smit, 1996).

2.3.1 Defining psychometric assessment

Differential Psychology provides the basic assumptions underlying psychometric assessment, i.e. that individuals differ both psychologically as well as physically. Unlike physical measurement that is relatively simple and uncomplicated, psychological measurement, is more complex. Psychological variables cannot be directly measured, rather deductions and abstractions regarding the nature and extent of certain psychological variables are made through the process of measurement (Cascio, 1998; Kerlinger, 1986; Smit, 1996).
The first step in measuring psychological differences would be to adequately define the variable/s in question. In order for such measurement to be considered useful, the variable under consideration should be relatively stable over time and situations. Furthermore they also need to be suitable for accurate assessment (Smit, 1996). For example, intelligence, the central construct in this investigation, is such a variable and may therefore be subjected to measurement.

Simply defined, measurement is the process of assigning numbers to objects or events according to certain rules (Babbie, 1992; Cascio, 1998; Smit, 1996). It simply answers the question, “How much?” This definition does however not bring the quality of measurement into consideration. Kerlinger (1986, p. 392) expresses these ideas in a very simple yet elegant manner:

“Measurement is a game we play with objects and numerals. Games have rules. It is, of course, important for other reasons that the rules be “good” rules, but whether the rules are “good” or “bad”, the procedure is still measurement.”

The purpose of the game is to find similarity between the tests scores and reality. The greater the similarity, the more effective the measuring procedure.

The Society for Industrial Psychology (1998, p. 13), defines an assessment instrument as “a measure based on a sample of behaviour or attributes used for making decisions about people in the workplace”. Smit (1996, p. 24), defines a psychological test as “an objective and standardised measuring instrument for a limited defined area of human behaviour”. Brown (as cited in Smit, 1996, p. 24), states that it is “a systematic procedure for measuring a sample of behavior”. Regardless of the specific purpose, all psychological tests need to adhere to certain basic principles and requirements. Following, is a discussion on the most important requirements or the “good” rules of psychological measurement.
2.3.2 Reliability

In its simplest form, reliability means that if a psychological test is repeatedly administered to the same person, it would yield the same results (Babbie, 1992). According to Huysamen (1983, p. 24), the reliability of a test refers to “the consistency of its scores over different administrations involving different occasions, test forms, scorers, etc.” However, there is no point in obtaining reliable test results if the test is not valid. In other words reliability does not ensure accuracy.

Any particular testing produces an observed score, which consists of the true score and an error score. An individual’s true score is his or her mean score on an infinitely large number, or universe, of independent administrations of the test. Since no test is perfectly reliable, an individual’s scores will not always be equal to his or her true score. The quantity by which an individual’s observed score on a particular administration of the test deviates from his or her true score is known as his or her error score for that administration (Huysamen, 1988).

The higher the reliability, the better observed scores correlate with true scores. The more the scores fluctuate from one test administration to the next as a result of unsystematic sources of variation or sources of error variation, the lower is the reliability of the test. Further, a low reliability coefficient does not only undermine the usefulness of the measure, but also tends to lower the correlation of that test score with any other variables. Different kinds of reliability are associated with different sources of error variation. Reliability, which refers to test-score consistency over different occasions, is referred to as test – retest reliability. The consistency in test scores over different parallel test forms is known as parallel – forms reliability. The agreement in scores assigned by different scorers is called inter-scorer or observer reliability (Babbie, 1992).

2.3.2.1 Procedures for the determination of reliability

- Test-retest reliability or temporal stability
Temporal stability is determined by administering the same test to the same group of respondents at different times. The correlation between the two sets of data is a reliability coefficient known as the coefficient of stability (rtt) (Smit, 1996). An important assumption underlying this method is that the behavior being measured is stable over time.

♦ Parallel/equivalent form reliability

Parallel form reliability is determined by administering equivalent forms of the same test to the same group of respondents at different times. When the time interval between the application of the two forms is short, the calculated reliability coefficient is known as the coefficient of equivalence. If the time interval between the application of the two forms is long, it is known as the coefficient of stability and equivalence (Babbie, 1992; Huysamen, 1994).

♦ Observer reliability

Observer reliability is determined by calculating the relationship between the evaluations of the same test by different observers. This method does not require two applications of a test. Different observers evaluate the same set of answers, after which the researcher calculates inter-correlations between the evaluations (Babbie, 1992).

♦ Content reliability/item homogeneity

Content reliability gives an indication of the homogeneity of a test. Homogeneity refers to the consistency estimates that are primarily concerned with the internal structure of the test. Although items in a test may differ, all items on the test should be centered on the same general content or factor being measured by the test (Smit, 1996).
The homogeneity of a test is normally determined by analysing the internal consistency of performance on the test items. The homogeneity of a test is optimal when inter-item correlations are high, item variance is high and the difficulty value of the items are of the same magnitude. Item homogeneity can be determined through split-half methods and internal-consistency formulas (Babbie, 1992; Huysamen, 1994).

* **Split-half methods.** When applying a "split-half" method, the test is administered just once. Instead of working from item scores, the researcher splits the test into two roughly equivalent parts. The shorter a test, the lower its reliability. It is thus necessary to correct this reliability index, and this is usually done by applying either the method of Spearman-Brown or the method of Guttman (Cronbach, 1990; Smit, 1996).

* **Internal consistency methods.** The two most often used formulas are the Kuder-Richardson formula 20 (KR-20) and Cronbach's alpha. Cronbach's alpha describes how much each item in the test is correlated with all others - i.e. the overall consistency of the test or the extent to which high responses go with highs and lows with lows over all the test items. The KR-20 is a specialised version of Cronbach's alpha and applies when all the test items are dichotomous (such as yes-no or true-false). When coding the results of dichotomous data, every right answer usually counts one point and a wrong answer counts zero. However, the result with KR-20 is the same as with Cronbach's alpha (Aron & Aron, 1994).

**2.3.2.2 Practical implications**

Generally, in psychology, a test should have a reliability of at least 0.7, but preferably closer to 0.9 to be considered useful (Aron & Aron, 1994). When determining the test-retest reliability or temporal stability of a test, it is important that a few days should pass before administering the post test. It is also important that the researcher has knowledge regarding the aspect being measured when he/she interprets rtt. If this were not the case, the researcher would not know if rtt is low because of error factors, or because the behavior being measured
is not stable over time. When determining parallel/equivalent form reliability, Smit (1996) is of the opinion that equivalent forms of tests should constitute the following:

* Consist of the same number of items.

* Items of the two forms should be uniform concerning item difficulty, content and representativeness.

* The item homogeneity of the two forms should be the same.

* The mean scores of the two tests should be the same.

* Scoring procedures should be the same.

2.3.3 Validity

The validity of an instrument gives an indication of whether a psychometric instrument measures what it is supposed to measure (Babbie, 1992). Validity can be determined in a variety of ways. Each type of validity has a different meaning and use. The following provides a description of the definition, purpose, procedure and practical implications of each type of validity.

2.3.3.1 Content validity

Babbie (1992, p. 133) defines content validity as “the degree to which a measure covers the range of meanings included within a concept”. It can therefore be described as a method to determine whether test or scale items, represent the behavioural aspects they are supposed to measure.
The evaluation of the content validity of a psychometric instrument thus implies the logical investigation of the test content as well as the methods used in constructing the instrument. It is more appropriate to investigate the content validity of an instrument before construction is completed.

It is advised that any new instrument that is developed within an organisation, be revised in conjunction with the developer, in order to ensure that all tests have the necessary content validity. If the developer wants to determine the content validity of his/her instrument himself/herself, the process must be recorded and reviewed by the organisation using the test in conjunction with the developer.

According to Babbie (1992) and Smit (1996) the following steps are necessary in determining the content validity of a test:

* The relevant universe of items must be defined in terms of tasks and situations with which the subject (testee) may be confronted.

* The total universe of items must be systematically divided into subdivisions.

* A probable sample of tasks or situations for each category must be assembled.

* The selected tasks or situations must be written as questions.

2.3.3.2 Construct validity

In psychology, research is done on the functional relationship between relevant variables in the field. A construct is an imperceptible, hypothetical variable which forms part of a theory, developed to explain observable behavior. Almost all psychological concepts, for example, intelligence, interest, attitude and performance motivation are hypothetical constructs. These constructs must be measured or quantified before any assumption (hypothesis) concerning
relationships between these constructs can be tested. "Construct validity is based on the way a measure relates to other variables within a system of theoretical relationships" (Babbie, 1992, p. 133). In simple terms, construct validity can be defined as the extent to which a test measures the theoretical construct it is supposed to measure.

Construct validity is important when a test is developed (or an existing test is evaluated) to measure certain attributes or constructs which varies between individuals. Research would be aimed at determining whether the instrument measures the constructs it claims to measure, when tested on a specific sample. The same research could also be aimed at determining whether the test functions effectively amongst different cultural groups.

Construct validity cannot be determined by means of one single numerical index. A wide variety of methods are used to determine construct validity. These methods can be divided into two categories, namely intra test methods and inter test methods. A brief description of each is given.

**Intra test methods**

These methods are aimed at the investigation of the internal structure of the test (Smit, 1996). In other words, the researcher takes a look at the expected pattern of responses, the internal structure of the instrument as well as the relationship between items or sub-scales of the instrument. These methods provide information regarding the area of behaviour measured by the instrument and is usually obtained by means of factor analysis. This method, however, does not provide any information regarding the relationship between this construct and other variables. It is also possible to investigate whether an instrument measures the same construct in different groups, for instance, different cultural groups.
Inter test methods

These methods imply the evaluation of the intercorrelations of several tests simultaneously. They are aimed at the identification of commonalities and determining whether tests measure the same construct (Smit, 1996). These tests have to be administered at the same time as the newly developed instrument. Two inter test methods can be distinguished:

* Method of congruent/convergent validity

The newly developed test is correlated with the existing test. High correlations give an indication that the two instruments measure the same construct (Babbie, 1992).

* Method of discriminant validity

According to this viewpoint, a test is therefore not only invalid when it does not correlate well with a test that measures the same construct, but is also invalid when it correlates too high with a measure with which it is supposed to differ (Babbie, 1992). For example, if a certain ability is supposed to differ between groups, t-tests or one-way analysis of variance can be used to confirm the construct validity of a test.

2.3.3.3 Criterion-related validity

Criterion-related validity stands in relation to the use of measuring instruments that are used to make practical decisions concerning aspects like the selection of applicants for positions etc. “Criterion-related validity is studied by comparing test or scale scores with one or more external variables, or criteria, known or believed to measure the attribute under study” (Kerlinger, 1986, p. 418). This type of validity can be divided into two types namely concurrent validity and predictive validity. Both types are based on the same principle, namely the comparison of test data with independent criterion data.
• Concurrent validity

It “can be described as a form of empirical validity which is determined by correlating test scores with criterion scores obtainable at the same time” (Plug, Meyer, Louw & Gouws, 1986, p. 266). It refers to the accuracy with which a test can identify or diagnose the current status of an individual’s behavior. Concurrent validity differs from predictive validity in the sense that the criterion data is available at the same time as the test data (Plug et al, 1986).

This type of validity plays an important role when new instruments are developed in order to measure the same constructs measured by other older instruments. In order to compare new and old tests, they need to be administered to a large representative sample of the population the test is developed for. The correlations between the scores on the new and old test are then computed in order to determine to what extent the two tests measure the same construct. Another procedure used to determine the concurrent validity, entails determining how well a test distinguishes between individuals who are known to differ on a specific criterion (Babbie, 1992).

If a new instrument is developed which aims at measuring the same constructs included in one or several other tests, all the instruments involved have to be administered to the same sample. The sample must be relatively large (at least 100 respondents) and representative of the population for which the new instrument is intended (Babbie, 1992; Huysamen, 1994).

• Predictive validity

This refers to the accuracy with which a test or instrument enables one to predict some future behaviour or status of individuals (Huysamen, 1994). The purpose of this type of validity is summarised in its definition. It is of the utmost importance in, for instance, the determination of the effectiveness of selection batteries in the prediction of success in training or job performance. It is however, imperative that the measures of the criteria such as success
during training or job performance are valid measures of the criteria themselves. Also, the results of such a study may only be generalised to people and criteria, which correspond to those used in the validity study (Babbie, 1992; Cascio, 1998; Huysamen, 1998). In order to determine the predictive validity of any test/instrument the following steps in the process of validation serve as a guideline (Cascio, 1998; Smit, 1996; Society for Industrial Psychology, 1992).

Step 1: A valid criterion (e.g. academic performance in the final year matriculation examination) must be identified and available in numeric format.

Step 2: Each individual's performance on the psychometric instruments should be linked to their performance or rating on the criterion. In this regard identification numbers may be used.

Step 3: The test / instrument must be administered to a relatively large sample (± 100, depending on the number of tests or instruments included in the battery) which is representative of the population for whom the test/instrument will be used.

Step 4: The data is then statistically manipulated by means of, amongst other techniques, correlations, regression analysis and discriminant analysis.

The following aspects can jeopardise all efforts at validating any instrument if not taken into account (Cascio, 1998; Smit, 1996).

* A distinction is made between the true or conceptual criterion and the available or operational measure of the criterion. This standard is often not directly measurable. An indirect measure of career success could for instance be salary. It is clear that this measure is only an indication of the criterion and not the criterion itself. It is therefore important that the criterion is reliable and valid.
When an objective measure of the criterion behavior is not readily available, ratings of individuals’ behavior are often used. In such cases, it is important to try and limit the subjectivity of raters. Usually more than one observer and multiple evaluation scales are used in an attempt to limit this problem. Criterion contamination is one of the most common problems experienced in this regard. Cascio (1998, p. 51) describes criterion contamination as “the situation in which a person’s criterion score is influenced by the rater’s knowledge of his predictor score”. This problem may lead to an artificial increase in the validity coefficient.

2.3.4 Standardisation of testing procedures

In order to conduct proper research in the validation and standardisation of psychometric instruments, in line with the Employment Equity Act (No. 56 of 1998) and other relevant legislation, testing procedures must be uniform. The following guidelines are provided (Cascio, 1998; Smit, 1996).

• The same psychometric instruments must be administered for the same purposes (e.g., the same selection battery should be used to select potential bridging candidates across the country).

• The instructions specified in manuals must be followed precisely. This should be done in order to ensure that differing instructions do not act as a confounding factor.

• All psychologists must familiarise themselves with the correct scoring and interpretation procedures as stipulated in the manual.

• Care must be taken that the correct norms are used when interpreting results. The nature of the population according to which the norms were developed must be taken into account. For example, stanine scores determined for black, white, coloured and Asian groups must be applied to members of these culture groups accordingly.
2.3.5 The conversion and interpretation of psychological test results

Norms are usually applied when psychological test results are interpreted, raw scores have little or no meaning in this regard. According to Smit (1996), more information is needed to make the interpretation meaningful. For example, the domain covered by the test should be well defined and absolute standards in relation to the test results should be available. Raw scores can be transformed into either norm-referenced scores or criterion-referenced scores.

- **Norm-referenced scores**

With a norm-referenced test score, the individual score is interpreted relative to the scores obtained by other testees. Thus, a comparison is made between the individual and other comparable persons, i.e. with individuals who belong to the same age, education and/or population represented by the normative standardisation sample (Smit, 1996).

- **Criterion-referenced scores**

According to Smit (1996), a criterion-referenced test measures the individual's ability in terms of a particular criterion that is predetermined in a specific field, for example, a driver's license test. However, it should be noted that in almost all psychological tests, the performance in the test is interpreted by comparing individual performance in the test with that of a representative sample which provides a basis for the evaluation of the individual case. According to Smit (1996), it is important that the meaning attached to the norms is limited to a specific population, i.e. that population for which the test was standardised. It is therefore of the utmost importance for the normative sample to be described in specific and accurate terms. In this way many problems surrounding the interpretation of test scores can be avoided.
It is crucial that the psychologist or psychometrist administering the psychological test is thoroughly acquainted with the contents of the test manual. Furthermore, every instruction in the manual regarding the administering and interpretation of psychological test results should be followed (Society for Industrial Psychology, 1998).

When interpreting test results which require the use of norms, the following standards stipulated by the Standards for Educational and Psychological Testing of the American Psychological Association (as cited in Smit, 1996, pp. 135-136), should be adhered to:

* In order to facilitate test interpretation, scales in which the test results are reported, as well as the underlying rationale should be accurately described in the test manual. In the manual, the method by which the raw scores were transformed to a new scale should be indicated.

* When use is made of the scale scores, the scale used should be compatible with the envisaged objective of the test and it should be described precisely.

* Norms presented in the manual should refer to clearly defined groups. These groups must consist of those persons to whom the testee is usually compared.

* In the reporting of normative data, the year when it was first gathered as well as descriptive statistics, should be given. The method of sample selection must also be given in sufficient detail to ensure appropriate evaluation.

2.3.6 Test bias

Psychological testing in South Africa is continuously subject to debate with regard to the issues of test bias and culture fair tests. In order to understand the various arguments, defining the purpose of a test acts as a point of departure. According to Lynch (1997), a test is quite clearly defined as having the central function of discrimination, a word that may seem
inappropriate. As will be discussed later, discrimination *per se* is not at issue, *unfair* discrimination however is.

A further concern is the fact that blacks tend to perform at consistently lower levels than other race groups (Cascio, 1998; Smit, 1996). Group differences may be explained in two ways:

+ There is a *real* difference in the ability being tested (which may be attributed to factors outside the test - whether social, cultural or historical).

+ There are confounding variables within the test (e.g. method effects, background knowledge) which systematically mask or distort the ability being tested.

Test bias studies are therefore directed at identifying and, where possible, reducing the effect of any confounding variables on test scores, by making changes to the scores. For this reason statistical investigations of test bias typically include controls for the ability differences between testees, so that any remaining variation in group behavior can confidently be attributed to the properties of the test itself, rather than to factors outside it (Elder, 1997).

A biased test is one where "two persons who have equal probabilities of getting the same score on a criterion, have different dimensions or combinations of dimensions of the relevant underlying ability, but the test items are selected in a way that favors one person's particular mix" (Elder, 1997, p. 263). In view of creating and sustaining fair and unbiased psychological testing, all tests which are used must be standardised. Furthermore, test users should refer to the test manuals to ensure that the correct norms are available and used for the correct groups (Schaap, 1998).

Existing psychological tests should not be totally discarded, as there they may still be valuable within the selection process. They should instead be suitably changed and refined, bias items should be removed and norms must be developed for other groups. Bias investigations like all construct validation exercises, depends on decisions regarding what a
given test is designed to measure and about the kinds of inferences that can legitimately be made from test scores (Elder, 1997). Therefore, a potential pitfall in any bias detection procedure lies in the choice of suitable control measures of ability. This means that the results of both external and internal test bias detection must be interpreted with caution.

This growing awareness of issues relating to bias, fairness and ethics, implies that further research into the effects of contextual variables such as method, background knowledge, text of item and task types is essential (Shohamy, 1997). Furthermore, the responsibility of test developers, researchers and users should not be restricted to the verification of high reliability coefficients and a statistically confirmed model of the construct being assessed. Part of the goodness of a test, as well as its validity, is dependent on the consequencess of its use (Lynch, 1997). As will be discussed in the following section, avoiding adverse impact due to psychometric assessment, is paramount.

2.4 LEGAL IMPERATIVES IMPACTING ON SELECTION

Within the South African context, psychometric assessment should not only focus on ensuring a competent and efficient human resources, but the methods and techniques used should also be legally and ethically defensible. The manner in which information is obtained about candidates, as well as the type of information obtained, have very important legal implications within the South African context.

The Constitution of the Republic of South Africa (Act No 108 of 1996) places some constraints on all aspects of the legal process. Important here is the emphasis on social justice and redressing past imbalances and injustice. The most important thread running through the following sources of law is the focus on the prevention of unfair discrimination.
2.4.1 The Health Professions Act

The Health Professions Act (1974) makes provision for the creation of the Professional Board for Psychology which through the Psychometrics Committee, ensures that professional standards are adhered to. The following are some of the professional requirements that must be adhered to by psychometrists/psychologists in the administration of psychometric testing:

♦ The test administrator must be properly trained and registered with the Health Professions Council of South Africa in order to administer, score and interpret test results.

♦ The Psychometrics Committee of the Professional Board for Psychology must classify specific tests in terms of how and for what purpose it should be used. Periodically a list of classified tests are released by the above-mentioned committee.

♦ Psychometrists/psychologists are responsible for ensuring that a specific test is used for the purpose intended by the Psychometrics Committee of the Professional Board for Psychology.

♦ A further aspect, implied by the previous requirement, is that psychometrists/psychologists should ensure that a specific test is valid for the purposes for which it is being used.

♦ Psychometrists/psychologists should ensure that appropriate norms are consulted.

♦ If foreign instruments are used, appropriate research should be conducted to assess whether the test is culturally biased and great care should be taken when scores are interpreted.
2.4.2 The Employment Equity Act

The Employment Equity Act (1998, p. 15) states that: "Psychological testing and other similar assessments of any employee are prohibited unless the test or assessment being used has been scientifically shown to be valid, reliable; can be applied fairly to all employees; and is not biased against any employee or group. It further states that unfair "discrimination refers to the unfair treatment of a person or group of persons either directly or indirectly, intentionally or unintentionally, based on, but not limited to, his or her race, gender, sex, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, political opinion, culture, language, marital status, pregnancy or family responsibility." Furthermore, "Discrimination on one or more of these grounds is unfair unless it is established that such discrimination is fair." Also mentioned is the fact that it is unfair to discriminate when "... making distinctions or choices based on factors other than individual merit or established standards that have an adverse impact on an individual or members of a group" (Employment Equity Act, 1998, p. 14).

2.4.3 The Labour Relations Act

The Labour Relations Act (1995, p. 13) states its overriding purpose as being "the advancement of economic development, social justice, labour peace and the democratisation of the workplace". The following objectives are stated as follows:

- To give effect to and regulate the fundamental rights contained in Section 27 of the Constitution;

- To give effect to the duties of the Republic as a member of the International Labour Organisation;

- To provide a framework in which employees and their unions, employers and employer organisations can bargain collectively to determine the wages, terms and
conditions of employment and other matters of mutual interest, and formulate industry policy; and

- To promote orderly collective bargaining.

Amongst other requirements, organisations need to ensure valid and fair selection and recruitment practices in order to comply with the Labour Relations Act (1995). Failure to comply, would result in the commission of an unfair labour practice. The Labour Relations Act (1995, p. 253), defines it as “any unfair practice or omission which arises between an employer and employee, involving ... the unfair discrimination, either directly or indirectly, against an employee on any arbitrary ground, including, but not limited to race, gender, sex, ethnic or social origin, colour, sexual orientation, age, disability, religion, belief, political opinion, culture, language, marital status or family responsibility”.

For the purposes of this Act, the term employee also refers to a job applicant. Therefore, its implications for recruitment and selection are clear in terms of preventing access discrimination.

2.4.4 Implications for psychometric assessment

The above discussion may create the impression that any form of discrimination is unlawful. This is, however not the case. Fair discrimination is permissible when it “is based on the requirements of a job or post”. In fact, all the above mentioned sources of law recognise genuine job requirements as justification for differentiation between candidates. Clearly, the only manner in which the “genuine job requirements” can be determined is by conducting proper job analysis (Cascio, 1998; Taylor, 1994). This aspect was dealt with at length in previous sections.

What the above implies is that the I/O Psychologist has to act in a very circumspect manner when eliciting information regarding a selection candidate. Eliciting information of this
nature should be clearly linked to the inherent requirements of the job and not merely reflect the wishes and/or preferences of decision makers. Furthermore, the onus or burden of proof rests on the decision maker to prove that this kind of information did not have an adverse impact on the candidate's prospects of being selected for a particular job or position.

2.5 SUMMARY

In this chapter, it was emphasised that the selection of personnel needs to be pursued in a holistic and scientific manner. It was further emphasised that psychometric assessment is the crux of the selection process. The limitations of psychometric tests were also briefly discussed. In this regard, perhaps other forms of assessment (e.g. in-basket exercises and simulation exercises) should be utilised in addition to psychometrics in order to compensate for these limitations. The conclusion can be drawn however, that despite these limitations, psychometric assessment may still be a useful tool within the broader selection process. Furthermore, despite trade union objections to its use, it is suggested that the baby not be thrown out with the bath water, as it were. If approached in a scientific manner, i.e. if the prerequisites of psychometric assessment as discussed are adhered to, psychometric tests may contribute significantly to the effective selection and utilisation of human resources.

Finally, in this chapter, the legal prerequisites regarding the use of psychometrics within the selection context was discussed. Despite the objections by business that certain legal requirements are unnecessarily intrusive, the conclusion is drawn that these laws (especially the labour related laws) do not contradict sound scientific practice. These laws should instead be viewed as complimentary to the dictates of sound science as well as a means of improving the quality of the selection process and the decisions that emanate from it.

This chapter serves to provide a scientific and legal backdrop against which psychometric assessment should proceed within the South African context. It demonstrates that, instead of being a mere formality or typical rubber-stamping activity, psychometric assessment should rather serve as a useful and complimentary activity to the entire selection process. It can,
however, not be overstated that it is not a panacea for all selection problems. If scientifically structured and based on the inherent requirements of the job as well as the needs of the organisation, it could be one of the most useful tools in personnel assessment. Approached in this manner, psychometric measurements could potentially satisfy the demands of the business and labour sectors, as well as those of the broader scientific community.
CHAPTER 3 THE NATURE AND ASSESSMENT OF INTELLIGENCE

3.1 INTRODUCTION

The history of intelligence testing is inextricably linked to that of psychological assessment and indeed the field of Psychology in general. In fact, the development of intelligence testing is viewed by many as the genesis of psychometric assessment (Smit, 1996; Walsh & Betz, 2001).

Throughout history, intelligence has been conceptualised in various ways. The conceptualisation and theories of intelligence are discussed as they directly impact on the measurement of intelligence. For the purposes of this investigation, the inclusion of historical developments in intelligence testing, is motivated by the fact that the independent variables included in the empirical study (i.e. the APIL Battery and Raven's APM) are notably informed by these developments. Furthermore, much of our current understanding of intelligence can be attributed to the work of earlier theorists (Raven & Court, 1998; Walsh & Betz, 2001; Weiten, 1994). In this regard, current trends in terms of defining intelligence and determining its dimensions are considered.

Internationally, as well as in the South African context, the concept intelligence and its measurement has been subjected to a considerable amount of polemic within both scientific and non-scientific circles. This is mainly due to the fact that blacks (in South Africa) and minorities (in the American context) generally obtain lower scores than their white counterparts. When intelligence testing is conducted for selection purposes, it may translate into these groups experiencing large-scale adverse impact. This state of affairs has led to questions regarding the possibility that traditional intelligence tests are culturally biased in favour of whites (Lopes, Roodt & Mauer, 2001; Van Rooyen, 2001; Zwick, 1999). In this regard, the culture fair measurement of intelligence is afforded attention. Finally, in keeping with the aims of this investigation, the relationship between intelligence and academic achievement is discussed. Also, factors that may potentially have a mitigating impact on this relationship, especially emotional intelligence, are discussed.
3.2 DEFINITION OF INTELLIGENCE

As was pointed out in chapter 1, intelligence is an abstract concept and not an entity in its own right. It is a characteristic that cannot be possessed, but merely a manifestation in behaviour. As such, it can therefore not be measured directly. However due to the fact that intelligence is relatively stable over time and situations, its manifestations in behaviour in a specific situation can be measured (Cascio, 1998; Smit, 1996).

Baron (1995) and Smit (1996) distinguish between the following approaches to the classification of intelligence:

- The first approach defines intelligence in terms of the ability to adapt to the environment. Intelligence is therefore “the ability to adapt to new situations with the help of conscious thought processes or to satisfy the requirements that have been established” (Smit, 1996, p. 178).

- Secondly, intelligence is equated with learning ability. Based on the fact that intelligence has been shown to be positively correlated with learning ability, the assumption is made that they are the same.

- Finally, there are those definitions that emphasise the abstract thinking processes. This would include “the ability to handle symbols effectively in the consciousness in order to solve problems, to see relationships, to abstract general principles, and to make deductions” (Smit, 1996, p. 178).

According to Wechsler (as cited in Smit, 1996, p. 178), intelligence is “the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment”. He sees it as global to the extent that it is characteristic of the individual’s behaviour as a whole. It is an aggregate to the extent that it consists of clearly distinguishable qualities.
Psychologists tended to disagree quite sharply regarding the issue of whether intelligence is a single characteristic or whether it consists of distinct parts. Some, e.g. Spearman supported the former view. He expressed intelligence as a function of a general ability $g$, which manifested itself in all forms of intelligent behaviour (Baron, 1995). Others (e.g. Sternberg) tend to believe that intelligence is composed of many separate mental abilities that function independently.

According to Baron (1995), Owen and Taljaard (1996) and Smit (1996) intelligence is not merely the sum of a series of intellectual abilities. Various reasons for this exist:

- Firstly, the end – product of intelligence also deals with the manner in which various components are *combined*.

- Other factors also play a role in intellectual performance such as motivation, drive and mood.

Clearly, there is still considerable disagreement regarding the exact definition of intelligence. However, for the purposes of this investigation the following conclusions will be drawn. Firstly, intelligence is multi – faceted. Secondly, it is a combination of general reasoning and problem solving abilities as well as specific subcomponent abilities (Baron, 1995; Owen & Taljaard, 1996; Smit, 1996).

Different intelligence tests emphasise the above to different degrees. Therefore, although a wide variety of psychometric instruments may be categorised as intelligence tests, they may not measure exactly the same thing. Therefore a clear understanding of the different approaches to the conceptualisation of intelligence becomes essential as it directly impacts on the manner in which a specific test is constructed as well as the way in which a specific test should be used and interpreted (Smit, 1996).
3.3 THE CONCEPTUALISATION AND THEORIES OF INTELLIGENCE

As was mentioned earlier, the theories of and approaches to the conceptualisation of intelligence have greatly influenced its measurement. All of these theories on the nature and measurement of intelligence originate in Differential Psychology. This area of psychology specifically focuses on the measurement of individual differences in terms of mostly latent, psychological constructs such as intelligence. This approach makes use of statistical techniques such as correlation and factor analytical methods (Baron, 1995; Smit, 1996; Walsh & Betz, 2001; Weiten, 1994). The following is a brief discussion of the most important developments in the description and measurement of intelligence.

3.3.1 The Galton – Cattell approach

The end of the nineteenth century scientists marked the onset of exploration into the nature and measurement of intelligence in the field of psychology. Galton is considered to be the first individual to attempt the scientific measurement of intelligence. He proposed the existence of a general mental ability. The theory is based on the assumption that information reaches humans through the senses and that intellect is the sum of all the constituent parts of sensory functioning (Owen & Taljaard, 1996; Walsh & Betz, 2001; Weiten, 1994).

According to Galton, intelligence is a function of the speed and refinement of sensory responses to environmental stimuli. James Cattell, a student of Wundt’s, became interested in the measurement of intelligence. He collaborated with Galton on a number of anthropometric laboratory studies focussing mainly on sensory functioning as the key to intelligence. Cattell termed these mental tests, the first time that this concept was used (Owen & Taljaard, 1996).

The results and indeed these experiments themselves, were later discredited due to the fact that scores obtained on various tests did not correlate highly with each other nor with other intuitive measures of intelligence such as teacher’s ratings. Due to this and other criticisms, this approach was abandoned. Later studies however, found support for Galton’s “general mental ability” concept and Cattell’s conceptualisation of fluid and crystallised intelligence (Walsh & Betz,
Cattell considered intelligence as consisting of two main components i.e. **fluid** and **crystallised** intelligence. Fluid intelligence refers to “the ability to form concepts, reason and identify similarities; it is more intuitive and is active in forming new mental structures” (Baron, 1995, p. 417). Crystallised intelligence deals with aspects “that involve drawing on previously learned information to make decisions or solve problems” (Baron, 1995, p. 417). This aspect is especially used in classroom learning as well as in vocabulary tests. Baron (1995) states that research has shown that fluid intelligence peaks in early adulthood, while the latter develops throughout the lifespan.

### 3.3.2 Binet – Simon’s conception of intelligence

At more or less the same time, Binet and Simon developed a distinctly different approach to intelligence testing. They viewed intelligence as involving “higher” mental processes of judgement and reasoning rather than the sensory capabilities emphasised in the former approach. This approach placed greater emphasis on sound reasoning, judgment and common sense capabilities (Walsh & Betz, 2001).

A further shift was the assumption that these higher mental processes increase as a child ages and stabilises in adulthood. This gave rise to the concept **mental age** as opposed to chronological age. This is a notion that is still prevalent in the thinking on intelligence testing today (Walsh & Betz, 2001). In addition to the above, one of their greatest contributions to the measurement of intelligence is the Binet – Simon test that was later revised, augmented and standardised by Lewis Terman at Stanford University (Owen & Taljaard, 1996; Weiten, 1994).

### 3.3.3 Spearman’s two – factor theory

Charles Spearman’s studies of Galton – Cattell’s research found that there were correlations between some of the mental tests. Using factor analysis, conceived by himself, Spearman came to the conclusion that there was evidence for the existence and measurability of Galton’s “general
Spearman, in his two-factor approach to intelligence, proposed that every mental test measured a general factor $g$ and a specific factor $s$ that was unique to specific test items. In collating the scores on a number of tests the $s$ factors could be cancelled out, leaving a “residue” or indicator of general mental ability (Smit, 1996; Walsh & Betz, 2001).

Spearman’s main contribution lies in his assumption that the specific test content is not actually significant in the measurement of intelligence. This so-called indifference of the indicator idea states that test content is merely a vehicle for eliciting and measuring $g$. Spearman stated that essentially, intelligence contains two elements, i.e. education of relations (inductive reasoning) and education of correlates (deductive reasoning). These concepts currently remain the basis for intelligence measurement. Also, most intelligence tests in use currently are considered to be $g$-saturated tests. Of significance, for the purposes of this investigation, is the fact that both independent variables measured are considered to be $g$-saturated (Raven & Court, 1998; Taylor, 1997; Walsh & Betz, 2001).

### 3.3.4 Sternberg’s triarchic theory

Sternberg identified three types of intelligence. The first, componential intelligence, is the ability to think in an analytical and critical manner. Furthermore, it emphasises effectiveness in information processing. This ability is crucial in terms of academic achievement. Secondly, experiential intelligence focuses on insight and the ability to formulate novel ideas. This ability is useful in integrating a number of unrelated facts and to focus on what information is crucial in a specific situation. Finally, contextual intelligence relates to practical and adaptive qualities in a changing environment. It involves the ability to prepare for problem solving in a given situation (Baron, 1995; Walsh & Betz, 2001; Weiten, 1994).

Sternberg expanded his theory to cover the link between intelligence and personality. He developed the concept intelligence style to describe the manner in which the above mentioned three types of intelligence are used in everyday life (Baron, 1995).
3.3.5 Multiple factor theories

As a direct result of the aforementioned factor analytical approach, multiple factor theories came into being. Essentially this approach focuses on the exploration of commonalities or shared variance in the previously mentioned factors. Thurstone subjected a number of mental tests to factor – analysis and concluded that there is not just one general ability or \( g \) – factor which accounts for success in all intellectual pursuits. Instead, he found that there are a limited number of group factors that consistently occurred. He identified seven factors or primary mental abilities (Baron, 1995; Smit, 1996; Walsh & Betz, 2001). These are:

- **Verbal Comprehension** (V) is measured by a vocabulary test, reading comprehension and verbal analogies.
- **Word Fluency** (W) is measured by anagrams and/or rhyming tests.
- **Numerical Ability** (N) entails the speed and accuracy with which arithmetic computations are made.
- **Spatial Ability** (S) entails the ability to see spatial relations and visualise imagery.
- **Associative Memory** (M) normally deals with the ability for memory of paired associates.
- **Perceptual Speed** (P) deals with the ability to quickly grasp details, similarities and differences in visual images.
- **Reasoning Ability** (R) is normally measured by arithmetic reasoning tests, number series as well as inductive and deductive reasoning tests.
The result of this conceptualisation was the so-called primary mental abilities (PMA) test (Smit, 1996; Walsh & Betz, 2001). In this regard, Cronbach (1990, p. 326) states that by using this term, Thurstone “intended to suggest that these abilities combine to produce success in any intellectual performance”. When conducting later research, Thurstone came to the conclusion that the PMA’s were not independent of each other and that g still played a role (Smit, 1996). In the final analysis, it would seem that Thurstone and Spearman’s theoretical constructions do not differ significantly (Smit, 1996). Spearman started out with g and later acknowledged group factors, whilst Thurstone started with group factors and was later convinced that g does play a role.

Another major contributor to multiple factor theories, Guilford, proposed the structure of the intellect (SI) model which is an expansion of Thurstone’s work. This model suggests that the intellect can be described and classified along three basic dimensions; operations (intellectual processes), content and product (Smit, 1996; Walsh & Betz, 2001). It deviates from other theories in the sense that it demonstrates that intellectual factors do not necessarily have to be independent. Secondly, this model provides for all the possible factors that may be involved in intelligence (Smit, 1996).

Although in totality this model is relatively cumbersome for general purposes, it is quite useful for the measurement of very specific abilities. The value of multiple factor theories, lies in the fact that it broadened the view of intelligence to also include factors other than g. Individuals who perform equally in terms of g may diverge in terms of the specific pattern of their abilities. These findings and the various test batteries that developed as a result, have had very important implications for education and occupational selection purposes as success in these environments often rely on very specific capabilities (Baron, 1995; Walsh & Betz, 2001; Weiten, 1994).

3.3.6 The hierarchical model: an integration

A further development was the development of hierarchical models of the nature of mental abilities. The basic assumption of this model is that intelligence is represented in a hierarchical form “with one or more broad general factors ... with group factors and more specific factors” (Smit, 1996, p. 186). Various mental abilities are arranged according to generality versus
specificity. Spearman's $g$ is seen as the most general component of intelligence, followed by two broad major group factors, verbal - educational ($v : ed$) and spatial - mechanical ($k : m$). These two group factors are then further divided into minor group factors. The two major group factors closely resemble Cattell's conception of intelligence. Currently, most psychologists are of the opinion that intelligence is composed of both general ability as well as specific abilities (Smit, 1996; Walsh & Betz, 2001).

Smit (1996) considers this model to be a modernisation of Spearman's work. In terms of its practical value, Brown (as cited in Smit, 1996, p. 187) states that:

"The hierarchical approach represents an intuitively satisfying collection of data and, since tests can be constructed to represent any level in the hierarchy, it is a useful tool for guiding test construction".

3.4 THE MEASUREMENT OF INTELLIGENCE

At the outset it needs to be re-emphasised that the measurement of intelligence is based on a process of deduction and abstraction. Intelligence is not directly measurable. A person's intellectual potential can however be measured. Intelligence tests merely provide an operational definition of intelligence (Smit, 1996).

Intelligence tests can be divided into two categories, i.e. individual and group tests. Although individual tests are in general considered to be more valid, they tend to be costly and time consuming. The Stanford-Binet and Wechsler tests are both individual intelligence tests (Baron, 1995; Walsh & Betz, 2001; Weiten, 1994). The Raven's APM and the APIL Battery are included here as they form an integral part of this investigation. Furthermore, for obvious reasons, the emphasis here is on tests that have demonstrated high predictive validity in terms of scholastic/academic performance.
3.4.1 Individual intelligence tests

Individual intelligence tests are in general considered to be more valid measures of intelligence. In addition, they can often be used to provide clinical information about, for example, intellectual deterioration and psychotic behaviour (Baron, 1995; Weiten, 1994).

3.4.1.1 Binet – Simon's Intelligence Quotient

Since the development of the Binet–Simon intelligence test for the identification of mental retardation in children in 1905, many developments have occurred in the measurement of intelligence. Terman expanded the work of Simon and Binet and published the Stanford–Binet test in 1916 (Weiten, 1994). The fact that this test yielded a single score, i.e. the well-known Intelligence Quotient (IQ) contributed to its popularity.

The IQ score is calculated by dividing the mental age by the individual's chronological age multiplied by 100. If a person's mental age is equal to the chronological age the IQ is equal to 100. A score above 100 indicates a level of intelligence higher than that of the average person of that age. Contrastingly, scores below 100 indicate that the individual is less intelligent than his/her peers (Baron, 1995; Weiten, 1994).

Intelligence increases with age, with the result that there is a constant ratio between mental and chronological age up to age 15. At this stage, the mental age of a person does not increase any longer. As a result of this problem, Wechsler devised the so-called deviation IQ that is currently in use. This means that subjects' scores are compared to the normal distribution. In this system, the mean of the distribution is set at 100 and the standard deviation at 15. For example, a score of 115 indicates that a person scored exactly one standard deviation above the mean. Conversely, a score of 85 implies that a person scored one standard deviation below the mean. A score of 100 indicates average academic performance. Currently the term IQ is still used to denote cognitive ability in both academic as well as popular contexts (Baron, 1995; Weiten, 1994).
Currently, the Stanford–Binet test measures four types of reasoning that is seen as representative of the construct, intelligence. These components are verbal reasoning, quantitative reasoning, abstract visual reasoning, and short-term memory. It has also been updated regularly since 1916 (in 1937, 1960 and 1986) and remains one of the most popular measures of intelligence today (Owen & Taljaard, 1996; Smit, 1996; Weiten, 1994).

3.4.1.2 The Wechsler Scales

Wechsler found the Stanford–Binet test unsatisfactory in the assessment of adults. A further drawback of this test was the fact that it was verbal in nature and neglected the fact that intelligence can also manifest non-verbally. In 1939, Wechsler published the first intelligence test for adults, the Wechsler Adult Intelligence Scale (WAIS).

The WAIS incorporated at least two innovations. Firstly, non-verbal or performance items were incorporated. In order to emphasise the verbal and non-verbal components, different scores for verbal, non-verbal and total or full-scale IQ were computed. Secondly, the intelligence quotient was discarded in favour of a scoring scheme that made use of the normal distribution.

Today, the Wechsler scales are the most commonly used individual measure of intelligence worldwide. An overview of the dimensions of the Wechsler Adult Intelligence Scale–Revised (WAIS–R) is presented in Table 3.1 below.

**TABLE 3.1 SUBTESTS OF THE WAIS–R**

<table>
<thead>
<tr>
<th>TEST</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Tests</td>
<td>Examinees are asked to answer general information questions, increasing in difficulty.</td>
</tr>
<tr>
<td>Information</td>
<td>Examinees are asked to repeat series of digits read out loud by the examiner.</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Examinees are asked to answer questions requiring detailed answers. Answers indicate their comprehension of the question.</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Similarities</td>
<td>Examinees indicate in what way two items are alike.</td>
</tr>
<tr>
<td>Performance Tests</td>
<td></td>
</tr>
<tr>
<td>Picture Completion</td>
<td>Examinees indicate what part of each picture is missing.</td>
</tr>
<tr>
<td>Picture Arrangement</td>
<td>Examinees arrange pictures to make a sensible story.</td>
</tr>
<tr>
<td>Block Design</td>
<td>Examinees attempt to duplicate designs made with red and white blocks.</td>
</tr>
<tr>
<td>Object Assembly</td>
<td>Examinees attempt to solve picture puzzles.</td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>Examinees fill in small boxes with coded symbols corresponding to a number above each box.</td>
</tr>
</tbody>
</table>

(Adapted from Baron, 1995, p. 423)

3.4.2 Group intelligence tests

In the interest of financial and time efficiency, group tests were designed. They are normally paper – and – pencil tests that can be used to test large groups of people at a time. Administration procedures and instructions are strictly standardised in order to ease administration and the training of administrators. These factors do have positive implications for the reliability of group tests (Baron, 1995; Walsh & Betz, 2001; Weiten, 1994). Disadvantages associated with the use of some group tests are that they require a certain level of linguistic ability as well as physical, emotional and mental normality (Baron, 1995; Walsh & Betz, 2001; Weiten, 1994).

Although group tests are very popular, especially within the educational and selection contexts, they are also subject to a great deal of criticism. This is mainly due to the suggestion that they unfairly discriminate against candidates from disadvantaged backgrounds (Baron, 1995).
These and other aspects surrounding the validity of intelligence tests for people from disadvantaged backgrounds will be addressed in a later section.

3.4.2.1 Army Alpha and Army Beta

Movements in this direction first started during World War I, when Americans needed to psychologically screen thousands of combatants. This led to the development of the Army Alpha (for persons who could read) and Army Beta (for those who were illiterate or not fluent in English). Subsequently, many other group tests were developed. Some of the more popular tests are the Otis, Henmon - Nelson and the Cognitive Abilities Test (Baron, 1995).

3.4.2.2 Raven’s Advanced Progressive Matrices

Raven’s APM proved to address many of the concerns raised above. It is a non-verbal test of an individual’s capacity to apprehend figures presented for his/her perception, see the relations between them, and conceive the correlative figures completing the system of relations presented. To assess a person’s intellectual efficiency, i.e. the speed with which the individual can do accurate work, the test is administered within a specified time. Administering the test in an untimed manner, gives an indication of intellectual capacity (Raven & Court, 1998).

The test consists of Set I and II and is used for people over 11 years of age and of average or above average intellectual ability. In the first set there are 12 problems. They are designed to introduce a person to the method of working. In the second set there are 36 problems. As this test is designed to differentiate between people of superior intellectual ability, it is often used to select staff for high-level technical or managerial positions (Bors & Stokes, 1998; Raven & Court, 1998).

3.4.3 South African - based intelligence tests

Considering the diversity of the South African population as well as the need to ensure
measurement validity, various intelligence tests were adapted and/or developed for the South African context. The following represents an overview of the most important individual and group tests in this category.

3.4.3.1 Individual intelligence tests

- The Senior South African Individual Scale (SSAIS)

This scale was developed by the national Council for Education and Social Research and released in 1964. From 1980, it became known as the Senior South African Individual Scale (SSAIS). This scale is meant for individuals between the ages of five and seventeen who are proficient in either Afrikaans or English (Smit, 1996). Owen and Taljaard (1996, pp. 107 - 108) state that the main aims of the SSAIS are:

- To provide a valid and reliable instrument to measure a testee's general intellectual ability;
- To provide a separate measurement of a child's ability to solve intellectual problems involving words and symbols, as well as problems involving the observation of visual patterns and the handling of objects; and
- To provide separate measurements of a number of facets of intelligence that have been demonstrated to have clinical significance.

This scale is based on Wechsler's work and is considered to be $g$ - saturated. A central assumption is that "intelligence comprises related problem-solving abilities, some of which are more closely associated with effective functioning at school and consequently the prediction of scholastic achievement" (Owen & Taljaard, 1996, p. 108). The SSAIS contains both verbal and non-verbal subtests (Owen & Taljaard, 1996; Smit, 1996). The verbal scale comprises:

- Vocabulary - which focuses on knowledge of words without utilising reading and writing abilities.
➢ **Comprehension** - performance is based on *common sense* or judgement.

➢ **Verbal reasoning** - focuses on the ability to distinguish between essential and superficial similarities as well as abstract thinking.

➢ **Problems** - measures the ability to handle numerical concepts and concentration.

➢ **Memory** - aimed at providing a measure of concentration and short-term memory.

The non-verbal scale consists of:

➢ **Pattern Completion** - measures visual perception, orientation and logical thinking.

➢ **Blocks** - aims at measuring synthesising, analysing and spatial reasoning abilities.

➢ **Absurdities** - measures basic perceptual and concept-formation abilities in terms of visual recognition, identification and comprehension of known objects, forms and situations.

➢ **Form Board** - entails the synthesis of concrete visual forms, i.e. it measures effectiveness regarding visual orientation and perception.

It should be pointed out that these subtests were named for convenience and do not necessarily reflect the contents of each subtest. Most of the scales contained in the SSAIS are considered to show high reliability and predictive validity in terms of scholastic achievement (Owen & Taljaard, 1996; Smit, 1996).

◆ **The Junior South African Individual Scales (JSAIS)**

The aim of this instrument is to measure as many of the aspects of mental abilities as possible (Owen & Taljaard, 1996). The main focus is on intelligence in terms of scholastic performance. This *g*-saturated test is further aimed at identifying mental retardation in order to place children
into special education facilities. The JSAIS is applicable to children between the ages of three and seven who are either English or Afrikaans-speaking. When used on other language groups, the validity of the test seems to decrease.

The JSAIS consists of 22 separate subtests. However, based on factor-analysis, 12 of the subtests were selected to create the so-called Global IQ (GIQ). The GIQ is dependent on performance on all 12 subtests (Owen & Taljaard, 1996; Smit, 1996). The 12 subtests that form the GIQ are:

- **Vocabulary** - indicates the ability to recognise words and understand their meanings.

- **Ready knowledge** - measures the testee's general knowledge of his/her immediate environment.

- **Picture riddles** - indicates the ability to understand and reinterpret the spoken word.

- **Word association** - measures the ability to think rationally.

- **Story memory** - measures the ability to remember meaningful verbal material on a short-term basis.

- **Form discrimination** - measures an important part of visual – spatial reasoning ability, i.e. the visual discrimination.

- **Absurdities A** (Missing parts) - measures the ability to evaluate the correctness of visual stimuli.

- **Absurdities B** (Absurd situations) - measures the ability to perceive absurdities in visual stimuli.

- **Block patterns** - a non-verbal measure of intelligence, i.e. visual – spatial reasoning ability.
form board - entails the synthesis of concrete visual forms, i.e. it measures effectiveness regarding visual orientation and perception in terms of three-dimensional figures.

number and quantity concepts - measures the manner in which the candidate handles simple arithmetical and calculation problems.

memory for digits - measures auditory short-term memory.

A shortened combined scale, GIQ 8 Scale, is also available. Owen and Taljaard (1996) report satisfactory reliability figures as well as a high agreement with teachers' ratings of the intellectual capability of children.

- the south african wechsler adult intelligence scale (SAWAIS)

The National Institute for Personnel Research (NIPR) adapted and standardised the WAIS for the South African context. The main deviation from the original test is the fact that an IQ measurement can be directly obtained by adding the subtest standard points. In terms of the dimensions measured, it is similar to the original test.

- other individual tests

Other South African individual scales include the following:

- the individual scale for Indian South Africans (ISISA);

- South African Individual Scale for the Blind (SAISB); and

- the snijders – oomen non – verbal Intelligence Scale (SON-scale)
3.4.3.2 Group intelligence tests

♦ The New South African Group Test (NSAGT)

The NSAGT measures the general level of mental functioning. The NSAGT series consists of three versions: the junior series (8 - 11 years), the intermediate series (10 - 14 years) and the senior series (13 - 17 years). The tests include both verbal and non-verbal subtests (Owen & Taljaard, 1996; Smit, 1996). According to Smit (1996, p. 192) the "reliability of the test is generally high and the concurrent validity also appears to be high". The value of the test lies in its ability to accurately predict scholastic performance (Owen & Taljaard, 1996; Smit, 1996).

♦ The APIL Battery

The APIL Battery is a group test designed by Taylor (1997) to measure a person's underlying ability to learn cognitively challenging new skills. It assesses an individual's fundamental cognitive capabilities and potentialities. It is considered useful in identifying candidates who are likely to master new cognitively demanding material in a formal training context, especially with the view of identifying individuals with potential for tertiary level development.

It provides a profile of 8 scores and a learning curve. These are fundamental building blocks of intellectual competence - the cognitive materials needed to develop skills required to master new challenges. It further indicates the individual's intellectual adaptability rather than his or her already acquired skills or abilities (Taylor, 1997).

♦ Other group tests

Other group tests include the group test for Coloured pupils, the group test for Indian pupils, the test for mental alertness, the general Scholastic Aptitude Test (GSAT) and the NB - group test for five and six - year olds and for seven and eight - year olds (Owen & Taljaard, 1996; Smit, 1996).
3.5 THE RELATIONSHIP BETWEEN INTELLIGENCE AND ACADEMIC ACHIEVEMENT

Researchers have often investigated the relationship between intelligence scores and indicators of academic performance, such as examination results. As a result, the role of intelligence tests in academic achievement has been established by a number of such studies (Baron, 1995; Weiten, 1994). As was pointed out earlier, a number of intellectual dimensions have been found to correlate significantly with academic achievement. Some of these are Cattell's *crystallised intelligence* and Sternberg's *componential intelligence*. Most of the measuring instruments discussed also seems to be positively correlated with academic performance. In terms of using intelligence testing for academic selection, the predictive validity of these has been established. According to Fourie (1991), there is a direct relationship between intelligence and academic achievement. Intelligence plays a decisive role in determining the quality of learning and academic achievement (Ausubel, 1968; Fourie, 1991; Garbers & Van Aarde, 1978; Viljoen & Vos, 1998). Van Rooyen (2001) reports that step-wise regression analysis demonstrated that a combination of cognitive factors has the greatest predictive validity in terms of predicting academic success at university level. Schmidt and Hunter (1998) conducted a meta-analytical study on the impact of cognitive mental ability tests on performance, spanning 85 years of research. Their results indicate that on average general mental ability (Spearman's \( g \) - factor) correlates 0.54 with training success especially where learners are expected to absorb knowledge that is directly presented to them. In this regard, Weiten (1994, p. 236) unreservedly states that intelligence tests "are valid measures of the kind of intelligence that's necessary to do well in academic work".

Despite the abovementioned findings, the possibility that traditional intelligence tests may adversely impact on disadvantaged groups has repeatedly been raised (Foxcroft, 1997; Nainaar, 1999; Smit, 1996; Society for Industrial Psychology, 1998; Weiten, 1994). For decades, the validity of various cognitive measures for predicting the academic performance of "non-white" testees have been questioned (Van Rooyen, 2001). A number of international as well as South African research studies have found that blacks and members of other ethnic groups, routinely achieve lower scores on intelligence tests compared to their white counterparts (Goldman &
Hewitt, 1976; Pfeifer & Sedlacek, 1971; Van Rooyen, 2001; Viljoen & Vos, 1998). Furthermore, research results indicate that intelligence tests have a lower predictive validity when applied to members of disadvantaged groups (Baron, 1995; Smit, 1996; Van Rooyen, 2001; Viljoen & Vos, 1998; Weiten, 1994). Researchers such as Fourie (1991) and Van Rooyen (2001) therefore caution against the indiscriminate use of cognitive measures in academic selection. Although there is a relationship between intelligence and academic achievement, the specific intelligence measure used in selection, needs to be scrutinised very carefully for potential cultural bias.

3.6 FACTORS MODERATING THE RELATIONSHIP BETWEEN INTELLIGENCE AND ACADEMIC ACHIEVEMENT

The fact that intellectual ability has often been confirmed to be positively related to academic performance, has led to a situation where it is often used as the only predictor of success. This section however, attempts to highlight additional factors that may moderate the relationship between intelligence and academic success, especially within the South African context.

3.6.1 Bias in the assessment of intelligence

As was mentioned earlier and in chapter 2, psychometric assessment in general has been exposed to severe scrutiny over the last few decades. Intelligence testing has not been overlooked in such debate. Much of the criticism levelled against it is based on the perception that intelligence tests place testees from previously disadvantaged groups at an unfair disadvantage. This situation is exacerbated by South African practitioners’ dependence on psychometric tests that were developed abroad as well as a lack of research capability in order to conduct proper validation studies (Nainaar, 1999; Smit, 1996; Society for Industrial Psychology, 1998).

Furthermore, the adaptation of foreign intelligence tests and its adaptation for the South African population is also viewed with suspicion (Baron, 1995; Nainaar, 1999). Considering this state of affairs, as well as the legal constraints impacting on psychometric assessment in general, the culture – fair assessment of intelligence becomes essential (Potgieter & Van der Merwe, 2001; Van Rooyen, 2001).
As was mentioned earlier, a long history of unequal treatment in South Africa has led to the development of a number of discrepancies between the various races. As a result, blacks are currently confronted with deficiencies in terms of socio-economic status as well as educational backlogs. All of these factors may have a negative impact on the performance of disadvantaged group in intelligence tests (Nainaar, 1999; Smit, 1996; Taylor, 1994).

* Ethnicity and culture

Many critics maintain that the reason for racial discrepancies in performance in intelligence tests is the fact that cultural bias is built into intelligence tests (Claassen, 1995; Foxcroft, 1997; Owen, 1991). This implies that because tests were developed by and for a specific cultural group, individuals from other cultural groups may be at a disadvantage when taking them (Baron, 1995; Smit, 1996, Weiten, 1994). Within the American context, Mercer (as cited in Weiten, 1994, p. 244) states that “when IQ tests are given to minorities, they measure both mental ability and assimilation into the mainstream”. Therefore, differing values seem to be a more subtle form of cultural bias. Considering the unequal development between various ethnic and racial groups within the South African context, this reality is also prevalent (Foxcroft, 1997; Lopes et al, 2001). This is illustrated by Claassen (1995, p. 8) when he states that “testing and measurement have as objective to reflect some aspect of the world. It can be expected that tests will reflect the nature of the society in which they are used”.

* Language

Research has shown that language impacts directly on performance in intelligence tests. For instance, studies conducted by Owen (1991), Van Rooyen (2001) and Viljoen and Vos (1998), have found that speaking English as a home language acts as a decisive advantage in terms of academic achievement. For example, Viljoen and Vos (1998) found a significant significant (p<0.01) and relatively high positive correlation (r = 0.622) between home language and academic performance. Van Rooyen (2001) also found that speaking English as a home language, has the highest significant (p<0.01) correlation (r = 0.3287) with the average academic achievement of
students in a university bridging—course. In South Africa most intelligence tests and their instructions are only available in Afrikaans and/or English. Potgieter and Van der Merwe (2001) note that translation may not necessarily be an adequate response to these difficulties. The following are some of the problems associated with translation:

* Translation may add to or lessen the meaning of the test content that was not intended by the developer.

* Some words, expressions and/or phrases do not have an equivalent in another language.

Education and age

"It is generally accepted that the academic achievement of students from the previous black educational departments is impaired because of educational and academic disadvantages" (Van Rooyen, 2001, p. 187). Many intelligence tests assume a certain amount of pre-knowledge that is expected at certain ages. Due to discrepancies between various races regarding the age at which schooling starts as well as qualitative differences in education, members from disadvantaged groups may not have had the opportunity to acquire such knowledge. This may place them at a distinct disadvantage compared to their white peers (Taylor, 1997). However, research conducted by Van Rooyen (2001) as well as Viljoen and Vos (1998) did not produce significantly different academic results between students of different age groups.

3.6.2 Optimalising the culture fair assessment of intelligence

Intelligence testing will probably never be completely culture-free, nor can cultural bias be completely eliminated (Owen & Taljaard, 1996; Smit, 1996; Weiten, 1994). Given this, the following measures are suggested in order to reduce the adverse impact that intelligence testing may have on candidates from disadvantaged groups:

* Attempts should be made to minimise the potential impact of language as a confounding factor by using non-verbal instruments.
• The role of previous learning within the assessment process needs to be reduced. Tests like the APIL Battery and the APM attempt to reduce this by incorporating a learning component into the administration of the tests, i.e. candidates are given practice opportunities before commencing with the test (Bors & Stokes, 1998; Raven & Court, 1998; Taylor, 1997).

• Attempts should be made to limit culture-specific test content and speed as factors within the assessments.

• If foreign instruments are used, appropriate research should be conducted to test for cultural bias (Society for Industrial Psychology, 1998).

• Appropriate norms for the various population groups should be generated and consulted (Society for Industrial Psychology, 1998).

• If possible, attempts should be made to ensure that the test administrator and the candidates are from the same cultural background. Research has shown that giving test instructions in both English and the candidates' home language produces positive results (Owen & Taljaard, 1996).

• Tests should be administered and explained to candidates in a culturally sensitive manner to ensure their understanding of what is expected of them (Society for Industrial Psychology, 1998).

Both the APIL Battery and the Raven's APM are considered to be culture-fair as both attempt to diminish the effects of previous learning and language on performance on the tests (Bors & Stokes, 1998; Raven & Court, 1998; Taylor, 1997).

3.6.3 Emotional intelligence

Within the field of psychology the term emotional intelligence (EI) has become increasingly
Emotional intelligence is defined in a number of ways. For this purpose, it is defined as “the ability to understand and relate to others” (Ostrowski, DiGiacomo, Medaglia, & Ortiz, 1999, p. 1). This definition emphasises being in touch with one’s emotions as well as the ability to monitor them. A broader definition is provided in Tucker, Sojka, Barone, and McCarthy (2000, p. 331) where EI is defined as “the ability to monitor one’s own and others’ emotions, to discriminate among them, and to use the information to guide one’s thinking and actions”.

Goleman (1998), the person responsible for popularising this concept, states that EI is a far better and more reliable predictor of success than IQ. According to Goleman (1998) it entails five aspects or competencies. The first is self-awareness defined as the ability to “understand and recognize your own behavioral and emotional patterns and moods” (Ostrowski et al., 1999, p. 3). A person measuring high on self-awareness is capable of controlling her/his emotions in order to positively impact on others, views her/himself realistically and possesses self-acceptance. Secondly, self-regulation is described as the ability to monitor and control one’s emotions so that it may have productive consequences. It helps with being reasonable, calm, thoughtful, reflective and adaptive to change. The third component of emotional intelligence is motivation which is defined as the ability to set and strive for attainable goals. The first three aspects are collectively termed personal competencies. Fourthly, empathy entails attempting to truly understand and feel compassion for others. Lastly, social skill is the culmination of all the other components of emotional intelligence. The latter two aspects are clustered as social competencies. Goleman (1998) is further of the opinion that unlike many other human characteristics, like intellectual ability, which stabilises in childhood, EI can be developed throughout life. It should and can be developed on a daily basis (Goleman, 1998).
Differing views exist however. For instance, Conger (1993) and Tossman (1999) share the view that EI is relatively stable in adults and can only be changed as a result of major life experiences. They further state that the changes brought about during EI training are not enduring, tending to last only a few months. Further criticism entails the fact that the so-called “big five” personality characteristics, i.e. conscientiousness, extroversion, openness to experience, emotional stability and agreeableness, have long been recognised and researched. EI is therefore nothing but the proverbial old wine in a new bottle. Lastly, not all cultures may be accepting of public displays of (some of) the EI competencies; EI may therefore not be such a good predictor of success in these circumstances (Conger, 1993; Tossman, 1999).

According to Johnson and Indvik (1999), techniques to enhance EI can only be effective, if they encourage participants to have a desire to change, to use self−reflection, to listen to the continuously playing internal script, to develop emotional control, to practice empathy and active listening skills and to validate the emotions of others.

3.7 SUMMARY

The above discussion by no means assumes to provide an exhaustive list of all the developments within the field of intelligence testing. For example, in the interest of brevity and economy, developmental theories as well as cognitive psychological theories were not discussed at length in this chapter. The primary focus would be the major contributors to and developments in the field. Also, specific emphasis is placed on those developments that stimulated the development of the two independent variables (psychometric instruments) under investigation.

In summary, it should be stated that although divergent in terms of approach, all the above theories and models have contributed to our current understanding of the construct intelligence and the identification of its dimensions. The Binet−Simon emphasis on judgement and reasoning, Spearman’s education of correlates and relations, and Cattell’s fluid and crystallised intelligence, form the basis of most intelligence tests currently in use. The multiple factor approach has led to the realisation that intelligence is not a single, unitary ability but rather a constellation of various abilities such as verbal and numerical abilities.
As was pointed out earlier, this suggestion has had a very important impact on occupational and educational selection and development. Especially Cattell’s expansion of intelligence to include learning ability has very positive implications for the assessment of the potential of previously disadvantaged individuals.

The existence of a wide array of opinions regarding the nature and measurement of intelligence, has important implications in terms of the use of intelligence tests. Different intelligence tests emphasise different components of intelligence. Therefore, although a wide variety of psychometric tests are grouped under the phrase intelligence, they may not measure exactly the same thing. As a result test users need to be very circumspect when selecting and interpreting a specific test, ensuring that they are completely aware of the definition of intelligence that was employed by the test developer. The onus rests on the test user to ascertain exactly how intelligence is defined within a specific test. I/O Psychologists therefore need to apply their minds in the context of psychometric assessment as it is essential to the appropriate and responsible assessment of intelligence in industry.

Furthermore, considerable attention is devoted to the development of culture – fair assessments of intelligence. Within the South African context, this aspect is considered crucial in affecting social justice. As was mentioned in Chapter 1, the primary aim of the university bridging course that serves as the context for this investigation, is aimed at achieving social justice in education. Therefore, continued efforts need to be made to ensure that intelligence testing do not hinder candidates from disadvantaged backgrounds from entering into academic institutions.

Finally, emotional intelligence as a possible mitigating factor has been discussed. Considering the above discussion of EI as well as its detractors, the conclusion may be drawn that it is a concept well worth exploring within the field of selection for academic development programmes. This may be especially relevant in terms of the academic development, as well as the possible lack thereof, of disadvantaged groupings. Having stated this however, the empirical investigation of this concept and its relevance fall outside the stated scope of this research project.
This chapter serves as a broad overview of developments within the field of intelligence especially regarding its measurement and predictive properties. Research has shown that measures of intelligence do have predictive validity regarding the prediction of academic achievement. Within the context of apparent cultural bias, it has however, been pointed out that a number of mitigating factors exist which may alter the relationship between intelligence and academic achievement. The often-vitriolic debate that exists around these issues, often does not question the necessity of assessing intelligence, but rather the manner in which it is currently being assessed. The following chapter, highlights attempts to eradicate or control for these mitigating factors within the research methodology followed in this research project.
CHAPTER 4 RESEARCH METHODOLOGY

4.1 INTRODUCTION

This chapter entails a discussion on the methodology followed in the empirical investigation. The characteristics of the population and sample are defined and discussed. As a logical extension of chapter 3, the psychometric properties of the APIL Battery and Raven's APM, the rationale for its inclusion, as well as the reliability and validity of these instruments, are attended to. Furthermore, the selection and relevance of the criterion measure are discussed. Research hypotheses, based on the research aims, are also presented. Finally, the steps followed in gathering and analysing data regarding the predictor and criterion variables, are discussed.

4.2 PURPOSE OF THE STUDY

The purpose of this investigation is the validation of a selection battery used in the selection of university - bridging course students. Of pivotal importance here is the compilation of a scientifically and legally defensible selection battery. To this end, the relationship between various psychometric tests and academic performance needs to be determined.

4.3 DESCRIPTION OF THE POPULATION AND SAMPLE

The population from which the validation sample was drawn is black scholars who have passed final year matriculation examinations but were unable to enter university due to the fact that they did not attain university exemption. The aim of the bridging programme, as discussed in chapter 1, is to provide these students with a conducive environment in which they will have the opportunity to improve on their final year matriculation marks which would assist them in gaining university entrance. This sample represents the first group of individuals who participated in this programme. As the sample included all applicants to the programme, it may be typified as a convenience or purposive sample. Babbie (1992, p. 207)
defines purposive sampling as "a type of non-probability sampling in which the researcher uses his or her own judgement in the selection of sampling members".

The sample consists of 167 participants and includes candidates from across South Africa as illustrated in Table 4.1. Research has shown that biographical variables play a significant role in academic achievement. Therefore, data regarding gender, ethnicity and home language were collected for comparative purposes. Research conducted by, for example, Owen (1991), Van Rooyen (2001) as well as Viljoen and Vos (1998), found that these factors significantly affect academic achievement. Although subjects' age is routinely incorporated into research of this nature, it was not the case here. Data regarding age was not collected because of previous research findings, which indicated that it does not play a significant role in terms of academic achievement in university bridging course. For example, internationally, Astin (1976) could not find a significant and direct relationship between age and academic performance within the American context. South African research conducted by Van Rooyen (2001) as well as Viljoen and Vos (1998), also did not produce significantly different academic results between students of different age groups.

Table 4.1  GEOGRAPHIC DISTRIBUTION OF THE SAMPLE

<table>
<thead>
<tr>
<th>GEOGRAPHIC ORIGIN</th>
<th>FREQUENCY</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Cape</td>
<td>24</td>
<td>14.34%</td>
</tr>
<tr>
<td>East London</td>
<td>6</td>
<td>3.59%</td>
</tr>
<tr>
<td>Eastern Transvaal</td>
<td>32</td>
<td>19.16%</td>
</tr>
<tr>
<td>Kwa – Zulu Natal</td>
<td>44</td>
<td>26.35%</td>
</tr>
<tr>
<td>Pretoria</td>
<td>29</td>
<td>17.90%</td>
</tr>
<tr>
<td>Potchefstroom</td>
<td>32</td>
<td>19.16%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>167</td>
<td>100%</td>
</tr>
</tbody>
</table>

In terms of gender, the sample consists of 127 males and 40 females (see Table 4.2). Furthermore, although the encompassing term Black was used to select candidates, ethnic
origin was also considered in order to investigate its influence (see Table 4.3). The following tables indicate the distribution in terms of gender, ethnicity and home language.

Table 4.2  GENDER DISTRIBUTION OF THE SAMPLE

<table>
<thead>
<tr>
<th>GENDER</th>
<th>FREQUENCY</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>127</td>
<td>77.0%</td>
</tr>
<tr>
<td>Female</td>
<td>40</td>
<td>23.0%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>167</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 4.3  ETHNIC DISTRIBUTION OF THE SAMPLE

<table>
<thead>
<tr>
<th>ETHNICITY</th>
<th>FREQUENCY</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>African</td>
<td>148</td>
<td>88.7%</td>
</tr>
<tr>
<td>Coloured</td>
<td>14</td>
<td>8.3%</td>
</tr>
<tr>
<td>Indians</td>
<td>5</td>
<td>3.0%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>167</td>
<td>100%</td>
</tr>
</tbody>
</table>
### Table 4.4  HOME LANGUAGE DISTRIBUTION OF THE SAMPLE

<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>FREQUENCY</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IsiXhosa</td>
<td>29</td>
<td>17.9%</td>
</tr>
<tr>
<td>Venda</td>
<td>5</td>
<td>2.9%</td>
</tr>
<tr>
<td>IsiZulu</td>
<td>39</td>
<td>23.3%</td>
</tr>
<tr>
<td>Setswana</td>
<td>38</td>
<td>22.7%</td>
</tr>
<tr>
<td>Sesotho</td>
<td>8</td>
<td>4.7%</td>
</tr>
<tr>
<td>English</td>
<td>7</td>
<td>4.1%</td>
</tr>
<tr>
<td>Afrikaans</td>
<td>13</td>
<td>7.7%</td>
</tr>
<tr>
<td>Sepedi</td>
<td>3</td>
<td>1.7%</td>
</tr>
<tr>
<td>IsiSiswati</td>
<td>25</td>
<td>14.9%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>167</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

#### 4.4 MEASURING INSTRUMENTS

As was discussed in chapter 3, cognitively - loaded psychometric tests are often used in selection for academic purposes because they tend to be the most valid predictors available (Sackett, Schmitt, Ellingson & Kabin, 2001). These findings serve as the rationale for the inclusion of the APIL Battery and Raven's APM in this selection battery. Furthermore, cost factors were taken into account considering that the financial implications of conducting a selection process are pivotal (Cascio, 1998). Research has shown that both instruments are reliable, valid and relatively culturally fair and unbiased measures of cognitive ability within an academic context (Arthur & Day, 1994; Assessment and Development Services, 2001; Bors & Stokes, 1998; Lopes et al, 2001; Raven & Court, 1998; Taylor, 1997; Viljoen & Vos, 1998). Both tests being investigated are considered to be relatively inexpensive and cost effective. In terms of administration time, Raven's APM takes approximately 40 to 60 minutes to administer (Arthur & Day, 1994; Bors & Stokes, 1998; Raven & Court, 1998). It is therefore more economical in contrast to the APIL Battery, which often takes approximately 4 to 5 hours to administer. Despite its lengthy administration time, it was
included, as the entire battery does not have to be used. If shown to be valid predictors, individual subtests may be used separately (Taylor, 1997). Lastly, the main motivating factor for the inclusion of these tests was the fact that they are available in the organisation. Following is a discussion of the psychometric properties of the two tests under investigation.

4.4.1 The Abilities, Processing of Information and Learning (APIL) Battery

The APIL Battery is "a set of tests designed to assess an individual's core or fundamental cognitive abilities and potentialities" (Taylor, 1997, p. 1). It represents a movement away from specific skills assessment, which to a great extent reflects the socio-economic and educational history of a person. Instead, the battery aims at identifying those individuals that are likely to master certain skills in future if provided with proper training and/or education (Taylor, 1997).

Reliability figures vary between 0,82 and 0,92 for the various subtests. Its validity in various contexts has also been established, ranging from 0,21 to 0,89 (Lopes, et al, 2001; Taylor, 1997). In terms of its prediction of MBA results in Gauteng, it has shown to correlate significantly (p< 0,05) with academic performance. Correlations range between 0,30 and 0,59 (Assessment and Development Services, 2001). Furthermore "statistical analysis has shown no evidence of predictive bias" (Assessment and Development Services, 2001, p. 1).

4.4.1.1 The aim of the test

The APIL Battery is designed to measure a person's underlying ability to learn new, cognitively challenging skills. It assesses an individual's fundamental cognitive capabilities and potentialities (Taylor, 1997). It is considered useful in identifying candidates who are likely to master new cognitively demanding material in a formal training context, especially with the view of identifying individuals with potential for tertiary level development. It provides a profile of 8 scores and a learning curve. These are fundamental building blocks of intellectual competence - the cognitive materials needed to develop skills required to master new challenges. It further indicates the individual's intellectual adaptability rather than his or
her already acquired skills or abilities. A computerised report with stanines and stens is produced (Taylor, 1997).

4.4.1.2 Description of the constructs measured by the APIL Battery

The APIL Battery consists of tests which are administered consecutively. The constructs measured are as follows:

♦ The Concept Formation Test

This test measures an individual's capacity to think abstractly and conceptually. This entails the ability to reason in a hypothetical manner, to deal with abstract concepts, theorise, build scenarios and trace causes. This ability is especially needed when higher level managerial skills and technical expertise are required. It furthermore plays a crucial role within an education and training environment (Taylor, 1997). Concept formation is similar to Cattell's fluid intelligence as described in chapter 3, which is defined as the ability "to form concepts, reason and identify similarities; it is more intuitive and is active in forming new mental structures" (Baron, 1995, p. 417).

♦ The Flexibility - Accuracy - Speed Test

The APIL Battery manual (Taylor, 1997) describes this test as a battery within a battery as it measures intellectual speed, accuracy of information processing and cognitive flexibility. In terms of all of the above - mentioned components, scores are expressed as stens. Stens 1 - 3 are indicative of mental "slowness, inaccuracy and/or low flexibility", whereas stens between 8 - 10 indicates mental "quickness, high accuracy and/or flexibility" (Taylor, 1997, p. 5).

♦ The Curve of Learning

Taylor (1997) is of the opinion that conventional cognitive ability tests assume that intelligence stabilises at a given age. Such an assumption may therefore render such measures
inflexible and unfair. The author disagrees with this stance, stating that "this assumption is highly questionable in a society where great discrepancies in opportunity have existed and continue to exist" (Taylor, 1997, p. 8). This argument led to the establishment of a measure of the curve of learning (COL) of an individual on both practical (manpower utilisation) and ethical (fairness in opportunity) grounds. In essence, the COL test measures the individual's learning potential or future achievement capacity. The testee is presented with a novel task and given four sessions in which to master it. In this manner, learning speed is assessed by measuring the gains or improvement that has occurred over the four sessions. It is important to note that learning or gain is measured against a baseline that has been established by the testee and not an external source (Taylor, 1997).

♦ The Memory Test

Immediately after completing the COL test, the memory test is administered based on the COL material. The individual's memory is tested on the symbols contained in the COL test. A high score on the memory test indicates that the testee has processed the information contained in the COL on a deeper level (Taylor, 1997).

♦ The Knowledge Transfer Test

The Knowledge Transfer Test measures the individual's ability to apply and adapt knowledge. This ability is particularly relevant within an education and/or training setting. It is also useful in a work setting where individuals are expected to transfer experience gained in one context, to another. This test measures the degree of knowledge transfer by exposing testees to a series of related and increasingly complex problems.

♦ Primary and Secondary Scores

APIL scores are divided into primary and secondary scores. The latter are derived from the primary scores. Primary scores are determined by adding correct answers using scoring masks. These primary (raw) scores attained by participants on the subtests were utilised for
the statistical analyses in this investigation. Secondary scores are calculated by means of computer software. A computerised report is generated which provides indicators of the following:

* Speed;
* Accuracy;
* Cognitive flexibility;
* The COL improvement scores; and
* The COL total output scores.

4.4.1.3 Administration

The test battery is administered in the sequence of constructs discussed in the previous section. This is motivated by the fact that some of the tests build directly onto a previous test. For example, the memory test makes use of material used in the COL test (Taylor, 1997).

4.4.2 Raven's Advanced Progressive Matrices (APM) Test

Raven's APM is described as a $g$-saturated test, i.e. it is a measure of general intelligence. Factor analytical studies conducted by Alderton and Larson (1990) and Arthur and Woehr (1993), confirmed that Raven's APM is a single – factor measure of Spearman’s $g$. The test consists of a series of perceptual analytical reasoning problems that are presented in a matrix format. Following, is a discussion of the aims, description and administration of Raven's APM (Alderton & Larson, 1990; Arthur & Woehr, 1993; Bors & Stokes, 1998).

4.4.2.1 The aim of the test

Raven's APM is a non-verbal test of an individual's capacity to comprehend figures presented for her/his perception, see the relations between them, and conceive the correlative figures completing the system of relations presented. It is used for people over 11 years of age and of average or above average intellectual ability. Raven's APM is designed to assess a
person's intellectual "efficiency". This refers to the speed with which the individual can perform certain tasks in an accurate manner. Within a selection context, knowing a person's intellectual efficiency is quite useful in terms of jobs that would require quick and accurate judgement. Within a clinical setting the Raven's APM is quite useful in order to determine mental retardation (Bors & Stokes, 1998; Raven & Court, 1998).

In order to measure intellectual "efficiency", the test is administered within a specified time. When it is administered without attaching a time limit to it, it gives an indication of the individual's intellectual "capacity" (Raven & Court, 1998). As this test is designed to differentiate between people of superior intellectual ability, it is often used to select staff for high-level technical or managerial positions (Bors & Stokes, 1998; Raven & Court, 1998).

4.4.2.2 Description and administration

Raven's APM consists of Sets I and II. In the first set there are 12 problems. They are designed to introduce a person to the method of working and to familiarise candidates with the manner in which the test should be completed. For these reasons, set I is normally administered without attaching a time limit and is not scored. In the second set there are 36 problems. These items are arranged in an increasing order of difficulty. Set II is normally completed within a time limitation of 45 minutes. A mask is used to score Set II and a single score is obtained (Bors & Stokes, 1998; Raven & Court, 1998). In this investigation, this score was used in the subsequent data analysis.

Research has shown that Raven's APM is equally difficult for people from different socio-economic and racial backgrounds. This implies that the test works the same way for people from different backgrounds. It would therefore seem to be "culture-fair". This aspect, as well as the relative ease with which it is administered accounts for much of its popularity (Arthur & Day, 1994; Bors & Stokes, 1998; Raven & Court, 1998; Viljoen & Vos, 1998).

The validity and reliability of Raven's APM are well established (Arthur & Day, 1994; Bors & Stokes, 1998; Raven & Court, 1998; Viljoen & Vos, 1998). Raven's APM has been found
to correlate 0.74 with the full Wechsler Adult Intelligence Scale and 0.75 with the Otis I.Q.
test. It has furthermore shown substantial internal consistency, with split - half reliabilities
ranging from 0.8 and 0.9 (Alderton & Larson, 1990; Arthur & Day, 1994; Bors & Stokes, 1998). According to Muller (2002) Raven’s APM declared 32.1 percent of the observed variance in academic performance. Several other South African researchers, for example, Viljoen and Vos (1998) have also found that Raven’s APM manifests sufficient validity and reliability to be used, with some degree of confidence, within the South African academic context.

4.5 CRITERION DESCRIPTION

A distinction can be drawn between the true or conceptual criterion and the available or operational measure of the criterion. The conceptual criterion refers to a standard in terms of which an individual’s behaviour must be evaluated as successful or unsuccessful. In terms of academic performance, the choice of criterion should proceed with a great deal of circumspection in order to prevent considerable adverse impact (Cascio, 1998; Huysamen, 1998; Sackett et al, 2001). Sackett et al (2001, p. 315) emphasise that it “is important to measure what is relevant, not what is convenient, easy or cheap”.

The dependent variable or criterion used in this research is academic achievement at the end of the bridging programme. Robinson (as cited in Van Rooyen & Huysamen, 2000, p. 93) states that “academic performance is clearly the sine qua non for the validation of remedial courses” and by extension the selection instruments used. Robinson (as cited in Van Rooyen & Huysamen, 2000, p. 93) further states that, “in the final analysis, remedial instruction must necessarily stand or fall on the basis of this single criterion”. At the heart of this investigation is the determination of the predictive validity of a selection battery used to predict success in the university bridging - course. Success, in this context, implies academic improvement in a final year supplementary matriculation examination. Although matriculation results have long been questioned as a valid predictor of success at university, it remains the most widely used criterion used by universities to determine entrance (Curtis & De Villiers, 1992; Webb & Erwee, 1990). With due consideration of these factors, it therefore logically follows that the
most plausible criterion measure would be academic performance as measured by matriculation results. Considering this state of affairs, it would be quite ineffectual to select any other criterion measure, bearing in mind that the university bridging – course that is considered here, is ultimately aimed at preparing candidates for university selection.

4.6 DATA COLLECTION REGARDING INDEPENDENT VARIABLES

Huysamen (1998) states that in order to investigate the predictive validity of psychometric tests, they first need to be administered to a large representative sample of the intended population. Scores obtained in this manner should not be used in decision-making regarding participants. The APIL Battery and Raven’s APM were administered to participants before the start of the programme according to the procedures prescribed in their respective manuals (Raven & Court, 1998; Taylor, 1997). In keeping with the dictates of validation studies individuals were not selected based on their results in the various selection instruments (Huysamen, 1998; Society for Industrial Psychology, 1998; Tabachnick & Fidell, 1996). Instead, other criteria, such as race and marks obtained in their previous matriculation examinations, were used for selection purposes. This was mainly due to the fact that the selection battery used has not yet been shown to be a valid predictor of performance in the bridging programme. The organisation determined that only black matriculants who have not achieved matriculation exemption previously, were to be admitted to the programme.

4.7 DATA COLLECTION REGARDING THE DEPENDENT VARIABLE

Data regarding participants’ performance in Mathematics (n = 137), Science (n = 137) and Biology (n = 47) as well as the average mark obtained in a final year supplementary matriculation examination, was collected upon completion of the university bridging – course. These results were obtained from the training institution they attended.
4.8 RESEARCH HYPOTHESES

Research hypotheses have been formulated in accordance with the six empirical aims stated in chapter 1. Babbie (1992, p. G3) defines a hypothesis as "a conjectural statement of something that ought to be observed in the real world if the theory is correct". Rosnow and Rosenthal (1996) state that within the behavioural sciences, a hypothesis needs to conform to three criteria, i.e. correspondence with reality; coherence and parsimony; and falsifiability.

Tabachnick and Fidell (1996) state that in order to get a holistic picture of the function of independent variables in a regression it is important to compare:

- The total effect of the independent variables (the total selection battery) with the dependent variable (average academic performance),

- The unique relationship of every independent variable (the psychometric tests individually) to the dependent variable, and

- Correlations of the independent variables with each other.

As far as possible, the first three (core) hypotheses were designed to meet all these requirements. However, in terms of the last requirement, the various psychometric tests and its subtests were merely screened for multicollinearity and singularity. Multicollinearity refers to a situation where various subtests may be too highly correlated (0.90 and above). Singularity refers to redundancy in variables, i.e. one of the variables is a combination of two or more other variables (Tabachnick & Fidell, 1996). No such cases were found and therefore a separate hypothesis would not need to be formulated in this regard. The last three hypotheses, deal with the effect of biographic factors on performance.

In the following hypothetical statements, performance denotes the average performance achieved by individual participants and the total selection battery refers to the combined effect of the APIL Battery and Raven's APM.
• H1₀: The APIL Battery does not predict academic performance in a final year supplementary matriculation examination.

• H1₁: The APIL Battery predicts academic performance in a final year supplementary matriculation examination.

• H2₀: Raven's APM does not predict academic performance in a final year supplementary matriculation examination.

• H2₁: Raven's APM predicts academic performance in a final year supplementary matriculation examination.

• H₃₀: The total selection battery under investigation does not declare the variance observed in academic performance in a final year supplementary matriculation examination.

• H₃₁: The total selection battery under investigation declares the variance observed in academic performance in a final year supplementary matriculation examination.

• H₄₀: No significant differences in academic performance in a final year supplementary matriculation examination between male and female candidates exist.

• H₄₁: Significant differences in academic performance in a final year supplementary matriculation examination between male and female candidates exist.

• H₅₀: No significant differences in academic performance in a final year supplementary matriculation examination between candidates from different ethnic groups exist.

• H₅₁: Significant differences in academic performance in a final year supplementary matriculation examination between candidates from different ethnic groups exist.
- H$_{60}$: No significant differences in academic performance in a final year supplementary matriculation examination between candidates from different language groups exist.

- H$_{61}$: Significant differences in academic performance in a final year supplementary matriculation examination between candidates from different language groups exist.

4.9 STATISTICAL ANALYSIS

For the purposes of this study, the SPSS package was used to analyse the data. The predictive validity of the above mentioned selection battery was determined. Predictive validity refers to the accuracy with which a test or instrument enables the researcher to predict some future behaviour or status of individuals (Huysamen, 1994). The purpose of this type of validity is summarised in its definition. It is often used to determine the effectiveness of selection batteries in the prediction of success in training or on the job. However, it is imperative that the criterion measures used are valid measures of the criteria themselves (Cascio, 1998). Data will also be statistically manipulated by means of, amongst other techniques, correlations and regression analysis.

4.9.1 Correlation coefficient

Correlation deals with measuring the relationship or strength of association among sets of variables. It implies an inter - relationship or correspondence between two variables. It is expressed as a single index ranging from -1 to +1. This indicates the direction and degree of correspondence or co - variance between variables (Babbie, 1992; Huysamen, 1998; Pietersen & Damianov, 2001).

Further reasons for using correlations are that:

- The correlation coefficient is a familiar concept within the scientific community;

- It is a convenient way of summarising a relationship between variables;
• It is less susceptible to sampling fluctuations;

• It can be directly linked to predictive accuracy by using $R^2$ (Babbie, 1992; Huysamen, 1998; Pietersen & Damianov, 2001).

In order to determine the correlation between the variables relevant to this investigation, Pearson's product-moment correlation ($R$) is determined. According to Huysamen (1998, p. 35), "the higher the validity coefficient, the better the agreement between decisions based on individuals' test scores and those that would have been made in terms of their criterion scores, had the latter been known". In order to predict the precise criterion scores of individuals, a validity coefficient of 0.8 is required. Huysamen (1998) however states that, within the selection context, coefficients as low as 0.2 and 0.3 are acceptable if there are few vacancies relative to the number of applicants.

In order to determine the percent of variance in the dependent variable that is explained by the independent variables, rather than other unknown factors, the coefficient of determination ($R^2$) is calculated. The determination of $R^2$ is critical, considering the fact that "if the validity coefficient is equal to 0.70 only about one half ($0.70^2 = 0.49$) of the criterion-score variance can be attributed to test-score variance" (Huysamen, 1998, p. 35).

Furthermore, due to the number of independent variables involved, the calculation of the adjusted $R^2$ is also indicated. The validity coefficient differs from $R^2$ depending on the number of independent variables involved (Huysamen, 1998; Pietersen & Damianov, 2001).

4.9.2 Regression analysis

Regression deals with prediction that is the ability to build a statistical model which uses information about a set of independent or predictor variables in order to estimate the expected value of some dependent or response variable. Univariate and multiple regression analysis are used in order to determine the nature of the relationship between variables. Simply put, it
assists the researcher in empirically determining the most effective predictor or set of predictors for a specific criterion (Howitt & Cramer, 2000). Essentially, it answers the question: How does one variable (e.g., academic performance) depend on other variables (e.g., psychometric tests) (Pietersen & Damianov, 2001).

4.9.3 Comparisons of independent groups

The T-test is used in order to determine differences that may exist between gender groups. T-tests are used to compare two independent groups in terms of their scores measured against a criterion (Pietersen & Damianov, 2001). Also, differences between various ethnic as well as home language groups, respectively are investigated by means of a one-way analysis of variance (ANOVA). Considering the fact that some of these groups are relatively small in size, Scheffe’s technique was used (Tabachnick & Fidell, 1996).

4.10 SUMMARY

In this chapter, the characteristics of the population and sample were discussed and the distribution schematically illustrated in terms of gender, ethnicity as well as home language. The psychometric properties of the APIL Battery and Raven’s APM, procedures followed for administering them, as well as the manner in which data regarding these two tests was collected, were discussed. Furthermore, the various research hypotheses were stated. These hypotheses are based on the research aims discussed in chapter 1. Lastly, the statistical procedures used to analyse data were discussed.

This chapter serves as a blueprint for the empirical component of this research effort. The following chapter will deal with the reporting and discussion of research results, research limitations and recommendations emanating from the former.
CHAPTER 5 RESEARCH RESULTS, CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

5.1 INTRODUCTION

In this chapter, the results of the empirical study are presented and discussed. Results are reported based on the six hypothetical statements presented in chapter 4. The nature of the results facilitates the rejection of three out of the six null hypotheses. Based on the discussion of results, various conclusions pertaining to the predictive validity of the selection battery under investigation are made. Conclusions are informed by the literature review conducted in conjunction with the results of the empirical investigation. With due consideration of the practical constraints inherent in a study of this nature, research limitations are discussed. Finally, recommendations regarding the literature review, empirical study, future recruitment and selection efforts, as well as future training are suggested.

5.2 RESEARCH RESULTS

Research results are based on statistical analysis as discussed in chapter 4. Results are first presented, significance levels (p) indicated and then discussed. Although the stated hypotheses only deal with average academic performance, results are also presented in terms of the specific marks obtained in Biology, Science and Mathematics. As was stated in chapter 4, this was done in order to enrich the conclusions that can be drawn from the results of the empirical study. Furthermore, only those results that are statistically significant are presented in table form. The observed significance levels often referred to as the p-value, facilitates or negates the rejection of the null hypotheses. If the observed significance level is sufficiently small (normally less than 0.01 or 0.05), the null hypothesis is rejected (Babbie, 1992; Huysamen, 1998; Rosnow & Rosenthal, 1996).
5.2.1 Results and discussion regarding the predictive power of the APIL Battery

The first set of hypotheses (H_0 and H_1) deals with the predictive validity of the APIL Battery in terms of the average mark obtained by individuals (see Table 5.1 below). Results are presented in terms of the coefficient of determination (R^2) as well as the adjusted R^2. The levels of significance (p) regarding all the stated correlations are also indicated.

<table>
<thead>
<tr>
<th>TABLE 5.1</th>
<th>PREDICTIVE POWER OF THE APIL BATTERY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R^2</td>
</tr>
<tr>
<td>AVERAGE MARK OBTAINED</td>
<td>0.352</td>
</tr>
<tr>
<td>Biology</td>
<td>0.504</td>
</tr>
<tr>
<td>Science</td>
<td>0.186</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0.347</td>
</tr>
</tbody>
</table>

As reported in Table 5.1, all the above results are statistically significant at the 1 percent significance level (p < 0.01). The predictive power of the APIL Battery is lowest regarding academic performance in Science relative to that of the other academic performance measures. The APIL Battery's ability to predict academic performance in Biology (37.1 percent) is higher than for Science (11.5 percent) and Mathematics (28.9 percent). Considering the nature and aim of the APIL Battery, as discussed in chapter 4, this result is to be expected. Although the APIL Battery is a g - loaded test, it also includes extensive memory testing (Taylor, 1997). As academic performance in Biology relies more heavily on the capacity to remember than Mathematics or Science, these findings therefore make logical sense. Due to the fact that there are relatively few participants who wrote Biology examinations (n = 47), these results need to be interpreted with care as they may have been artificially inflated.

The APIL Battery declares 29.4 percent of the variance observed in average academic performance. All the above findings fall within the range of the MBA study (see chapter 4) reported by Assessment and Development Services (2001), which found that the APIL Battery declares between 9 percent and 34.8 percent of the variance in academic performance. The
correlations of Mathematics and average scores are quite similar. With the exception of Biology, these findings are slightly lower than those reported by Lopes et al (2001) in their study. In the aforementioned study, the APIL Battery declares 36.6 percent of variance in academic performance. Although the declared variance may seem low at face value it should be borne in mind that normally, cognitive measures only declare approximately 9 percent of observed variance (Lopes et al, 2001). Also, Huysamen's (1998) statement that, within the selection context, percentages as low as 4 percent and 9 percent are acceptable if there are few vacancies relative to the number of applicants, needs to be considered.

These results facilitate the rejection of H10. Results regarding the correlation between average academic performance and the APIL Battery are statistically significant ($p = 0.000 < 0.01$). This implies the existence of a linear relationship between the APIL Battery and average academic performance. The rejection of H10 is therefore justified, which implies that the APIL Battery does indeed have predictive power in terms of academic performance.

5.2.2 Results and discussion regarding the predictive power of Raven's APM

The second set of hypotheses (H20 and H21), deals with the predictive validity of Raven's APM regarding the average mark obtained by individuals (see Table 5.2 below). Results are presented in terms of the coefficient of determination ($R^2$) as well as the adjusted $R^2$. Both are presented in terms of marks obtained in Biology, Science, Mathematics and the individual participant's average mark, respectively. The levels of significance ($p$) regarding all the stated correlations are also indicated.
As reported in Table 5.2, the predictive power of Raven’s APM is lower in terms of all of the above performance criteria. Raven’s APM’s ability to predict academic performance in Mathematics and overall academic performance is high relative to the other academic performance indicators. Raven’s APM’s ability to predict Biology is the lowest (5.3 percent). As is the case regarding the APIL Battery, Raven’s APM declares approximately the same amount of variance for Mathematics (11.0 percent) and average academic performance (11.5 percent). The fact that Mathematics as well as overall academic performance relies less heavily on memory than academic performance in Biology, could potentially explain the correlational pattern observed.

It supports the previously mentioned assertion (see chapters 3 and 4) that Raven’s APM is one of the best measures of Spearman’s g – factor (Arthur & Woehr, 1993; Bors & Stokes, 1998; Raven & Court, 1998). Although the above correlations are very low, they are all significant (p < 0.01), except for Biology (p = 0.065 > 0.05). The latter can be explained by the fact that the number of participants who completed the final year Biology examination are relatively low (n = 47). This state of affairs may serve as an alternative explanation for the correlational pattern observed.

All the above-mentioned correlations are lower than that of the APIL Battery. Once again the “9 percent - rule”, as mentioned by Lopes et al (2001) as well as the cautionary note by Huysamen (1998) need to be applied in terms of interpreting these findings. In contrast to the APIL Battery, Raven’s APM only declares 11.5 percent of the observed variance in academic performance. This

<table>
<thead>
<tr>
<th>TABLE 5.2 PREDICTIVE POWER OF RAVEN’S ADVANCED PROGRESSIVE MATRICES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R</strong></td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>AVERAGE MARK OBTAINED</td>
</tr>
<tr>
<td>Biology</td>
</tr>
<tr>
<td>Science</td>
</tr>
<tr>
<td>Mathematics</td>
</tr>
</tbody>
</table>
renders Raven's APM much less effective in predicting academic performance than the APIL Battery in this context. These findings are lower than research results reported by Viljoen and Vos (1998) and Muller (2002) which indicated that Raven's APM declares 31.2 percent and 32.1 percent respectively of the perceived variance in academic performance.

Despite the above statement, results regarding the correlation between average academic performance and Raven's APM are statistically significant \( (p < 0.01) \). These findings point to the existence of a linear relationship between Raven's APM and average academic performance. Therefore, the rejection of \( H_2_0 \) is justified. This implies that although Raven's APM is not as effective as the APIL Battery, it may still be considered to have predictive validity in terms of academic performance.

5.2.3 Results and discussion regarding the predictive power of the total selection battery

The third set of hypotheses (\( H_3_0 \) and \( H_3_1 \)), deals with the predictive validity of the total selection battery regarding the average mark obtained by individuals (see Table 5.3 below). Results are presented in terms of the coefficient of determination \( (R^2) \) as well as the adjusted \( R^2 \). Results are presented in terms of marks obtained in Biology, Science, Mathematics and the individual participants' average mark, respectively. The levels of significance \( (p) \) regarding all the stated correlations are also indicated.

<table>
<thead>
<tr>
<th>TABLE 5.3 PREDICTIVE POWER OF THE TOTAL SELECTION BATTERY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Mark Obtained</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Average Mark Obtained</td>
</tr>
<tr>
<td>Biology</td>
</tr>
<tr>
<td>Science</td>
</tr>
<tr>
<td>Mathematics</td>
</tr>
</tbody>
</table>
As reported in Table 5.3 the predictive power of the total selection battery regarding academic performance in Science is low compared to the other academic performance criteria. The total selection battery’s ability to predict academic performance in Biology (46.1 percent) is very high relative to the other performance indicators. Once again, caution needs to be applied when interpreting correlations involving Biology. Although the other correlations are very low, they are all significant ($p < 0.01$), except for Science ($p = 0.137 > 0.05$). The correlations of Mathematics and average mark are quite similar. All the above-mentioned correlations, except for Biology, are lower than that of the APIL Battery.

The total selection battery declares 27.1 percent of the observed variance in academic performance. Considering the fact that in isolation, the APIL Battery is a better predictor than the total selection battery, it would seem that the Raven’s APM does not add much value to the predictive efficiency of the selection battery.

Results regarding the correlation between average academic performance and the total selection battery are statistically significant and therefore the rejection of $H_{30}$ is justified. This is due to the fact that a linear relationship between the total selection battery and average academic performance does exist ($p < 0.01$). Considering the views of Huysamen (1998) and Lopes (2001) regarding what constitutes an acceptable correlation, it would seem that the total selection battery under investigation is a valid predictor of academic performance.

5.2.4 Results and discussion regarding the effect of gender on academic performance

The fourth set of hypotheses ($H_{40}$ and $H_{41}$) deals with the effect of gender on academic performance. A $T$ - test conducted indicated that no significant difference in terms of academic performance exists between male and female participants. The mean performance of males were 3.59 percent better, but these results are however not significant ($p = 0.137 > 0.05$). This result may be due to the fact that the female group ($n = 40$) was much smaller than the male group ($n = 137$). These findings contrast sharply with research conducted by Van Rooyen (2001) where it was found that women performed significantly better than men in a university bridging —
course.

Based on the results of this investigation, H4 cannot be rejected with confidence. The results indicate the absence of a linear relationship between gender and academic performance. This implies that no statistically significant difference exists between gender groups in terms of academic performance.

5.2.5 Results and discussion regarding the effect of ethnicity on academic performance

The fifth set of hypotheses (H50 and H51), deals with the effect of ethnicity on academic performance. The ANOVA performed revealed that no significant difference (p > 0.05) exists between the three ethnic groups in terms of their academic performance. However, Indian candidates tended to outperform other ethnic groups in the final year supplementary examination. Indians achieved an 11.42 percent (p = 0.255) higher mean performance than Africans and a 4.64 percent (p = 0.841) better mean academic performance compared to Coloureds. On average, Coloureds outperformed Africans by 6.79 percent (p = 0.279). As was indicated, none of these results are statistically significant at the set 5 percent level of confidence.

Care should therefore, be applied when interpreting findings, as Indians comprised only 3 percent and Coloureds only 8.3 percent of the sample. These findings however replicate those by Viljoen and Vos (1998) and Muller (2002). Findings also tentatively support the assertion (see chapter 3) that cultural background moderates the relationship between intelligence and academic performance (Claassen, 1995; Owen, 1991; Van Rooyen, 2001).

Based on the results of this investigation, H5 cannot be rejected with confidence. This assertion is based on the fact that the results indicate the absence of a linear relationship between ethnicity and academic performance. This implies that no statistically significant differences between different ethnic groups were found regarding their academic performance.
5.2.6 Results and discussion regarding the effect of home language on academic performance

The sixth set of hypotheses (H$_{60}$ and H$_{61}$), deals with the effect of home language on academic performance. An ANOVA performed revealed that English speakers routinely outperformed all other language groups, except for the Venda-speaking group. None of these results are however statistically significant at the 5 percent level of significance. Also these results need to be interpreted carefully considering the unequal group sizes as presented in Table 4.4 (see chapter 4).

These findings replicate other South African research findings. As was indicated in chapter 3, research has shown that language impacts directly on academic performance. For instance, studies conducted by Owen (1991), Van Rooyen (2001) and Viljoen and Vos (1998), have found that speaking English as a home language acts as a decisive advantage in terms of academic achievement. Viljoen and Vos (1998) found a significant ($p < 0.01$) and relatively high positive correlation ($r = 0.622$) between home language and academic performance. Van Rooyen (2001) also found that speaking English as a home language has the highest significant ($p < 0.01$) correlation ($r = 0.3287$) with the average academic achievement of students in a university bridging course.

These findings are to be expected considering the fact that English is the medium of instruction in this university bridging course. These results are interesting in terms of offering a possible explanation for the findings regarding the effect of ethnicity on academic performance. In light of the fact that all Indians who participated in this study are also English speakers, it may be argued that it is their home language, instead of their ethnic make-up, that accounts for their superior academic performance. Although this argument needs to be subjected to further research, it may have important implications in terms of determining which biographical factors are in fact responsible for the variance observed in academic performance.

Based on the results of this investigation, H$_{60}$ cannot be rejected with confidence. Research results indicate the absence of a linear relationship between home language and academic
performance. This means that no statistically significant difference exists between different language groups in terms of average performance. This may imply that a linear relationship does not exist between home language and academic performance.

5.3 CONCLUSIONS

The following conclusions as informed by the research results, are made below. Conclusions are based on the stated general and specific aims regarding the literature review as well as the empirical investigation (see chapter 1).

5.3.1 General conclusion

Based on the results of the empirical study, it may be concluded that the selection battery under investigation is valid for the prediction of academic performance when used for the selection of university bridging - course students.

5.3.2 Specific conclusions in terms of the literature review

5.3.2.1 Conclusions in terms of the role of psychometric assessment in selection

In terms of the role of psychometric assessment in selection, it may be concluded that the selection of personnel need to be pursued in a holistic and scientific manner. It may also be concluded that although psychometric assessment is pivotal to selection, it has limitations. Therefore, other forms of assessment should be utilised, in addition to psychometrics, in order to compensate for these limitations. The fact that the total selection battery declares only 27,1 percent of the variance observed in academic performance confirms the discussion on the role of psychometric assessment within the selection context. It is clear from these findings that measures other than cognitive assessment need to be incorporated within the selection process in order to provide a more complete profile of candidates. As will be discussed later in this chapter, some of these factors that may need to be assessed are emotional intelligence and proficiency in English (Goleman, 1998; Van Rooyen, 2001; Viljoen & Vos, 1998).
Despite these limitations, psychometric assessment may still be viewed as a useful and complimentary tool within the broader selection context if approached in a scientific and legally defensible manner.

5.3.2.2 Conclusions in terms of the nature and assessment of intelligence

The discussion on the theoretical development of the construct intelligence can by no means be considered an exhaustive list of all the developments within the field of intelligence testing. The literature review primarily focussed on the major contributors to and developments in the field. Also, specific emphasis was placed on those developments that stimulated the development of the two independent variables (psychometric instruments) under investigation. It may be concluded that although divergent, all the theories and models discussed, have contributed to our current understanding of the construct intelligence and the identification of its dimensions. It can be concluded that these have had a very important impact on occupational and educational selection and development (Owen & Taljaard, 1996; Smit, 1996).

It may be concluded that different intelligence tests emphasise different components of intelligence and therefore, although a wide variety of psychometric tests are grouped under the phrase intelligence, they may not measure exactly the same construct. As was pointed out earlier, the APIL Battery seems to measure more than Spearman’s $g$ – factor. In contrast, Raven’s APM seems to be a purer measure of $g$ as was indicated earlier (see chapter 4). The implication is therefore, that psychometric test users need to select and interpret psychometric tests with great care.

In the final analysis, it is imperative that the test user ascertains exactly how intelligence is defined within the context of a specific test. I/O Psychologists therefore need to apply their minds in the context of psychometric assessment, as it is essential for the appropriate and responsible assessment of intelligence in Industry (Nainaar, 1999; Raven & Court, 1998; Taylor, 1997).

Considering the overall aim of the university bridging – course under investigation, continued
efforts are needed to ensure that intelligence testing does not hinder candidates from disadvantaged backgrounds from entering academic institutions. Within the South African context, the culture - fair assessment of intelligence is crucial in affecting social justice (Nainaar, 1999; Van Rooyen, 2001).

5.3.3 Specific conclusions in terms of the empirical study

It would seem that the selection battery under investigation has predictive power in terms of the prediction of academic performance in a university bridging - course. The results of the empirical study lead to the conclusion that conceptual ability does indeed have a significant impact on academic performance, as was discussed in chapter 3. As both psychometric tests are measures of intelligence albeit not to the same extent, these findings tentatively confirm the role of intelligence in academic performance as was discussed in the literature review.

The APIL Battery seems to be a relatively good predictor of academic performance, declaring 29.4 percent of the variance observed in terms of average academic performance. Raven's APM is a less effective predictor of academic performance as it only declares 11.5 percent of the variance observed in academic performance. The selection battery in total declares 27.1 percent of the observed variance. It would therefore seem that although Raven's APM is a valid predictor of academic performance, not including it in the selection battery, would not radically alter the predictive power of the selection battery.

Considering the fact that this cognitively - loaded selection battery in total only declares 27.1 percent of the observed variance in academic performance, the conclusion may be drawn that there are other factors that also impact on academic performance. It would seem that assessing conceptual ability in isolation is not sufficient in terms of predicting academic performance. The role of factors that potentially mitigate the relationship between intelligence and academic performance was explored in chapter 3. It can be concluded that these research findings confirm the existence of moderating factors that do indeed affect the variance observed in academic performance (Foxcroft, 1997; Hunter, 1989; Sackett et al, 2001; Van Rooyen, 2001).
The predictive power of the selection battery in terms of academic performance in Biology cannot be confirmed due to the fact that too few cases \((n = 47)\) are available to confirm the results. Furthermore, it would seem that all the predictor variables, show a similar correlation pattern for Mathematics and average academic performance. The conclusion can be drawn that perhaps Mathematics and average academic performance share the same underlying construct. The possibility exists that academic performance in Mathematics may be used as a predictor for overall academic achievement.

Due to the fact that distinct differences in terms of the size of the different biographical groups exist in the sample, firm conclusions regarding the effect of biographical variables on academic performance cannot be made. Results regarding the effect of biographical variables therefore need to be viewed cautiously.

It was stated that Indians tended to perform better than other ethnic groups and English speakers outperformed other language groups. The fact that all Indians in the sample are also English speaking, leads to speculation as to the cause of the observed variance in academic performance. A preliminary conclusion may be drawn that speaking English as a home language and not ethnicity, may be at the heart of the observed variance in academic performance. It may therefore be concluded that speaking English as a home language serves as a distinct advantage in this particular context.

This notion is supported by the literature discussion, which indicated that home language plays a decisive role in the determination of the success of students in university bridging course. Previous studies (Van Rooyen, 2001; Viljoen & Vos, 1998) indicated that because training and examinations are normally conducted in English, first language English speakers are normally advantaged. In agreement with previous research results, English speaking students, tended to perform better than other language groups. Although this interpretation makes logical sense, one should bear in mind that these results need to be interpreted with caution considering the fact that these results are not statistically significant. This aspect is addressed in both the limitations and recommendation sections later in this chapter.
Although the sample was drawn nationally, it was not done according to probability techniques. Therefore it may be concluded that results cannot be generalised nationally. The conclusion may therefore be drawn that using a more representative sample may lead to more useful and statistically significant results.

5.4 LIMITATIONS

At the outset, it was stated that the trait perspective would serve as an encapsulating framework for both the literature review and the research methodological aspects. In terms of the research methodology, this paradigm suggests that a strictly positivist stance be assumed. To some degree, this stance implied that quantitative methods be used. The trait model does not necessarily permit the exploration of variables that are not measurable in a quantitative manner. Therefore, the paradigm perspective followed may have limited the value of the research findings as a qualitative approach may have served to resolve the research problem more fully.

5.4.1 Limitations in terms of the literature review

The discussion of the construct intelligence may not have been a complete representation of the available body of knowledge regarding developments within the field of intelligence testing. This is due to the existence of a wide array of opinions regarding the nature and measurement of intelligence. Theories that were not discussed at length include developmental and cognitive psychological perspectives. In the interest of brevity, the primary focus was on the major contributors to and developments within the field. Also, specific emphasis was placed on those developments that directly stimulated the development of the two psychometric instruments that were investigated.

5.4.2 Limitations in terms of the empirical study

According to Huysamen (1994), the combination of quantitative and qualitative techniques may serve to enrich the research results. Suggestions that can be made in this regard are that interviews and/or focus groups are conducted to gather information of a more qualitative nature.
Participants’ experience of the programme may, for instance, be explored in this manner. Perhaps such an approach, used in conjunction with quantitative methods, may yield “richer” information. This could potentially identify the variables that would declare the variance observed in terms of academic performance more fully.

Due to the fact that a convenience sample was used, it is skewed in terms of gender, ethnicity and home language. In terms of all these factors, the sample does not reflect the distribution in the broader population. Females, Coloureds, Indians, as well as several language groups are proportionally underrepresented in this study. As a result, the effect of these factors on academic performance could not be adequately explored in this investigation. As was mentioned previously, it also placed limits on the generalisability of the research results.

5.5 RECOMMENDATIONS FOR FURTHER RESEARCH

The following recommendations emanate from the research results as well as the above-mentioned limitations.

5.5.1 Recommendations in terms of the literature review

It is recommended that the role of extraneous variables that mitigate the relationship between intelligence and academic performance be investigated more fully. Some of these factors are emotional intelligence as well as cultural and socio-economic variables.

5.5.2 Recommendations in terms of the empirical study

Further research efforts are required in order to attend to the above-mentioned disproportional representation of females, Coloureds, Indians, as well as some language groups, in the sample. It is recommended that a more representative probability sample be used in future to address this aspect. In this manner, more meaningful results regarding the effect of certain biographical factors on academic performance may be generated.
The results of the empirical study indicate that conceptual ability is a valid predictor of academic performance. However, considering the relatively low predictive power of a cognitively - loaded selection battery, assessing conceptual ability in isolation is not sufficient in providing a complete profile of candidates. The effect of other constructs such as interest, personality, social skills and emotional intelligence need to be subjected to further research.

As discussed in the literature review, emotional intelligence may be especially important in predicting academic performance. These factors may have an impact on the adaptation of participants to a new environment. For instance, candidates from other provinces may have been adversely affected by the fact that they are far removed from family and friends. Whether it impacts on academic performance within this specific context, should however be subjected to further research. The impact of emotional intelligence is well worth exploring within the field of selection for academic development programmes. This may be especially relevant in terms of the academic development, as well as the possible lack thereof, of disadvantaged groups.

5.5.3 Recommendations regarding future recruitment and selection

The results point to some differences in academic performance between candidates from different ethnic groups. It may be argued that these differences may be attributed to differing education standards, as most schools are still not fully integrated in terms of ethnicity.

The results further indicate that differences may exist between candidates regarding home language. It is therefore recommended that some assessment of proficiency in English be included in the selection battery in order to obtain a more holistic profile of candidates.

Despite the fact that it is a valid predictor of academic performance, excluding Raven's APM would not significantly alter the predictive efficiency of the selection battery. It is therefore, recommended that it be removed from the selection battery. Considering the financial implications of conducting a selection process, such a step may be considered prudent.
5.5.4 Recommendations in terms of training

In order to secure a return on investment in terms of training expenditure, organisations need to ensure that the right people are trained and given an opportunity to utilise these skills. Thus enabling employees to perform, it may well lead to greater organisational efficiency. Perhaps more attention needs to be given to social skills training as well as emotional intelligence training. Further, incorporating English communication skills training within this university bridging - course may have a positive impact on academic performance.

5.6 SUMMARY

In this chapter, the results of the empirical study were presented and discussed. The nature of the research results facilitated the rejection of three out of the six stated hypotheses. Results indicate that the selection battery under investigation does have predictive validity in terms of academic performance in a final year supplementary matriculation examination. It would seem that the APIL Battery is a good predictor of academic success. Results also indicate that Raven's APM, although it does have predictive validity, does not add much value to the overall predictive power of the selection battery.

It was concluded that intelligence does have an impact on academic performance. It was emphasised however, that it is not the only factor that influences academic performance. Based on this assertion, it was recommended that the impact of other mitigating factors be explored in further research efforts. It was also stated that the fact that a non – probability sample was used, resulted in the underrepresentation of females, Coloureds, Indians, as well as several language groups, in the sample. This state of affairs placed limitations on the generalisability of research results. In the final analysis, it would seem that this investigation conformed to the stated research questions and hypotheses.
REFERENCES


Health Professions Act, No.56. (1974).


