Chapter 4
Psychometric testing

4.1 Introduction

Traditionally, scholastic achievements were used as selection criteria to gain access to South African higher education institutions. Kotze, Van der Merwe and Nel (1996: 39) state that

> "...mainly because of unequal educational opportunities, (this) placed many students at a severe disadvantage. Although research in South Africa and elsewhere has clearly indicated that matriculation results remain the best predictor of success at tertiary level, it was also found that matriculation results of matriculants from the previous Department of Education and Training and equivalent school systems, particularly at the lower ranges, are an inaccurate reflection of students' academic ability, or potential for success at tertiary level."

Van Aswegen (1997:14) supports the above authors by saying that the implications of the inequalities in the education system are clearly demonstrated by the patterns of access to higher education. He continues by referring to the current crises faced by universities and technikons namely of finding appropriate criteria for undergraduate admission of disadvantaged black students. This is because the difference in the allocation of resources between black and white schools influence the credibility of grades obtained as an accurate reflection of students' academic potential.

From the above, together with the aspects touched on by the White Paper on Higher Education, as discussed in Chapter 2, the need for a fair, reliable and valid

instrument to assess the ability of prospective students, in order to see whether they have the potential to be successful in their studies, has become clear.

In order to conceptualise the importance of a fair instrument, it is necessary to understand the current socio-political context of psychometric testing in South Africa. This chapter will first provide the aim of psychometric testing; secondly the current perceptions regarding psychometric testing in South Africa; and will then go on to discuss the nature of psychometric testing.

4.2 The aim of psychometric testing

According to Gregory (1996:34), psychometric testing can be compared to procedures in the physical sciences whereby numbers represent abstract dimensions such as weight or height. Similarly, every psychometric test provides evidence that a person belongs to a specific category and not another. Gregory (1996:35) continues:

> "The implicit assumption of the psychometric viewpoint is that tests measure individual differences in traits and characteristics that exist in some vague sense of the word. In most cases, all people are assumed to possess the trait or characteristic being measured, albeit in different amounts. The purpose of the testing is to estimate the amount of the trait or quality possessed by an individual."

We live in a specialised society and specialised roles should be filled. Cronbach (1990:9) feels that individuals should follow different paths in training in order to fulfil these roles. Psychometric testing, in measuring relevant characteristics, plays an important part in the differentiation of people into specific roles.

4.3 Psychometrics at a crossroad

According to Van Aswegen (1997:6), it is generally agreed that psychometric assessment in South Africa is at a historical crossroad as it tries to free itself from European theories. Nzimande (1995:1) comments that historically disadvantaged groups perceive psychometrics to symbolise the overt and covert racism of South African psychology, as they perceive the development of psychology in this country as closely related to the history of psychometric testing.

The use of traditional psychometric tests for the assessment and prediction of performance of members of disadvantaged groups has been criticised in the past. Van Aswegen (1997:1) states that, among other things, the criticism focuses on the frequent assumption that all people had equal opportunities to comprehend the knowledge and skills the tests require. Hessels (1995:35) makes it clear that this assumption cannot be maintained especially in the case of historically disadvantaged groups with a different linguistic and cultural background. He furthermore feels that these groups are confronted with tests of which the contents are attuned to the dominantly white culture.

Although Taylor (1987:2) acknowledges that some First World countries, for example, the USA, are becoming more and more culturally diverse and are therefore experiencing the same testing problems, he continues by saying that the problems in South Africa are more intense and that psychologists need to devote their attention to addressing those needs by developing appropriate models and instruments. According to Taylor (1987:3), social and political developments could, in future, drive psychological testing from the scene if it does not adjust its role.

At present it is generally known that this prophecy made by Taylor in 1987 has partly been fulfilled and that psychometric testing, particularly for selection and

promotion purposes, has been experiencing an enormous amount of pressure, especially from political circles. Schaap, (1997:36) states that criticism is often levelled at psychological instruments, and the discriminating effect of instruments on decisions regarding certain groups and individuals is often emphasised. In the South African, context a negative accent is often placed on the cultural differences as far as they affect validity and the results of measurements and interpretation.

The search for alternatives for psychological tests over the past years has caused a debate of its own. Murphy and Davidshoffer (1994:15) state that: "A critical issue in the debate over psychological tests is wether or not there are alternatives that would be better, fairer, or more economical." The authors continue by referring to the example of the debate on the use of tests in academic admissions in the United States of America. Some critics feel that grade point average (GPA), together with interviews and letters of recommendation, would lead to better decisions than would psychological tests, such as the Scholastic Aptitude Test. Murphy and Davidshoffer (1994:16) comment that the

> "GPA clearly reflects an assessment device that meets our definition of a 'test'. That is, it is based on somewhat systematic observations (typically, a large number of observations from several different courses) under reasonably standard conditions, employing systematic (if not always objective) scoring rules... Both interviews and letters of recommendation fall within the category of behavioral observations...These methods are not based on systematic samples of behavior, they lack standardization and are characterized by highly subjective scoring rules."

In the light of the above it is clear that these methods are less likely to be valid predictors of future performance than well-designed psychological tests. The authors conclude: "Although psychological tests are far from perfect, they represent the best, fairest and most accurate technology available for making

many important decisions about individuals."(Murphy & Davidshoffer, 1994:2.)

From the aforegoing it is clear that the discrepancy between the understanding of psychometrics lies in the terms ***differentiation*** and ***discrimination***. Groups in favour of psychometric testing argue that psychometrics categorise people by measuring certain traits and by summing up performance in numbers or classifications (Gregory, 1996:35; Murphy & Davidshoffer, 1994: 2; Cronbach, 1990:9). Groups opposing the use of psychometrics feel that tests assume that all people had the same opportunities to comprehend the knowledge and skills these tests require (Van Aswegen, 1997:1). They argue that psychometric tests put testees from a disadvantaged background in a situation where they are confronted with contents attuned to a white middle-class culture (Hessels, 1995:35).

In an attempt to find a solution to the difficulties facing psychometric testing in South Africa, the idea of assessing learning potential has gained increasing popularity among politicised academics, unionists and personnel practitioners alike. According to Van Aswegen (1997:7), the popularity the concept of learning potential in South Africa can be ascribed to its promise of providing a means of fair assessment, despite unequal educational opportunities. The National Union of Metal Workers (NUMSA) emphasises in its policy document (1992) that tests for learning potential ought to be used. Nzimande (1995:6), highly critical of psychometric testing in general, feels that the assessment and identification of potential might be useful in redressing historical inequalities.

Shirley (1994:1) states:

> "...in order to realise our ambitions of redressing the wrongs of the past without compromise in quality performance, we will have to come to terms with the following, in the cognitive domain:

- ⇨ Measures must concentrate on the identification of potential rather than on measures which reflect crystallized competencies or skills if they are to be fair to groups originating from impoverished educational and developmental contexts;

- ⇨ The implication is that we should look for modifiability in domain specific skills - assessment of potential which will also be good predictors of performance in actual work settings..."

In the debate on equity and efficiency, Anastasi and Urbina (1997:545) state the other side of the coin when they say that insofar as culture influences behaviour, this effect will and should be detected by psychometric tests. They continue:

> "If we rule out all cultural differentials from a test, we may thereby lower its validity as a measure of the behavior domain it was designed to assess. In that case, the test would fail to provide the kind of information needed to correct the very conditions that impaired performance."

To accommodate the demands placed upon psychometric testing, many test developers have worked towards changing the face of psychometric testing.

To Erasmus (1997:52), this *changed face* implies a transition from testing to assessment when he says that to "test" means to subject a person to a process in which he or she could either fail or pass, whereas assessment should be seen as the mere mental finger printing of a person's potential, with such an assessment only applicable at the point of assessment. He sees potential as a phenomenon which could be manipulated or altered over time through interventions (that is, turning potential into performance), such as training and exposure to a stimulating environment. This presupposes a dynamic approach which recognises the need for assessment and reassessment in order to establish an index of growth.

Schoombee (1992:20) suggests that the development of psychological tests in different parts of the world tends to be more and more scientific and research-orientated. Anastasi and Urbina (1997:27) add that psychometric testing is in a state of rapid change.

As new and revised tests are being published in abundance, it would be worth our while to have a theoretical look at the nature of a scientifically accountable psychometric test.

4.4 The nature of psychometric tests

4.4.1 Definition

A study of the literature reveals many definitions of psychometric or psychological tests.

Sax (1980:13) gives the following definition: "A test may be defined as a task or a series of tasks used to obtain systematic observations presumed to be representative of educational or psychological traits or attributes."

Walsh and Betz (1985:20) say "...a psychological test is a highly refined and systematized version of the ordinary process of observation of ourselves and the people around us."

Cronbach (1990:32) defines a psychological test as "...a systematic procedure for observing a person's behavior and describing it with the aid of a numerical scale or a category system."

Murphy and Davidshofer (1994:3) state that:

"A psychological test is a measurement instrument that has three defining characteristics:

1. A psychological test is a sample of behavior.
2. The sample is obtained under standardized conditions.
3. There are established rules for scoring, or for obtaining quantitative (numeric) information from the behavior sample."

Gregory (1996:33) describes a psychometric test as "a standardized procedure for sampling behavior and describing it with categories or scores. In addition, most tests have norms or standards by which the results can be used to predict other, more important behaviors."

Anastasi and Urbina (1997:4) define a psychological test as an objective and standardised measure of a sample of behaviour.

Smit (1991: 25) gives a comprehensive definition encompassing all the above when he says that "**a psychological test can be described as an objective and standardised measurement of a specific sample of human behaviour. A psychological test therefore measures an aspect of human behaviour from which estimations of the testee's natural abilities and personality characteristics can be made for diagnostic, prognostic or predictive purposes.**"

4.4.2 Standardisation

It is clear from the above definition of psychometric testing, by Smit, that the concept of standardisation is important in the understanding of the nature of psychological testing.

To Walsh and Betz (1983:21) standardisation means that constant procedures and materials are used and that the methods of scoring the test are consequent across all testings. Mehrens and Lehmann (1991:280) add to this definition by

commenting that the scoring in a standardised test is usually objective.

Gregory (1996:33) says that a "test is considered to be standardized if the procedures for administrating it are uniform from one examiner and setting to another." Anastasi and Urbina(1997:6) elaborate by placing emphasis on the administering of the test to a large representative sample, known as the standardization sample, in order to establish the norms. According to them this is important because, in most instances an individual's test performance is interpreted by comparing it with that of others.

4.4.3 Objectivity

According to the accepted definition (Smit,1991:25), psychological testing is characterised by objectivity.

Anastasi and Urbina(1997:7) consider a test to be objective if the administration, scoring and interpretation of scores are independent of the subjective judgement of a particular examiner.

Smit (1991:26) quotes various authors in order to explain the concept of objectivity. His conclusion can be summarised as meaning that a test is objective if -

- a testee is not conscious of the way in which his answers can influence the interpretation of the test;
- the correctness of an answer can be determined unambiguously; and
- the test results are independent of the subjective judgement of an examiner.

4.4.4 Validity

There seems to be consensus among writers that validity is the most important characteristic of a psychometric test (Brown, 1983:19; AERA, APA, NCME, 1985:9; Anastasi & Urbina, 1997:8). Validity is said to refer to the extent to which a test measures what it is designed or developed to measure (Brown, 1983:19; Walsh & Betz, 1985:56; Anastasi & Urbina, 1997:8; Kline, 1993:15). A number of theorists add that validity involves the extent to which appropriate and meaningful inferences can be made from test scores and other measurements (Sax, 1980:289; Brown, 1983:98; AERA, APA, NCME, 1985:9; Mehrens & Lehmann, 1991:265; Gregory, 1996:107).

Test validation refers to the process of accumulating evidence to support the inferences made from test scores, therefore the inferences regarding specific uses of a test are validated, not the test itself (AERA, APA, NCME, 1985:9; Gregory, 1996:107). Anastasi and Urbina, (1997:113) support this by saying that the validity of a test cannot be reported in general terms, as it has to be established with reference to the particular use for which the test is being considered. Mehrens and Lehmann(1991:266) stress this argument further by saying that it is only the degree to which the evidence supports the inferences that are made from the test scores that can be labelled valid or invalid.

Traditionally, validity is divided into three categories, namely, content-related validity, criterion-related validity and construct-related validity. (AERA, APA, NCME, 1985:9; Anastasi & Urbina, 1997:114; Murphy & Davidshofer, 1994: 107.)

4.4.4.1 Contend-related validity

According to Gregory (1996:108), content validity is "determined by the degree to which the questions, tasks, or items on a test are representative of the universe of behavior the test was designed to sample." The determination of content

validity is usually a judgmental process that provides no quantitative index or measure of validity (Brown, 1983:133).

In selection, inferences based on content validity are concerned with whether or not a test or measurement procedure contains a fair sample of the performance domain it is supposed to represent. If a selection procedure is thus content valid, it will give the applicant a fair chance to demonstrate his or her competence to perform the relevant task (Van Aswegen, 1997:19).

4.4.4.2 Construct-related validity

Construct validity can be described as focussing on the test score as a measure of the psychological characteristic of interest (AERA, APA, NCME, 1985:9). In determining construct validity, testable hypotheses are deduced from theory and data is then collected to test these hypotheses. If the hypotheses are substantiated, the conclusion can be drawn that the test measures the construct at interest (Brown, 1983:139). According to Anastasi and Urbina (1997:126), any data enlightening the nature of the construct under consideration represents appropriate evidence for this validation.

4.4.4.3 Criterion-related validity

Smit (1991: 58) states that criterion-related validity is specifically important in tests used for selection purposes. Therefore it is important for the purpose of this study. A detailed discussion will thus be given.

Criterion-related validity refers to the empirical technique of studying the relationship between test scores and an independent, external measure or criterion (Mehrens & Lehmann, 1991:268). Therefore criterion-related validity indicates the effectiveness of a test or measure in predicting an individual's performance in specific activities (Anastasi & Urbina, 1997:118). It is important to

remember that the inferences made from the test scores are being validated and not the test scores per se (Brown, 1983:100).

According to Brown (1983:99), the emphasis in criterion-related validity is fundamentally placed on the criterion performance, as we are interested in the test scores because they predict some important external behaviour. He further states that the idea that tests are used as part of decision-making processes is implicit in the concept of criterion-related validity. "Because the validity of a test is judged by its relationship to the criterion, if the criterion measure does not adequately reflect the desired outcome, the decision-making process will be less effective." (Brown, 1983:100.)

Authors distinguish between two types of criterion-related validity, namely, **predictive validity** and **concurrent validity** (Brown, 1883:99; Walsh & Betz, 1985:58; Anastasi & Urbina, 1997:118; Mehrens & Lehmann, 1991:268; Gregory, 1996:111).

Predictive validity refers to the accuracy with which early test data can be used to estimate criterion performance in the future (AERA, APA, NCME, 1985:11).

Concurrent validity is studied when the predictor and criterion scores are obtained simultaneously and the present status on the test and the present status on the criterion are being observed (Walsh & Betz, 1985:58).

4.4.4.3.1 Selecting the criteria

Brown (1983:101) states that if a test was developed to predict performance in a field, a standard measure of performance must be identified - the criterion. He continues by saying that in most cases, having multiple criteria is better than having one criterion only, as the latter causes valuable information to be neglected.

Latham and Wexley (1981:72) as well as Cascio (1991:68), refer to the debate between theorists on the utilisation of composite versus multiple criteria as measure of performance. The basic argument for composite criteria states that the measurement or criterion should provide an overall yardstick of success for each individual respondent. Latham and Wexley (1981:72) discuss three methods by which composite criteria can be established:

- Each criterion may be weighted equally. This practice assumes that each criterion is equally important for defining overall success on the task.

- The criteria can be subjectively weighted by so called experts.

- The criteria can be weighted in terms of their monetary value for the organization.

Advocates of multiple criteria argue that measures of demonstrably different variables should not be combined (Cascio, 1991:68). As this study is about the validity of a selection battery for an academic programme, multiple criteria will primarily be used. The reason for this decision is that the different subjects which students have to pass differ vastly from each other. Furthermore, none of these subjects are more important than the other, as no qualification can be obtained before all of the prescribed subjects have been passed.

Mehrens and Lehmann (1991:269) comment that, when studying criterion-related validity, both the conceptual and operational (measurement) aspects of the criterion should be examined. The criterion is the global concept of successful performance and this conceptual idea is identified by an operational measure (Brown, 1983:101). In this study, the conceptual criterion will be success in engineering studies and the criterion measure will be the academic marks obtained.

In order to be adequate, criterion measures ought to have certain characteristics. The most important feature of a criterion measure is **relevance**. This means that a criterion measure should reflect the important aspects of the conceptual criterion. (Brown, 1983:102; Mehrens & Lehmann, 1991:269).

Secondly, a criterion measure should be **reliable,** for if it varies from time to time, or from situation to situation it cannot consistently relate to other measures (Brown, 1983:102). Mehrens and Lehmann (1991:269) qualify this statement by saying that as the maximum relationship obtained between two variables is equal to the square root of their respective reliabilities, the reliability of the criterion affects the criterion-related validity as much as the reliability of the predictor.

A third characteristic of criterion measurement is that it should be **free from bias.** Bias is commonly caused by criterion contamination that occurs when the criterion score is influenced by the knowledge of the predictor score (Brown, 1983:102; Mehrens & Lehmann, 1991:269), for example, if a teacher knows the IQ of a pupil it might influence his or her rating of that pupil.

4.4.4.3.2 Validity coefficient

A validity coefficient is a correlation between test score and criterion measurement that gives a single numerical index of test validity (Anastasi & Urbina, 1997:141). The higher the correlation the more accurate can the scores on the criterion be predicted from test scores (Brown, 1983:104).

4.4.5 Reliability

Reliability refers to the extent to which scores obtained by the same person on the same test or a different form of the same test, administered on different occasions, remain consistent (Anastasi & Urbina, 1997:84). In these instances, differences among scores are generally known as errors of measurement and it

can thus be said that reliability refers to the degree to which a test score is free from errors of measurement (AERA, APA, NCME, 1985:19). These differences cannot be attributed to errors of measurement if maturation or intervention occurred, or if the inconsistency is relevant to the construct being measured (AERA, APA, NCME, 1985:19).

Theorists like Anastasi and Urbina (1997:91-99), Walsh and Betz (1985:48), Mehrens and Lehmann (1991:254), Kline (1993:5) and Gregory (1996:92) distinguish between the following types of reliability:

- **Test-retest reliability** refers to the extent in which the scores obtained on an identical test, administered at different occasions, are similar.

- **Alternate forms reliability** refers to the degree in which alternative forms of the same test yield similar results.

- **Split-half reliability** is obtained by procedures by which two scores are obtained for each testee by dividing the test in two equivalent halves. This provides a measure of consistency regarding the content sampling of the test (internal consistency).

- **Kuder-Richardson reliability and Coefficient Alpha** are methods by which reliability are determined by utilising a single administration of the test. The methods are based on the consistency of responses throughout the test.

- **Scorer reliability** is determined by calculating the relationship between scores given on the same test by different scores. Most tests provide standardised procedures for administration, and variance in the test scores due to the administrator are minimal. With individual tests, utilised for intensive examination, there is however, evidence of variance due to the

administrator. It is important to determine this type of reliability when using subjectively scored instruments in research.

Kline (1993:13) states that "high reliability, both test-retest and internal consistency, is essential for a test to be valid."

4.4.6 Test bias

In the literature many different definitions for test bias are found.

Brown (1983:224) defines test bias as follows:

> "**A test can be considered biased if it differentiates between members of various groups on bases other than the characteristic being measured.** That is, a test is biased if its content, procedures, or use result in a systematic advantage (or disadvantage) to members of certain groups over other groups and if the basis of this differentiation is irrelevant to the test purpose."

Taylor and Radford (1986:80) refer to several authors when making the statement that a test is biased if group differences in performance are too large or too small as compared to the group difference on the criterion measure which is used in predicting performance. They state that test bias may reside in test items, the test as a whole, within subjects, in the tester, in the testing context or in the interaction of these factors with other sources. They are of the opinion that test bias can never be eliminated, but that steps can be taken to minimise the effect of test bias, as its potential sources are known.

According to Rust and Golombok, as referred to by Geldenhuys (1996:20), a test is biased if the testing procedure is unfair to members of a specific group. These groups can be defined on grounds of race, gender, language and cultural differences. They continue by making a distinction between intrinsic and extrinsic test bias:

- **Intrinsic test bias** occurs when there is a difference in the mean performance of different groups and this difference can be ascribed to characteristics of the test itself and not to a difference in the trait being measured.
- **Extrinsic test bias** occurs if differences in test performance between groups can be ascribed to a genuine difference in the characteristic being measured.

Among others, Anastasi and Urbina (1997:164) and Brown (1983:224) oppose the view of Rust and Golombok. Brown states that it is important to realise that his definition does not imply that a test is biased merely because members of different groups perform differently. When a group performs differently on a test and the scores reflect a difference in the characteristic or trait being measured, the test is **not** biased. However, Brown (1983:224) concludes that as soon as a mean and/or distribution difference between groups is found, the possibility of test bias should be investigated.

Anastasi and Urbina (1997:164) state that:

> "If we want to use tests to predict outcomes in some future situation, such as an applicant's performance in college or job, we need tests with a high predictive validity against the particular criterion. This requirement is often overlooked in the development of so called culture-fair tests. In the effort to include in these tests only functions common to different cultures or subcultures, we may choose content that has little or no relevance to any criterion we wish to predict."

Brown (1983:227), as well as Walsh and Betz (1985:379), refers to different types of test bias:

- **Content bias** occurs when the content of test items gives a systematic

advantage to a specific group of testees, for example when the test contains questions that are more familiar to one group than another. Another form of content bias can be found in item format and presentation, when for instance pictorial material only depicts white males and never females or blacks.

Murphy and Davidshoffer (1994:286) refer to this type of bias as **cultural bias**, where a group of testees had the opportunity to become familiar with the test content and another group not. The example they use is that of test items that are highly academic in nature; the underlying assumption being that the school environment is more foreign to disadvantaged groups. Furthermore verbal items are more likely to be regarded as biased than nonverbal items - the rationale being that verbal items are likely to be presented in standard language (English, in their example), which more closely resembles the spoken language of the middle-class.

- **Internal structure bias** occurs if the internal or factor structure of a test and/or the behaviour of items in relationship to each other differ across cultural groups. This would imply that the test measures different things across groups.

- **Atmosphere bias** refers to the effects of the testing conditions on test takers' performance. For example the type of motivation elicited, factors related to the tester-testee interaction, and factors in the evaluation and scoring of responses.

- **Prediction/Selection bias** is caused when a test has different predictive validity across groups. Selection bias is examined through the comparison of regression equations and regression lines obtained with the different groups. This type of bias was discussed in detail in Chapter 3.

4.5 Conclusion

As there is ample evidence against the utilisation of traditional tests for the assessment and prediction of performance of members of disadvantaged groups, it was necessary to find an instrument which was proved to be valid and reliable in South African circumstances.

For this study, the Potential Index Batteries, developed by Erasmus and Minnaar, were used. Motivation for the selection of this specific instrument will be provided in Chapter 7.

Chapter 5
Criterion development

5.1 Introduction

According to Roe and Geuter (1991:197), the choice and definition of criteria is a traditional theme for psychologists involved in selection. Cascio (1991:50) supports this statement by saying that adequate and accurate measurement is a fundamental problem in personnel psychology. He continues "...although criteria are sometimes used for predictive purposes and sometimes for evaluative purposes, in both cases they represent that which is important or desirable" (Cascio, 1991:50).

5.2 Defining the concept

Cascio (1991:50) defines ***criteria*** as follows: "Criteria are operational statements of goals or desired outcomes." This definition, according to him, is applicable both when the criteria are used for predictive and for evaluative purposes. Criteria should always represent that which is important and desirable.

Pieters (1996: 297) refers to criteria as the specific behaviour, work process and/or outcomes that should be evaluated.

5.3 Criterion problem

According to Leap and Crino (1993:337), the ***criterion problem*** is created by an inadequate performance evaluation system used in a criterion-related study. This is caused by validating the selection device against performance information (criteria) that does not represent true levels of performance. Under these circumstances, the validity coefficient will be inaccurate and undefendable.

5.4 Criterion contamination

Anastasi and Urbina (1997:119) refer to the concept of ***criterion contamination*** as a possible source of error in the validation of selection procedures. According to them, criterion contamination occurs when the scores obtained on predictors themselves influence individuals' criterion status. For example, if a lecturer knows that a specific student has performed poorly on the potential assessment test, such knowledge might influence the grades given to that student. This would obviously raise the correlation between predictor scores and criterion scores in an artificial manner. It is therefore suggested that predictor scores are kept strictly confidential and that no person participating in the assignment of criteria ratings have any knowledge of the testee's assessment scores.

5.5 Dimensionality of criteria

According to Cascio (1991:53), operational measures of criteria vary along various dimensions of which the most important are discussed below.

5.5.1 Temporal dimensionality

Cascio(1991:54) states that the optimum time for criterion measurement varies from situation to situation and that conclusions might be influenced on when the criterion measurement was taken.

As far as this study is concerned the implication will be that academic results taken early in a semester might differ largely from academic results taken later in the semester, due to various factors. Failure to consider the temporal dimension of criteria may thus lead to misinterpretations.

Cascio (1991:55) continues by identifying two special cases of temporal dimensionality, namely:

➢ **Static dimensionality**

The nature of any task performance, if observed at any specific time, is multidimensional. That implies that a number of independent skills are involved in the performance of the task at hand. Leap and Crino (1993:341) support this argument by saying that there is no universal set of performance criteria for all tasks or jobs. If the criteria are not relevant to a specific task, they should not be examined.

In criterion research, the static dimensions are those typically investigated. A "photograph" of performance is taken at a single point in time, usually involving a single criterion, and the assumption is made that the performance of the employee or student has been "captured". Inevitably, this is not always the case.

➢ **Dynamic dimensionality**

As employees (and in the case of this study, students) develop their potential and abilities, the dimensions of performance that seemed important and valid early in their careers may change and seem irrelevant to their performance at a later

stage. Therefore it could be said that criteria may be "dynamic" - changing in importance over time.

Dynamic criteria might assume one of three possible forms (Cascio, 1991:56):

a. changes in average group performance over time;
b. changes in validity over time;
c. changes in the rank-ordering of scores on the criterion over time.

Changes in the validity over time, as seen by Cascio (1991:56), are of importance for this study.

Two possible explanations for the changing of validities over time has been given:

i. Changing task model

This model suggests that "while the relative amounts of ability possessed by individuals remain stable over time, criteria for effective performance might change in importance" (Cascio, 1991:56).

As the demands of the labour market change, especially the rapidly changing technological demands, the training of engineering students has to be adapted. This causes the desired performance of these students to change over time and therefore the validity of criteria (and predictors) will change.

ii. Changing subject model

The second model suggests that validities might fluctuate because an individual's level of ability may change over time, even though specific abilities required for effective performance may remain constant.

Referring to the issues of admission to higher education, raised in Chapter 2, the relevance of this model to the study at hand is clear. As admission to higher education has been made more accessible, the profile of the student undergoing

training has changed drastically. Due to historical inequality, a large proportion of current students lack certain academic skills that are necessary to achieve success in a higher education environment. This has a definite influence on the validity of predictors and criteria. It was, in fact, this situation which motivated this study.

5.6 Essentials of criteria development

Cascio (1991:57) refers to research done over 40 years ago by Stuit and Wilson, which demonstrated that continuing attention to the development of better performance measures leads to better prediction of performance.

He states that criteria should be developed and examined before predictors are selected to predict the criteria performance: "...if we use predictors with no criteria, we will never know whether or not we are selecting those individuals who are most likely to succeed" (Cascio, 1991:57).

He further identifies the following four basic problems which should be dealt with before human performance can be studied:

5.6.1 Reliability of Performance

This essential issue in personnel research refers, in this context, to the consistency or stability of job or task performance over time. According to Cascio (1991:58), the assumption of performance reliability is implicit in all predictive studies.

Cascio (1991: 58) summarises research done over 30 years of personnel psychology that shows that the reliability of performance varies greatly - individual output was found to be erratic and highly inconsistent. This could lead to

misinterpretation in research if the researcher happened to select a period of unusually high or low performance.

Cascio (1991:58) suggests that measures could be employed to compensate for the above inconsistency. One such measure could be to aggregate performance and thereby cancelling out or minimising incidental factors.

The implication for this study would be that final pass marks (a combination of marks obtained throughout the semester) would be used as performance measurements and not, for instance, the results of a single test series.

The main problem experienced when working with a set of final marks was that no mark was available for those students who had decided not to continue with a subject until at least predicate day.

5.6.2 Reliability of job performance observations

This aspect is of cardinal importance in prediction, as all evaluations of performance depend on an observation of some kind, and different methods of observation might lead to different conclusions (Cascio, 1991:59).

Since academic results will be used as criteria in this study, this issue might be addressed by the fact that all students are examined at the same time, by the same paper, scored by the same lecturer. The observations (marks) should therefore be consistent and comparable.

5.6.3 Dimensionality of job performance

Cascio (1991:60) states that research reveals that a variety of predictors are generally used in personnel decisions, but that the majority of studies use a single

or global measure of the job or task performance. He asks whether it is meaningful or realistic to reduce performance measurement to a single measure.

This problem is less relevant for this study, as academic achievement as expected from the selected students is not such a complex task that it needs more than academic performance as criteria measurement.

5.6.4 Performance and situational characteristics

According to Cascio (1991:60), this issue deals with the influence of environmental factors and conditions on individual levels of performance. In research it is necessary that attention be given to the possible moderating effects of variables other than those measured by predictors.

5.7 Steps in the development of criterion

Cascio (1991:62) refers to a five-step procedure for criterion development outlined by Guion, namely:

1. Analysis of job or task and/or organizational needs.
2. Development of measures of actual behaviour relative to expected behaviour, as identified in Step 1.
3. Identification of criterion dimensions underlying measures developed in Step 2.
4. Development of reliable measures, each with high construct validity of the identified elements.
5. Determination of the predictive validity of each independent predictor for each one of the criterion measurements.

Smith and Robertson (1986:50-51) identifies three steps in the development of criteria, namely:

1. Analysis of job or task which describes expected behaviour.
2. Analysis of expected outcomes.
3. Analysis of individual's contribution to the achievement of organizational goals.

The last step is not always practical, since it may be difficult to gather this data in time for selection decisions.

5.8 Evaluating criteria

Three concerns should be taken into consideration when criteria are evaluated (Cascio, 1991:63).

5.8.1 Relevance

Any criterion should be logically related to the conceptual criterion (that is, desirable outcomes based on the more general purposes or aims of the organisation). It is therefore important that the conceptual criterion is set clearly.

This point is also emphasised by the American Psychological Association (1985:60): "The rationale for criterion relevance should be made explicit. It should include a description of the job in question and of the judgements used to determine relevance."

Cascio (1991:64) continues by saying that if an important aspect of job or task performance is not assessed, an additional criterion measurement is required,

regardless of how many criteria are already used.

5.8.2 Sensitivity or discriminability

A criterion measure, in order to be useful, must be capable of discriminating between effective and ineffective employees (Cascio, 1991:64).

5.8.3 Practicality

Record-keeping and data collections cannot become impractical or interfere with ongoing operations, and therefore personnel researchers should keep criterion measures as relevant and practical as possible (Cascio, 1991:64).

The set of criteria selected for this validation study consists of all academic subjects taken by Civil and Mechanical Engineering Technology students at Technikon Pretoria, for the period 1996 - 1999.

As the general purpose of Technikon Pretoria, as institution for higher education, is to deliver qualified professionals to the labour market, academic success should surely be seen as a relevant criterion. Furthermore, academic performance discriminates and distinguishes clearly between successful and unsuccessful students. Academic records are kept for every student registered at Technikon Pretoria, thus the criteria remain practical as well.

5.9 Conclusion

In order to validate a selection battery for the selection of first-year students, the development of criteria and criterion measures is essential.

In the case of this study, the desired outcome is academic achievement, which will be measured by means of academic performance. In order to have criterion measures that are as valid and reliable as possible, academic performance will be aggregated to compensate for incidental factors. Care will also be taken that, as far as possible, all results used were obtained on the same tests, scored by the same lecturer.

It could be argued that academic performance is not the only relevant measurement when it comes to the selection of students. Issues such as performance on the actual job could also be regarded as relevant. The counter argument could be that a student who will not be successful in the job market should not pass the course.

As this study was initiated to address the specific problem of admitting students to higher education, academic success was deemed criterion enough. Future studies could focus on the predictive validity of the suggested selection battery used in this study for performance in the labour market.

Criteria measurements used in similar studies will be discussed in Chapter 6.

Chapter 6

Predictors of academic success

6.1 Introduction

Zaaiman (1998:61) states that the prediction of future behaviour is an important element of selection. She adds that the behaviour of each individual is influenced by a variety of interacting factors, and that, for selection purposes, one should identify those factors which can be measured in practice and which make significant contributions to the prediction of eventual academic performance.

In this chapter, attention will be given to such possible predictors, as identified and examined by previous research, namely, scholastic performance, learning ability, general cognitive ability, specific aptitudes and biographic factors such as gender and age. Subsequently, because of the focus of this study, possible predictors for success in engineering technician courses, as found in literature, will be discussed.

6.2 Scholastic performance

According to Zaaiman (1998:61), selection for higher education is internationally based mainly on academic criteria, where academic achievement is usually represented in the form of final school marks. This remark is supported by other researchers such as Kotze, Van der Merwe and Nel (1996:39) when they say that

matriculation results have internationally been proved to be the best predictor of academic success at higher education level.

The best known ways of quantifying matriculation symbols are a conversion to the so called M-score and the allocation of a Swedish Formula score. Table 6.1 illustrates the different approaches to quantification.

Table 6.1: Quantification of Matriculation Symbols

Swedish Formula		Matriculation Symbol	M-scores	
HG	**SG**		**HG**	**SG**
8	6	A	5	4
7	5	B	4	3
6	4	C	3	2
5	3	D	2	1
4	2	E	1	0
3	1	F	0	0

Over the past two decades a number of research findings showing the relationship between scholastic achievement and first-year higher education success have been published.

Stoker, Engelbrecht, Crowther, Du Toit and Herbst (1985:7) reported on twelve South African studies between 1957 and 1977 that " they all found that success at school did to a fairly large degree extend to achievement at university, particularly first year achievement". Bokhost, Foster and Lea (1992:59), however, comment that most of those studies were conducted at white, mainly Afrikaans-speaking universities.

In the study commissioned by the Committee of University Principals and conducted by the HSRC (Stoker et al, 1985), a number of factors were included, namely, age, school aggregate, gender, population group, examining body, home language and university. From this study it emerged that school aggregate was the best single predictor of university performance; specifically in the BA study field for both language and non-language directions.

In the science-related fields of study, research has also indicated that matriculation results have a strong relationship with university performance. In the case of the University of Cape Town, Moran (1987:2) expressed the following: "The evidence is clear and unequivocal. Matriculation point scores were a surprisingly excellent indicator of subsequent academic success for our first-year intake of students."

In a study performed by Bokhorst, Foster and Lea (1992:64) on factors affecting first-year students in Psychology at the University of Cape Town, it was found that the matriculation aggregate had an extremely high predictive effect, appearing as an almost perfect positive correlation. These results should however be interpreted with some reservation, as the number of non-white respondents was very low. Furthermore, significant effects for population group, language and matriculation authority were found, all pointing to a black disadvantage (Bokhorst et al, 1992:64).

At the Rand Afrikaans University, Zietsman (1985:22) found that matric symbols, weighted according to the Swedish Formula, were good predictors of academic success in Chemistry and Botany. Snyman (1987:9), at the same institution examined admission requirements for Law studies. He suggested a minimum of a higher grade D-symbol for first and second language in matric. Jacobs (1987:33) deduces from the above that students who performed well at school level have developed a study attitude and ability to work, which enable such a

student to perform well at higher education level.

On the opposite side, many researchers "have expressed grave misgivings about the reliability and validity of school performance as predictors, particularly in the case of black students" (Bokhorst, Forster and Lea, 1992:60). Among these are the Committee of University Principals (1987: 79) who noted the discrepancies within the differentiated educational systems and concluded that school examination results are not always a reliable reflection of a student's academic potential.

Kotze et al (1996:39) states that it was found that matriculation results from the previous Department of Education and Training and equivalent school systems, particular at the lower ranges, are an inaccurate reflection of student's academic ability. Zaaiman (1998:62) supports this by saying that school results are seen as being greatly dependant on educational opportunity and therefore refer more to the opportunity to learn than the ability to learn.

6.3 Learning potential

According to Kotze et al (1996:40) higher education institutions in South Africa have responded in different ways to the issue of alternative selection. One such framework that has been developed is that of the determination of learning potential. According to the aforementioned authors, this paradigm seeks to assess the ability of an individual to respond to actual learning challenges, by the inclusion of a teaching/learning element in the selection procedure.

Zaaiman (1998:64) adds that a dynamic testing framework usually consists of a test-teach-test situation and is aimed at evaluating the student's ability to benefit from instruction. The pretest provides data on the student's current level of functioning on a given task. Instruction is then given in order to teach the student

the necessary problem-solving skills. The student's level of functioning is assessed during a post-test session. An improvement in the post-test result is seen as an indication that learning has occurred.

According to Zaaiman (1998:64), dynamic testing instruments are often seen as more fair to use for selection in situations of unequal educational opportunity, as the intervention phase gives the student who has not had the opportunity to develop his or her academic potential a better chance to achieve a fair result.

In studies conducted at the Rand Afrikaans University (RAU) and the Potchefstroom University for Christian Higher Education (PU for CHE), Kotze et al (1996:51) endeavoured to determine the extent to which learning potential could be used to predict academic performance in first-year higher education studies. For the purpose of the investigation, the **Ability, Processing of Information and Learning Battery (APIL)** as developed by Taylor(1991), was used. According to Kotze et al (1996:63), the aim of this test is not to test the student's previous learning experience, but to determine the extent to which the individual has the capacity to acquire new concepts and skills of a cognitive nature, to process the information and apply it.

At the RAU the academic results of 2 336 students were correlated, by means of the Pearson product-moment coefficient, with the APIL scores obtained. Significant correlations were found between the total APIL scores and the academic performance of students in all faculties included in this study.

However, at the PU for CHE, a multiple regression analysis was performed on the APIL data, with first-year academic results as criterion predictor. The APIL was not found to make a meaningful contribution to the explained variance in the academic performance of first-year students (Kotze et al, 1996:79).

This finding is supported by Zaaiman (1998:65) who states that research results have not shown a significantly higher predictive validity for learning potential tests than for previous academic performance. She adds that a dynamic selection procedure is also not practical when a large number of applicants have to be evaluated, as it is resource intensive and time consuming.

6.4 General cognitive ability and specific aptitude

The concept of cognitive or mental ability has been central to the development of psychology over the past century. Despite this, the concept is not well defined and controversy surrounding it is rife (Van Aswegen, 1997:23). According to Zaaiman (1998:62), ability is usually defined "in general terms that refer to a wide range of mental capacities affecting all mental operations." She also states that, in the context of selection, ability tests usually refer to the traditional intelligence type tests.

Thorndike (1986:332) states that ability testing, in its beginnings, focused on providing a measure of general cognitive functioning. Zaaiman (1998:62) adds that ability tests are non-subject specific and gives the following examples of ability tests:

- verbal
- numerical
- pattern recognition
- spatial reasoning

The usefulness of assessing general ability as predictor of performance has widely been debated. In 1986, the Journal of Vocational Behavior (Gottfredson, 1986) devoted an entire issue to the topic of general ability and its role in the prediction of task performance. In this publication a number of distinguished psychologists

address the question of whether or not there is an association between mental ability and job performance. The conclusion that is reached in this volume is that general cognitive ability (or "g" as it is also referred to) emerges as the single most useful task performance predictor.

According to Zaaiman (1998:62), the opposite seems to be true within the academic sphere. She states that a consistent research finding is that previous academic achievement has the highest predictive validity, with traditional aptitude tests next and ability measures having the lowest predictive validity. A possible explanation for this could be that ability tests measure concepts that are only hypothetically related to academic achievement and are thus further away from the task to be predicted than either achievement or scholastic aptitude tests. Furthermore, academic performance usually constitutes the first predictor, but due to its high correlation with ability tests, the ability tests are not included in the prediction model. Suppressor variables are then included as second and third predictors in the prediction model. Zaaiman (1998:62) continues that one could expect that a minimum level of prior academic achievement would be necessary for a student to benefit from an academic programme, regardless of inherent ability.

From the early concept of general ability a movement towards more specialised ability tests developed and, in time, the rationale and procedures for combining the results of such tests, using the statistics of multiple regression to create task-tailored batteries were put forth (Thorndike, 1986:332). " Following up on the theoretical emphasis on specialized abilities, tests of such special abilities began to multiply and aptitude test batteries began to replace general ability tests" (Thorndike, 1986:333).

As early as 1937 Bingham defined aptitude as "...a condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with

training some (usually specific) knowledge, skill or set of responses such as the ability to speak a language, produce music etc" (1937:16).

Zaaiman (1998:61) says that scholastic aptitude tests are designed to evaluate subject-related skills, comprehension and insight, utilising non-routine problems which require little subject-specific knowledge.

Contradicting research findings as far as scholastic performance is concerned have been reported by researchers looking into the predictive validity of aptitude tests. Skuy, Zolezzi, Mentis, Fridjhorn and Cockcroft (1996:113) found that the **Patterns Relation Test** of Barker, which is a test of inductive reasoning and reasoning by analogy, predicted performance in Accounting for a group of students in the Faculty of Commercial Science at the Witwatersrand University, significantly (R^2=0.38; $p<0.05$).

Pienaar (1991:119) concluded that specific subtests of the Academic Aptitude Test (developed and standardised by the Human Sciences Research Council) differentiate between successful and unsuccessful learners at higher education level. He investigated the predictive validity of the Non-Verbal Reasoning, Verbal Reasoning and Reading Comprehension subtests and found that the academically successful students performed, on average, significantly better on all of the mentioned subtests than the academically unsuccessful students.

In a study conducted at the RAU, the Senior Aptitude Test (developed and standardized by the Human Sciences Research Council) was used to determine the profile of a potentially successful student in social work (Van Zyl, Terblanche & Jacobs, 1992:33). In this study, no significant difference could be found between the Senior Aptitude Test performance of successful students and that of unsuccessful students.

Kotze (1994:64) refers to various studies regarding the predictive validity of the SAT in academic environments, which all conclude that the SAT does not differentiate significantly between potentially successful and at risk candidates.

Samkin (1996:118) reports on a study conducted at the University of Durban-Westville where the performance of first-year Accounting students were predicted, using final school marks as well as subtests of the Academic Aptitude Test (Human Sciences Research Council, 1974). In this study it was revealed that the results of the English Reading and Number Comprehension subtests combined explained 14.4% of the variance of the final Accounting result for the total group of students. "For HED students both English Comprehension and Number Comprehension are individual reliable predictors of success in the first year. However, the results of English Comprehension and Number Comprehension cannot be used as reliable predictors of success for DET students" (Samkin, 1996:118).

6.5 Biographical factors

Various references can be found to the influence of biographical factors such as age and gender on the performance of students on higher education level are found in the available literature.

In this regard, Zaaiman (1998:67) states that gender inequality with respect to access to higher education is of international concern. Brusselmans-Dehairs and Henry (1994:353) wrote the introduction to a volume of the International Journal of Educational Research which is dedicated to the gender differences in mathematical abilities. There is general consensus that boys generally perform better than girls in Mathematics and Science at high school level.

As far as student age is concerned, Zaaiman (1998:68) declares that this might be a factor of importance in the South African context, as many pupils take longer than the norm to complete their schooling and are thus older than 18 or 19 years when applying for admission to higher education. In 1993, according to Zaaiman (1998:68), 37% of the total number of black candidates for the matriculation exam was 22 years and older. She refers to a trend in the UK where older applicants usually come from underrepresented groups in terms of ethnicity and social class. "In the UK the typical mature applicant is aged 21-25, with lower A-level qualifications often gained over several attempts" (Zaaiman,1998:68).

As this study is concerned with the development of a **selection battery** for first-year engineering technician students, biographical variables such as age and gender will not be included since it would be unconstitutional to deny applicants access to further studies on the grounds of their age and/or gender.

It can be seen from this overview that a large number of factors could influence academic performance. All of these cannot be tested or practically implemented in the process of selection, and not all will add to the fairness or predictive validity of selection procedures, but one of the aims of selection research is to identify the best possible combinations of predictors for future performance. Such combinations, specifically developed for the prediction of performance by engineering and engineering technology students will subsequently be discussed.

6.6 Predictors of success in engineering and engineering technician courses

In Chapter 2 reference was made to the statement of the Department of Education in its White Paper on Higher Education (1997:8) that there exists a

mismatch between the output of higher education and the needs of a modernising economy. A particular shortage in fields of science, engineering, technology and commerce is mentioned.

To higher education circles this remark came as no surprise. As early as in 1983 Smit(1983:1) stated that South Africa experienced general shortages in high-level artisans and acute shortages in engineers and engineering technicians. He added that in the midst of these shortages attrition rates, especially during the first year of training, remained alarmingly high.

In the sixteen years that have followed, research into the predictors of academic success in engineering and engineering technician courses continued. Those important for this study will be highlighted here.

In a study commissioned by Technikon Pretoria, Smit (1983:62) used a sample of 324 engineering technician students in order to determine predictors for academic success in Electrical, Mechanical and Civil Engineering technology courses. He examined the role played by scholastic performance in Mathematics, Science and Afrikaans (at that stage the medium of instruction at Technikon Pretoria). He further included scores on all of the subtests of the Senior Aptitude Test (SAT) as well as biographical information gathered by means of a questionnaire.

He found that performance in Afrikaans and Mathematics at school level played a prominent role in the prediction of success in technician training (keeping in mind that Afrikaans was the medium of instruction at that stage). Even though performance on subtests of the SAT contributed to a higher predictive validity it was not statistically significant. Non of the biographical factors (such as age, marital status, sex and home language) played a meaningful role in the prediction of students' academic success.

Fourie (1988:77) looked at the scholastic performance of first-year engineering students at the Rand Afrikaans University as predictors of their success at higher education level. Performance in Mathematics was found to be an important predictor of success in that particular group. It was expected by the researcher that mother-tongue performance at matric level would play a significant role in the prediction of higher education success. However, the results of the study showed that the performance in the home language in the final school exams did not predict the academic performance of first-year engineering students at the RAU. It further seemed that the Standard 9 final marks were a more accurate predictor of tertiary success than final matriculation results. It has to be mentioned that this sample was totally homogeneous and that the findings can therefore not be generalised.

Rutherford and Watson (1990:354) stated that selection procedures at the University of the Witwatersrand needed redesigning, since the use of final matriculation marks as sole selection criteria for students in the science and technology fields was unfair and ineffective. They wanted to develop a selection procedure that was acceptable to all, and at the same time admitted only students with a reasonable chance of graduating.

To achieve these objectives, it was deemed necessary to supplement matric results with further information about the candidates.

> "It was felt that strictly content tests, while standardizing achievement measures across the entire spectrum of applicants would further disadvantage the disadvantaged and so measures of ability which had been found elsewhere to be relevant to success in the sciences were sought" (Rutherford & Watson, 1990:355).

They identified the following factors as occurring frequently in relevant literature,

and measures thereof was included in the testing battery :

- Piagetian development level
- Spatial ability
- Maths and English ability

The Science Aptitude Test (Atlink, 1987), on close scrutiny, revealed many Piagetian-type questions - the principles of the conservation of mass and volume are, for example tested. Proportional reasoning is furthermore necessary to answer many of the questions (Rutherford & Watson, 1990:356).

The authors stated that, although the test was not purely a Piagetian questionnaire and did not allow deductions regarding the validity of Piagetian development levels in the selection for science courses, the fact that it contained Piagetian concepts that had been linked to performance in science gave it construct validity.

The second factor identified by Rutherford and Watson (1990:356) as being relevant to success in studies in the sciences was spatial ability. The Rotate and Flip Test, developed by the NIPR as a test of spatial visualisation, was selected to form part of the selection battery. In this test, the subject is provided with sections of a figure and has to manipulate them mentally by means of rotating and flipping them to form a given whole (Rutherford & Watson, 1990:356).

The third factor included into the selection battery by Rutherford and Watson (1990:356) was that of Mathematical and English ability - generally linked to success in the sciences. The authors added that competence in English was an obvious prerequisite to success in this particular course, as it was taught in English.

The Mathematics and English factors were combined in the selection of the third

test of the selection battery, namely, the Embedded Problems Test, also designed by the NIPR (Rutherford & Watson, 1990:356). The researchers criticised their own decision by saying that the Embedded Problems Test was not an ideal test as "the scoring does not allow credit for disembedding ability, English ability and Mathematics ability to be calculated separately, as only a global score on the final answer is given" (Rutherford & Watson, 1990:356).

In this study, final year-marks of the students in each subject were used as dependent variables and variance was studied for individual subjects rather than overall first year pass or fail, as it was felt that individual predictors might contribute differently for different courses (Rutherford & Watson, 1990:357).

The matric percentages obtained in English, Mathematics and Physical Science combined, explained 17-26% of the variance in the various first-year subjects. A multiple regression analysis was performed, using all of the mentioned predictors, and in this combination the variance accounted for ranges from 26-33% (Rutherford & Watson, 1990:358). The authors concluded that, the use of the selection battery in addition to the matric results, did increase the amount of variance in first-year grades explained, and have therefore increased the predictive validity of the selection process. They acknowledged however, that the particular tests used in the selection battery was not the best available and that first-order correlations of the results of these tests with end-of-the-year results proved insignificant. They admitted that new tests were being sought (Rutherford & Watson, 1990:358).

Smith (1992:5) proposed a model for the selection of civil engineering technician students at the Cape Technikon. In this study she randomly selected 145 students that were representative of the students in Civil Engineering at the Cape Technikon. She selected the following psychometric test battery:

- Senior Aptitude Test (SAT)
  - Verbal Reasoning
  - Three-Dimensional Spatial Perception
- Academic Aptitude Test (AAT)
  - Verbal Reasoning
  - Three-Dimensional Spatial Perception
- Gottschaldt Figures Test (for deductive reasoning)
- 16 Factor Personality Questionnaire (16PF)
- Self-Directed Research Questionnaire of Career Interests (SDS)
- Achievement Motivation Questionnaire (AMQ) by Pottas et al
- Job Satisfaction Index (JSI) by Brayfield and Rothe, adapted by Mauer
- A self-developed biographical questionnaire

Table 6.2: Correlations indicating factors with significant relation to the academic success of students in Civil Engineering (adapted from Smith, 1992:14)

Factor	Correlation r	p
Matric Mathematics	0.6183	0.0037**
Matric Physical Science	0.2731	0.0242*
Matric Mathematics and Physical Science	0.3607	0.0025**

Matric Mathematics, Physical Science and Languages	0.2746	0.0235*
Matric Mathematics, Physical Science and First Language	0.2748	0.233*
Matric Total Aggregate	0.5335	0.0044**
Three-Dimensional Spatial Perception (Raw score)	0.2259	0.0401*
AMQ (B)	0.3112	0.0039**
AMQ (AA)	0.3037	0.0326*
AMQ (BB)	0.2283	0.0404*
AMQ (PM)	0.3093	0.0050**
16 PF (E)	-0.2443	0.0432*
16 PF (I)	-0.2585	0.0336*
16 PF (L)	-0.3278	0.0210*
16 PF (N)	0.2703	0.0445*
16 PF (Q1)	-0.2621	0.0127*
16 PF (Q2)	0.6282	0.0016**

* ***Significant on 5% level of meaning***

** ***Significant on 1% level of meaning***

In Table 6.2, the factors that correlated significantly with the academic success of the sample are given. The matriculation results were converted to Swedish Formula scores by the method shown in Table 1 earlier in this chapter. In this study, Smith (1992:12) correlated the school results and psychometric data obtained with the final third-semester exam marks.

REMARK:

It is important to note that the sample for this study consisted of 91.7% male and 84.5 % White students (Smith, 1992:9). The sample is thus homogeneous and not representative of the national population.

From the results it is clear that, in academic terms, performance in matric Mathematics correlated the highest with the students' overall higher education performance (r=0.618; p=0.004). Furthermore, the total matric aggregate also correlated significantly (r=0.534; p=0.004) with the academic performance of the students.

As far as aptitudes are concerned, it seemed that only three-dimensional spatial perception, of those utilised, correlated significantly with the academic performance of the students (r=0.226; p=0.040). According to Smit (1996:234), the Spatial 3-D subtest on the Senior Aptitude Test measures primarily a "visualisation factor in which three-dimensional visualisation plays a role."

According to the results as shown in Table 6.2, four factors on Pottas, Erwee, Boshoff and Lessing's **Achievement Motivation Questionnaire** (1988) correlated statistically significantly with the academic performance of the students included in the study. Smit(1996:369) states that the AMQ was developed to, amongst others, measure the intensity of the motivation to achieve in adults.

> "Individuals who are achievement motivated strive to give of their best in whatever they undertake. They possess high personal standards of excellence and rely on their own abilities and skills to achieve success. In order to realise their goals, they persevere diligently, are action-orientated, and tend to manage time effectively and economically" (Smit, 1996:369).

In this particular study, the scores obtained for ***Time Consciousness*** (B) correlated the highest with the students' academic performance (r=0.311; p=0.003). Smit (1996:370) says that this subfactor indicates a person who plans and schedules his or her work. It furthermore seems that the score for ***Goal Orientation*** (AA) correlated significantly (r=0.304; p=0.032) with the academic performance of the students. People with high scores on this subtest tend to persevere despite obstacles and try to deal with matters as quickly as possible (Smit, 1996:370). In addition, the performance on the ***Personal Effectiveness*** subtest also correlated with the students' final exam marks (r=0.228; p=0.04). According to Smit (1996:370), high scorers on this factor usually set themselves very high standards of personal performance and have full confidence in their own abilities. The total ***Performance Motivation*** score also correlated significantly (r=0.309; p=0.005) with the academic performance of the respondents. Smit (1996:371) adds that it is advisable to take note of Factors AA and BB, as well as Factor PM, when using this questionnaire for selection purposes.

From Table 2 it can be seen that six of the factors of the **16 Factor Personality Questionnaire** (16PF) correlated significantly with the academic performance of the investigated group. The 16PF is claimed by Cattell (1961:2) to measure the total personality and was adapted by the HSRC for South African conditions and thus standardised (Smit, 1996:280).

The six factors or personality dimensions which correlated significantly with the academic performance of the students in Smith's study are briefly discussed below.

- **Factor E: Submissiveness versus Dominance**

  According to Smith (1992:14), Factor E showed an inverted correlation with the academic performance of students in her study (r=-0.244; p=0.043). This implies that the lower the score on Factor E, the better the academic

performance. According to Smit (1996:292), a person scoring low on this factor might tend to be obedient, accommodating, considerate and conventional. An explanation for this correlation could be that these students were conscientious and therefore more successful.

- **Factor I: Tough-mindedness versus Tenderheartedness**

The relation between the score obtained on Factor I by the students is again inverted to their academic performance (r=-0.259; p=0.033) (Smith, 1992:14). Smit (1996:294) states that "Persons who score low on this factor are realistic, accept responsibility, act on practical, logical grounds and are not artistic." In this case it seems as if the responsible, realistic students are the better performers in engineering courses.

- **Factor L: Trust versus Mistrust**

The students' score on Factor L of the 16PF correlated significantly with their academic performance (r=-0.329; p=0.021) - again the relation is inverted. The lower the score on Factor L, the better the academic success. Smit (1996:296) states that people with a low score on Factor L tend to be relaxed and easygoing as well as trusting and accepting. It could be possible that these students coped better with the pressure of higher education studies.

- **Factor N: Artlessness versus Shrewdness**

Smith (1992:14) reports a positive correlation between scores obtained on Factor N and the academic performance of engineering students (r=0.270; p=0.0445). Smit (1996:297) refers to people scoring high on this factor as being shrewd, emotionally detached and showing insight into self and others. This finding is not supported by previous findings, which revealed a negative relationship between this factor and success in education (Smit, 1996:297).

- **Factor Q1: Conservatism versus Radicalism**

  The relationship found by Smith (1992:14) between Factor Q1 and the academic performance of engineering students is inverted (r=-0.262; p=0.013). According to Smit (1996:298), persons measuring low on this factor are conservative and respect established ideas. This finding is again contrary to the expected, as research has shown that people with high scores on Factor Q1 manifest intellectual interests and tend to experiment.

- **Factor Q2: Group dependancy versus Self-sufficiency**

  Smith (1992:14) reports a significant correlation between Factor Q2 and the academic performance of the students taking part in her study (r=0.628; p=0.002). Smit (1996:298) sees people who score high on this factor as self-sufficient and resourceful; they furthermore tend to avoid social life, since they regard it as a waste of time. This may be the reason for their better academic performance.

Even though the sample used in this study was not representative of the general population and the results obtained therefore not generaliseable, some of the findings could serve as an indication of useful constructs that could be further investigated by this study.

Zaaiman (1998: 71) conducted a study in collaboration with the Free University of Amsterdam in which focus was placed on the selection of disadvantaged students for the ***University of the North Foundation Year*** (UNIFY). This foundation year is specifically aimed at preparing students for scientific and engineering courses by developing their Mathematical and Physical Science skills. According to Zaaiman (1998:74), 150 students are admitted to this course every year. Many more students apply and are evaluated. In 1993, 189 applicants were tested; from 1995 to 1997 the numbers varied from 600 to 700.

Zaaiman (1998:75) states that it was decided that the University of the North would use its own selection tests because of the uncertainty regarding the validity of the DET matric results for the prediction of future academic performance. These selection tests would be used to identify potentially successful students.

Critical performance areas were identified by relevant academic staff and test specifications were developed to match these. The test specifications were then used to develop the selection tests and procedures (Zaaiman, 1998:86). The areas that were assessed included the following:

- Mathematic skills (as evaluated by the UNIFY Mathematics Selection Test or UMST);
- Science skills (as evaluated by the UNIFY Science selection Test or USST);
- English proficiency (as evaluated by the UNIFY English selection Test or UEST);
- Basic arithmetic skills (as evaluated by the UNIFY Arithmetic Selection Test or UAST).

In this study, Zaaiman (1998:93) used the final average score obtained by students from the University of the North's Foundation Year as criterion for the calculation of the predictive validity coefficients. The year mark was used in cases where students were not allowed to write examination on the grounds of poor performance during the year.

Table 6.3 shows the predictive validity coefficients of the results of the selection tests for the final Foundation Year performance. (An initial English proficiency test was included in the table, since performance in this test was used to decide whether English language skills should be included in the test battery or not.)

Table 6.3: Predictive validity coefficients of the selection tests used at the University of the North's Foundation Year as calculated with the final academic performance as criterion (adapted from Zaaiman, 1998:93)

Predictor	1994 r	1995 r	1996 r
UNIFY Mathematical Selection Test	0.44**	0.39**	0.41**
UNIFY Science Selection Test	0.27**	0.38**	0.44**
Average Maths and Science	0.49**	0.49**	0.57**
UNIFY Arithmetic Selection Test	not available	not available	0.28**
English proficiency test	0.28**	0.27**	0.26**
UNIFY English Selection Test	not available	not available	0.30**

*** two-tailed significance less or equal to 0.01*

Further investigation showed that the best combination of predictors, as determined by a multiple regression analysis, were the Mathematics, Science and English Selection Tests. In 1996, the multiple R obtained by these predictors was 0.59, which was significant at $\propto = 0.01$ (Zaaiman, 1998:94).

The constructs identified in Zaaiman's study do in some instances correlate with those in Smith's study discussed earlier in this chapter, even though the type of assessment differed greatly. The role of all these constructs will be investigated thoroughly in this study.

6.7 Conclusion

From the literature reviewed it is clear that, although scholastic performance is shown as the single predictor with the highest predictive validity, a large portion of prospective students are placed at a severe disadvantage if this is used as sole selection criteria. As the majority of the applicants for the engineering technician courses at Technikon Pretoria come from the previously disadvantaged communities, and school performance is currently used as minimum admission requirement, the predictive validity of school performance as predictor of higher education performance will again be looked into.

The literature reviewed also revealed that the evaluation of learning potential does not significantly increase the prediction of academic success on higher education level. Furthermore, it is an impractical evaluation to carry out when large numbers of applicants have to be assessed. For these reasons, the evaluation of learning potential will not be included into the current study.

According to the literature studied, biographical factors have not been proved to be accurate predictors of academic performance at higher education level. However, it should be kept in mind, that it is difficult to subject these factors to statistical analysis, especially when the more conventional statistical techniques are used. It is thus possible that these results were not as accurate as they could have been. But, since this study is concerned with the development of a selection battery for first year engineering technology students, biographical variables such as age and gender cannot be included. It would be unconstitutional to deny applicants access to further studies on the grounds of age and/or gender.

As some success in the prediction of students' academic performance by means

of ability tests has been documented, this study will concern itself with the identification and determination of the predictive validity of a test of this kind. Chapters 8 to 11 will deal with the different aspects of the research methodology followed in executing the study. Chapter 7 will look into the specific psychometric test used in the study.

Chapter 7

Potential Index Batteries

7.1 Introduction

Since some success in the prediction of students' academic performance by means of ability tests has been documented (see Chapter 6), it was decided to use a test of this kind in this study to determine the potential for success of prospective students in the engineering technician courses at Technikon Pretoria.

It was decided to make use of the Potential Index Batteries (PIB), developed by Erasmus and Minnaar for the purpose of this study. The aforementioned instrument had previously been purchased by the Department of Student Counselling at Technikon Pretoria and all the student counsellors had been trained by the developers and were familiar with the testing procedures. The Department of Student Counselling was satisfied that the initial research yielded sufficient evidence towards the instrument's culture fairness and compliance with all relevant regulations and legislation.

In this chapter, the nature, purpose and relevant sub-indices of the instrument will be discussed in detail. Reference will be made to studies carried out on the validity and reliability of the instrument.

7.2 The nature of the Potential Index Batteries (PIB)

According to Erasmus et al (1997:2), PIB can be described as "an instrument for the screening of the potential of literate, semi-literate, literate and academically advanced persons."

According to the test developers, the principle according to which the test was developed, was that **the ability to perform a simple task, indicates the potential to master a related, but more complicated task.**

The instrument consists of three principle sections, namely, a Comprehensive Structured Interviewing for Potential system (CSIP) as well as a visual or non-verbal (V-PIB) and a verbal (PIB) assessment section. Each of the latter sections consist of smaller units or indices - each measuring a specific competency. In total, the instrument can evaluate a person's potential regarding 65 different fields or competencies. The fields relevant to this study will be discussed in detail later in this chapter.

The Comprehensive Structured Interviewing for Potential system (CSIP) is used in the execution of a job analysis in order to compile a comprehensive job description. According to Erasmus (1997:46), the competencies relied on by CSIP were identified and generically defined from 1993 to 1995, in a joint project of the developers (Potential Index Associates) and the National Productivity Institute. By using the Job Specification Index it is possible to establish the most important predictors necessary for success in a particular job or task.

> "A systemized procedure leads to the ranking of the 65 competencies of the CSIP in their order of importance for performance in a particular job. It then proceeds to select a predetermined number of competencies - a minimum of 8 competencies are suggested here - which are regarded as the most vital ones for success in this particular position" (Erasmus, 1997:47).

According to the developers this procedure customises the assessment

programme for each specific position, as candidates are only assessed on competencies vital for success in the particular position (Erasmus, 1997:47). It is furthermore essential that the persons involved in determining the relevant competencies should be familiar with the job (task) being analysed. The developers suggest that at least three, but preferably five or six people, should participate in the task. As soon as the most important competencies have been identified, the corresponding indices from the assessment sections (V-PIB and PIB) are included in the battery used for selection purposes.

7.2.1 Description of the relevant individual indices of the V-PIB and PIB

The following indices were identified as important for the prediction of the academic performance of first-year engineering technician students at Technikon Pretoria. The descriptions for the V-PIB indices are according to the User's Manual (Erasmus & Minnaar, 1995:1-6) and that of the PIB according to Erasmus et al (1997:17-20).

7.2.1.1 Visual Potential Index Batteries

- **Index 3 - Numerical Potential**

  This index comprises basic numerical calculations and is aimed at the respondent's mainly verbal potential to calculate. It is expected of the testee to reason along logical, mathematical lines, to apply reduction and induction, and it demands the capacity to add, subtract, divide and multiply.

- **Index 4 - Composition of Wholes**

  In this index, the testee has to synthesise small parts into an organised, integrated whole. It constitutes an important non-verbal component of potential. This capacity seems to involve both visual perception and logical reasoning. It furthermore demands a capacity for concrete

reasoning in terms of figures and forms and their logical place within a single structure. The testee has to be able to visualise the desired end-result in order to be able to perform the task.

- **Index 6 - Spatial Reasoning**

  In this index, the respondent is instructed to identify two identical illustrations amongst five almost similar ones. The capacity to perceive, analyse and synthesise is assessed, as well as the ability to select and categorise.

- **Index 7 - Perception**

  The respondent is expected to identify the one illustration that is different from four similar illustrations. This index assesses the potential to perceive detail as well as wholes in their specific, logical and sensible context.

7.2.1.2 Potential Index Batteries

- **Index 2 - Creativity**

  Creativity refers, in this context, to a person's need, willingness and ability to create new concepts and solutions to problems, rather than sticking to old and tried ways.

- **Index 3 - Reading Comprehension**

  The respondent is given five minutes in which to read a passage and is then asked questions on the contents of the passage. He or she is not allowed to return to the passage once the questions are being answered. This index assesses the testee's competency to read and understand clearly what the reading matter conveys.

- **Index 5 - Mental Alertness**

  This index requires the respondent to identify the sole deviation from five seemingly related objects. This ability is, according to the developers,

associated with the ability to classify objects correctly. Schaap (1997:70) refers to the fact that this index is also associated with the general mental ability (or G-factor) of the testee.

- **Index 12 - Vocabulary**
  The respondent's basic English vocabulary is assessed in this index. He or she is expected to select, from five possible answers, a synonym for a specific word given.

7.3 Validity and Reliability of PIB

The developers of PIB take great pride in the fact that they have made the instrument available for scrutinising by both its users and experts from the academic sphere. Studies for determining the validity and reliability of PIB have been undertaken by various persons from various institutions.

Correlations between PIB indices and other standardised psychometric tests were found to be high. In this respect, Schaap (1997:72) found correlations between PIB and the 16 Personality Field Questionnaire as high as 0.70 ($p = 0.01$).

In a run-up to the study done here, correlations were calculated between the PIB indices and the Academic Aptitude Test (AAT). An accidental sample of 500 psychometric records of prospective students was used for this study. According to Dane (1990:302), an accidental sample is based on availability or ease of inclusion. In the case of this study, these records were available at Technikon Pretoria, due to an organisational search for suitable psychometric instruments for use in selection.

It was found that Index 4 of the V-PIB (Composition of Wholes) correlated

significantly with Subtest 1 of the AAT (Non-Verbal Reasoning) and with Subtest 9 (Spatial Perception). In the case of Subtest 1, a correlation of 0.49 ($p = 0.01$) was found and in the instance of Subtest 9, a correlation of 0.48 ($p = 0.01$) was found. Index 6 (Spatial Reasoning) of V-PIB correlated at 0.39 ($p = 0.05$) with Subtest 9 (Spatial Perception) of the AAT. PIB's Index 5 (Mental Alertness) correlated with the AAT's Subtest 1 ($r = 0.58$; $p = 0.01$), Subtest 2 ($r = 0.71$; $p = 0.01$) and Subtest 9 ($r = 0.46$; $p = 0.01$).

The conclusion arrived at from the above results was that the instrument did show construct validity in previous studies and could therefore be used in this study. The calculation of the situation-specific validity for the current study will be discussed in a later chapter.

In a study executed by Schaap (1997:4), reliability coefficients of as high as 0.93 were found for Index 5 (Mental Alertness) of the PIB. Pilot studies undertaken by Technikon Pretoria rendered reliability coefficients for indices ranging from 0.69 to 0.96. The determinating of situation-specific reliability coefficients, relevant to the indices used in this study, will be discussed in Chapter 12.

7.4. Conclusion

Two factors influenced the decision to use the Potential Index Batteries in the development of a selection battery for engineering technician students at Technikon Pretoria. Firstly, the availability of the instrument was important, but even more important was the fact that the instrument was developed in South Africa and regarded as culture friendly. This made the instrument acceptable to student bodies and the Technikon Management, and also acceptable from a professional and legal point of view.

In Chapter 8 the methodology used in the research process will be discussed.