Chapter 8
Research Methodology

8.1 Introduction

As mentioned in Chapter 1, the objective of this study is the development and validation of a selection battery for first-year engineering technician students at Technikon Pretoria.

Possible predictor variables have to be identified, and, in order to determine the predictive validity of these predictor variables, a correlation has to be drawn between predictor scores and criterion scores. The predictive validity thus calculated will be an indication of the accuracy with which the predictors predict criterion performance.

In the following four chapters, attention will be given to the discussion of the research methodology followed in this study. Aspects which will be discussed in detail in this chapter will be the research design and the data collection techniques used in the study. In Chapter 9, the population from which the sample for this study was drawn will be illustrated, together with the sampling method and the sample itself. Chapter 10 will deal with the statistical techniques used in the data analysis process. Chapter 11 will explain the determination of the predictor variables.

8.2 Research Design

According to Guy, Edgley, Arafat and Allen (1987:92), a research design can be defined as "the plan of procedures for data collection and analysis that are undertaken to evaluate a particular theoretical perspective." McMillan and Schumacher (1993:157) add that research design refers to a plan for selecting subjects, research sites and data collection procedures to answer research questions. They emphasise that the goal of a sound research design is to ensure results that are credible.

The design selected for this study is a non-experimental, correlational design (Smit, 1983:70; McMillan & Schumacher, 1993:270). According to Smit (1983:70), this research technique is considered the best controlled and most accurate of all non-experimental designs. This design specifically deals with the determination of the extent and intensity of covariation or interdependency between variables in natural circumstances (Smit, 1983:70).

McMillan and Schumacher (1993:270) distinguish between two types of correlational research, namely, simple ***correlational studies*** and ***prediction studies***. They define a simple correlation study as a research study where the researcher obtains scores "from two variables for each subject and then uses the pair of scores to calculate a correlation coefficient" (McMillan & Schumacher (1993:270). In a prediction study, one variable (**predictor variable**) is used to predict performance on a second variable (**criterion variable**) and the predictor variable is usually measured before the criterion variable (McMillan & Schumacher, 1993:271). A ***positive correlation*** means that high values on one variable are associated with high values on a second variable, whereas a ***negative correlation*** is found when a high value on one variable is associated with a low value on a second variable (McMillan & Schumacher, 1993:35).

Smit (1983:70) states that correlative techniques deal with variables already existing within the subjects' repertoire and **not** with variables generated experimentally. He continues that correlational research differs from experimental research in so far as that in correlational research no manipulation of an independent variable occurs, and therefore no distinction can be made between an experimental and a control condition.

Smit (1983:74) regards psychometric test construction and research as the most general use of the correlational technique as a research method.

8.3 Data collection

According to Guy et al (1987:112), data collection refers to the tool or vehicle through which measurement is actualised. They add that even though the research design does not dictate the data collection procedures, certain data collection techniques usually accompany certain designs.

Smit (1983:145) as well as McMillan and Schumacher (1993:40) distinguish between quantitative and qualitative techniques of data collection. McMillan and Schumacher (1993:14) state that, on one level, the terms refer to distinctions made about the nature of knowledge and on another level they refer to the research method - that is, how data are collected and analysed - as well as the type of generalisations derived from the data. "**Quantitative research** presents statistical results represented with numbers; **qualitative research** presents facts in narration of words" (McMillan & Schumacher, 1993:14).

Quantitative techniques are used with experimental, descriptive and correlational designs (McMillan & Schumacher, 1993:41). As the design selected for this study is correlational, the data collection techniques will subsequently be quantitative.

McMillan and Schumacher (1993:167) name the following as common methods of gathering quantitative information:

- ⌘ Questionnaires
- ⌘ Standardised interviews
- ⌘ Tests
- ⌘ Standardised observations
- ⌘ Inventories
- ⌘ Rating scales

In the current study, two methods of data collection were used. The first involved a questionnaire that was used to identify the variables that were to be included in the prediction study. McMillan and Schumacher (1993:42) refer to questionnaires as a technique that encompasses a variety of instruments in which the subject reacts to written questions.

The questionnaire used in this study is a standardised questionnaire that forms part of the Potential Index Batteries discussed in Chapter 7. It was decided to analyse the requirements the Engineering Technology training programme expected the incoming students to meet by focussing on the actual performance of students in the programme. As a starting point in this process, the Comprehensive Structured Interviewing for Potential system (CSIP) was completed by a group of lecturing staff from the Engineering Faculty at Technikon Pretoria. This group of respondents consisted of 12 members of the lecturing staff, who are all directly involved with the training of Engineering Technology students.

The questionnaires asked the respondents to select, from a list of 65 competencies the eight which are indispensable for the success of an

engineering technology student. Subsequently the respondents were asked to rate these eight competencies in order of significance. These responses were then computer- processed and the results indicated to the researcher which predictive variables were thought important for the success of students by academic staff concerned with the training of engineering technology students. In this manner, the competencies to be included in the potential assessment battery were selected.

For the assessment of these competencies, psychometric testing was selected (as discussed in Chapter 6). This was the second data-collecting method to be implemented by the researcher. According to McMillan and Schumacher (1993:42), "testing" refers to the use of tests scores as data. "This technique involves subject response to either written or oral questions to measure knowledge, ability, aptitude, or some other trait. A numerical value is obtained as a result of each subject's answers to a standard set of questions" (McMillan & Schumacher, 1993:42).

According to Smit (1983:169), methods can be regarded as objective when different scorers give the same numerical values to the testees when following a prescribed set of rules. Psychometric testing and its characteristics were discussed in detail in Chapter 4 and only a summary of the uses of objective psychological tests - as stated by Smit (1983:169) - will be given here.

8.3.1 Uses of psychological tests in research

a. Psychological tests can be used for selection purposes. In other words, a person can, on the grounds of his or her performance on a test, be included or excluded from a study, grouped in the experimental or control group, etc.

b. Psychological tests can be used to determine the effect of the

experimental/independent variable.

The psychometric test used in this study was the Potential Index Batteries (PIB) of Erasmus and Minnaar (1995). (This test was discussed comprehensively in Chapter 7.) Students who applied for either the National Diploma In Civil Engineering Technology or the National Diploma in Mechanical Engineering Technology were evaluated by means of the PIB. These assessment sessions were run by staff members of the Department of Student Counselling at Technikon Pretoria. All of these staff members are registered with the Health Professions Council of South Africa as either psychometrists or counselling psychologists. They have also been trained in the administering of the PIB by the developers themselves. The psychometric results of the students who had been admitted to these courses were included in the study, since academic records of these students, which could be used as criteria variables in the study, would be available. These academic results consisted of the final first-semester mark obtained by students in all their required subjects.

The final matric symbols obtained in Mathematics and Natural Science were obtained from the Academic Administration Division at Technikon Pretoria. The matric symbols were quantified according to the Swedish formula discussed in Chapter 6, and added to the equation as predictor variable.

8.4 Data analysis

According to Guy et al (1987:308), the intent of the data analysis phase in research is to examine the body of data for the hypothesised relationships. It is in this stage of the research that answers to the research questions are gained.

Guy et al (1987:309), as well as McMillan and Schumacher (1993:191,) states that quantitative research findings, as in the case of this study, are usually presented

in statistical form. The statistical techniques used in the data analysis phase of this study will be described in Chapter 10.

8.5 Conclusion

In providing an answer to the research objective stated in Chapter 1, a research design was implemented which would best suit the type of investigation under way in this study. A nonexperimental, correlational design was selected, as this research technique is considered the best controlled and most accurate of all nonexperimental designs (Smit, 1983:70). Since a quantitative technique was selected for data gathering, the necessity for a statistical method in the data analysing process was obvious. The results of the data analysis will be discussed in detail in Chapter 12.

Chapter 9

The sample

9.1 Introduction

One of the first steps when designing quantitative research is to choose the subjects who will participate in the study (McMillan & Schumacher, 1993:159). Theorists seem to agree that this sampling process constitutes one of the most important steps in a research project, as it will determine the generalisability of the research findings (Smit, 1983:178; Guy et al, 1987:174; McMillan & Schumacher 1993:160). In this chapter the population from which the sample for this study was drawn will be illustrated, together with the sampling method and the sample itself.

9.2 The population

Dane (1990:289) describes a population as "all possible units or elements that can be included" in a study or research project. In the current study, the population comprised all mechanical, electrical and civil engineering technology students at Technikon Pretoria from, 1997 to 1999, of whom psychometric records were available.

9.3 The sample

Guy et al (1987:174) defines sampling as a method of selecting some part of a

group to represent the total. The objective of sampling is the estimation of population values from the information contained in the sample. Theorists such as Smit (1983:182), Guy et al (1987:184) and McMillan and Schumacher (1993:160) distinguish between two major sampling procedures: probability and non-probability samples. In the case of probability samples, the researcher can specify for each unit in the population the likelihood that it will be included in the sample. With non-probability samples, there are units of the population that have no chance of being included in the study.

9.3.1 Factors to be taken into consideration when sampling

Smit (1983: 179) regards the following factors as crucial when selecting a sampling procedure:

- The statistical definition of the research question as well as the type of information needed.
- The definition of the population in order that elements belonging thereto can be easily identified. This enhances the validity of the sample.
- Similar surveys ought to be studied to ensure that information does not already exist.
- After the nature of the data required has been identified the best process for sampling should be selected.

9.3.2 The sample size

According to McMillan and Schumacher (1993:163), the researcher must determine the size of a sample that will provide sufficient data to answer the research question. These authors, as well as Guy et al (1987:196), mention that the general rule for determining sample size is to use the largest sample possible. McMillan and Schumacher (1993:165) continue that in the case of correlational

research - as in this study - a minimum of thirty subjects should be included. If the researcher expects to find small differences or slight relationships, they feel that it is desirable to have as large a sample as possible. A large sample is also needed when a study concerns itself with a variety of dependent and/or independent variables. In the case of this study, it will thus be preferable to have the largest sample that is available.

9.3.3 The sample for this study

For the purpose of this study, a non-probability convenience sample was selected. Smit (1983:195) states that this method implies that the researcher uses whatever subjects are available for the study. Guy et al (1987:189) states that the major deficit of this method is that there is no formal procedure for generalising from sample to population, since sampling error cannot be determined. On the other hand, using a convenience sample could save time and money - thus, what is lost in accuracy would be gained in efficiency. Smit (1983:195) concludes that this sampling method is totally satisfactory if the researcher does not plan to generalise findings to the broad population.

The sample for this study consisted of a total of 732 Engineering Technology students at Technikon Pretoria. From these, 512 were Civil Engineering Technology students and the remaining 220 were Mechanical Engineering Technology students. These subjects were the total number of students from these two academic departments, enrolled from 1997 to 1999, of whom both psychometric and academic data were available.

As depicted in Figure 9.1, the sample consisted of 14.75% female and 85.25% male respondents. Figure 9.2 shows the frequency distribution of the home languages spoken by the respondents. The respondents indicated twelve different languages (frequency shown in brackets):

English	(18.72%)
Afrikaans	(29.64%)
Zulu	(4.92%)
Ndebele	(2.73%)
Tsonga	(4.37%)
Northern Sotho	(14.89%)
Venda	(3.83%)
Tswana	(9.29%)
Xhosa	(2.32%)
Southern Sotho	(4.23%)
Portugese	(0.96%)
Swazi	(0.68%)
Other	(0.96%)

Of the respondents, 5.24% indicated qualifications higher than a National Senior Certificate. The average age of the group was 18.08 years with a standard deviation of 2.04.

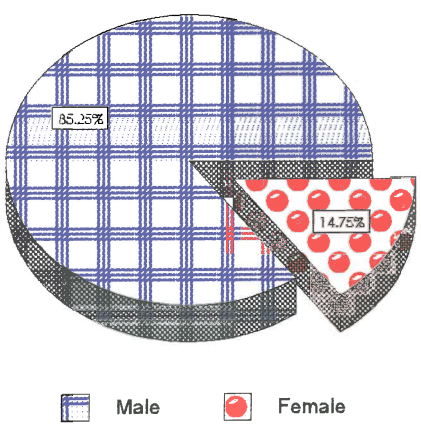

Figure 9.1: Frequency distribution of the gender of the sample used in the study

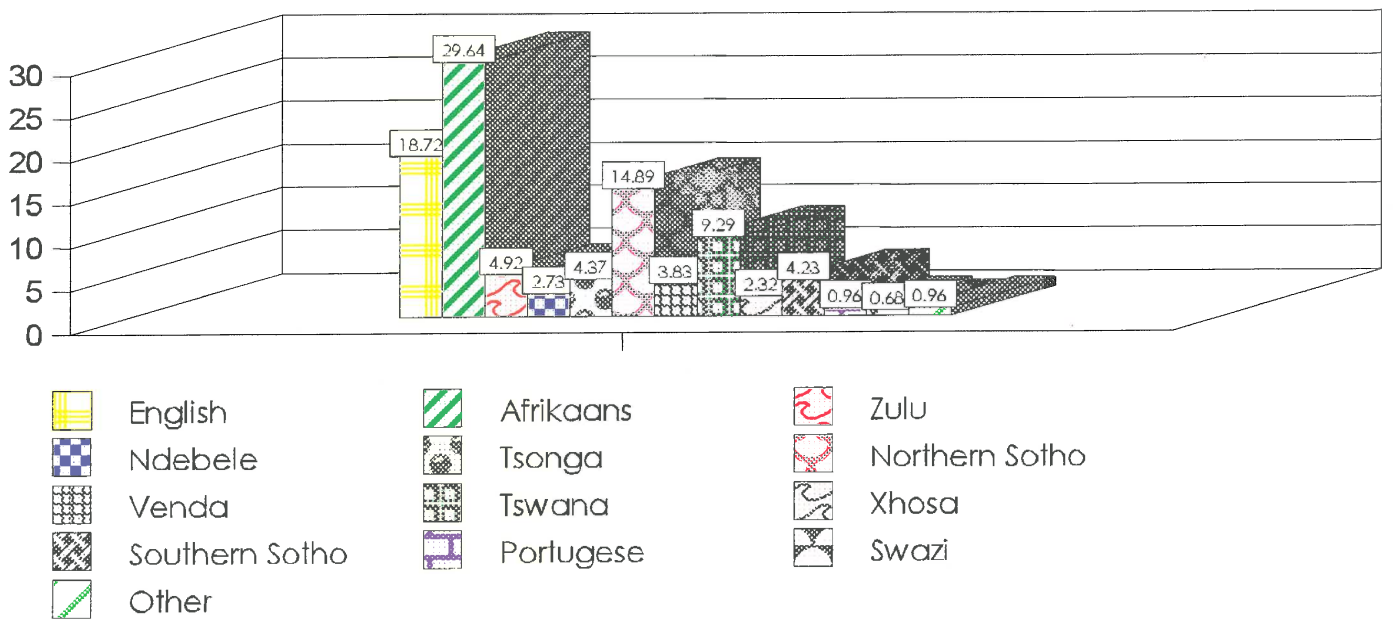

Figure 9.2: Frequency distribution of home languages spoken by respondents

9.4. Conclusion

As De la Rey (1976:159) states, sampling is decisive in the research process. It largely determines the success of a scientific study. In Chapter 9, the sampling method used in this study was described and motivated, and the sample itself was described. As a nonprobability sample was used, it was not possible to determine sampling error (Guy et al, 1987:196) and therefore as many subjects as were available were included in the sample. Chapter 10 will focus on the statistical techniques used in the data analysis process.

Chapter 10
Statistical techniques

10.1 Introduction

Guy et al (1987:309), as well as McMillan and Schumacher (1993:191), states that quantitative research findings, as in the case of this study, are usually presented in statistical form. Two major approaches comprise the subject matter of statistics, namely, **descriptive** and **inferential** (Howell, 1992:4).

McMillan and Schumacher (1993: 192) say that descriptive statistics transform a set of numbers and/or observations into indices which describe or characterise the data. Inferential statistics are used to make inferences or predictions about the similarity of the sample to the population from which the sample was drawn. The two approaches are interlinked and McMillan and Shumacher (1993:192) conclude: "...a researcher would first take a sample from a population, use descriptive statistics to describe the sample, and then use inferential statistics to estimate the true value of the test score for the population."

According to Dane (1990:243), statistical analyses used for predictive research are generally based on correlation, that is, the measure of the extent to which two or more variables are related. These methods of analysis, inter alia, include correlation coefficients, simple regression analysis and multiple regression analysis.

The descriptive and inferential techniques used in this study will subsequently be discussed.

10.2 Descriptive statistics

As mentioned in the introduction, descriptive statistics transform a set of numbers and/or observations into indices which describe or characterise the data. For the purpose of this study, measures of central tendency, measures of variability and frequency distribution are important.

10.2.1 Measures of central tendency

Bless and Kathuria (1993:35) indicate that measures of central tendency are values usually used to give a general description of a bulk of data. Theorists distinguish between three principle measures of central tendency, namely, mode, median and mean (Bless & Kathuria,1993:35; Howell, 1992:31-32; Popham & Sirotnik, 1992:14-16).

10.2.1.1 The mode

The mode is the score that occurs most frequently in a distribution (Howell, 1992:31; Popham & Sirotnik, 1992:16; Bless & Kathuria, 1993:35). According to Popham & Sirotnik (1992:16), the mode is the index of central tendency least used. The mode has, according to Howell (1992:32), the advantage of not being influenced by extreme scores.

10.2.1.2 The median

The median is the midpoint in a set of ranked scores (Howell, 1992:31; Popham & Sirotnik, 1992:15). Howell (1992:32) states that the median shares it major advantage with the mode in that it is unaffected by extreme scores on either end of the distribution. Guy et al (1987:322) add that the median becomes an especially important measure when there are a few extreme scores in a distribution.

10.2.1.3 The mean

The most often used measure of tendency is the mean, which is actually the arithmetic average of a set of data (Howell, 1992:32; Popham & Sirotnik, 1992:14). According to Howell (1992:33), certain disadvantages are associated with the mean. It is influenced by extreme scores and its value may not actually exist in the data. In its favour it can be said that it is unbiased and efficient and can be manipulated algebraically (Howell, 1992:33). Bless and Kathuria (1993:46) add that since the mean makes use of every score in the distribution, it is the most accurate measure of central tendency.

10.2.2 Measures of variability

Guy et al (1987:322) state that while measures of central tendency are concerned with how the scores in a distribution are grouped together, measures of variability are concerned with how the scores in a distribution are spread apart. McMillan and Schumacher (1993:205) add that measures of variability indicate how much scores, on the average, differ from the mean.

10.2.2.1 Variance

Bless and Kathuria (1993:63) as well as Guy et al (1987:325), define variance as the mean of the squared deviations from the distribution's mean. In formula form the variance can be expressed as:

$$s^2 = \frac{\sum (X - \bar{X})^2}{N}$$

10.2.2.2 Standard deviation

McMillan and Schumacher (1993:206) state that the standard deviation is a numerical index that indicates the average variability of the scores; it thus gives information about the distance, on the average, of the scores from the mean. Bless and Kathuria (1993:64) define the standard deviation as the positive square

root of the variance. In formula form the standard deviation can be expressed as:

$$s = \sqrt{\frac{\sum(X-\bar{x})^2}{N}}$$

10.2.3 Frequency distribution

In a frequency distribution the data are listed or grouped in order to assess how often a particular score or result occurs (Bless & Kathuria, 1993:vi). Usually the frequency distribution of a set of data is presented graphically.

10.3 Inferential statistics

According to Bless and Kathuria (1993:77), inferential statistics refer to techniques for making statements on the basis of partial information. They add that the aim of inferential statistics is to provide information that is not directly included in the data. For this study, the determination of correlation coefficients and the performing of multiple regression analysis are important and they will thus be discussed.

10.3.1 Correlation coefficients

According to Popham and Sirotnik (1992:65), the term "correlation", in statistical analysis, always refers to a quantifiable relationship between two variables. McMillan and Schumacher (1993:215) state that the most common correlation technique is the Pearson product-moment coefficient. The product-moment correlation is used when both of the variables use continuous scales.

10.3.2 Multiple regression analysis

Hair, Anderson, Totham and Black (1992:7) describe multiple regression as the method of analysis that is appropriate when the research problem involves a single metric dependent variable presumed to be related to one or more metric independent variables. The objective of multiple regression analysis is to use the independent variable of which the values are known to predict the single dependent value the researcher wishes to know. The result is a variate, a linear combination of the independent variables that best predicts the dependent variable (Hair et al, 1992:25).

Popham and Serotnik (1992:104) argue that the technique of regression enables investigators to make predictions regarding a person's performance on one variable, given that person's performance on another variable. In multiple regression then, two or more predictor variables, related to the criterion variable, are incorporated into a more complex prediction scheme. The authors motivate their argument as follows:

> "...by adding a second variable related to the criterion variable, it is possible to reduce the standard error of estimate that would be present in a single predictor scheme. The better additional predictor usually is one that is at the same time related to the criterion variable and not too strongly related to the predictor variable already used." (Popham & Serotnik, 1992:105.)

McMillan and Schumacher (1993:273) state that the combined effect of the independent variables, in terms of the predictive power, to the dependent variable is presented by **R**, the coefficient of multiple correlation. The coefficient of multiple correlation can be thought of as simple correlation of all the independent variables together with the dependent variable. According to Hair et al (1992:64,) this coefficient reflects only the degree of association between two

variables. **R square** (R^2) is the correlation coefficient squared, also referred to as the coefficient of determination. This value indicates the percentage of total variation in the dependent variable explained by the independent variable (Hair et al, 1992:64).

In most instances of multiple regression the researcher has various possible independent variables to select for inclusion in the regression equation. In the process of identifying the best regression model for the situation, the researcher can follow one of two general approaches, namely, a sequential search process or a combinatorial method (Hair et al, 1992:56).

Two general types of sequential processes are backward stepwise regression and forward stepwise regression (Hair et al, 1992:57).

Backward stepwise regression involves computing a regression equation with all the predictor variables and then going back and deleting independent variables that do not contribute significantly.

Forward stepwise regression also allows the researcher to examine the contribution of each predictor variable to the regression model independently. Each variable is considered for inclusion in the model prior to developing the equation.

According to Hair et al (1992:58), the combinatorial approach is primarily a generalised search process across all possible combinations of independent variables until such time as the best-fitting set of variables is identified.

As it was extremely important to know the contribution of each independent predictor variable to the regression model, the forward stepwise process were selected for this study. By being able to identify each individual predictor's

contribution to the regression model, it should be possible to select the predictors with the highest validity to include in the final assessment battery, which is the ultimate goal of this study.

10.3.2.1 The standard error of estimate

The **standard error of estimate** functions as a measure of accuracy of predictions. Hair et al (1992:64) describe the standard error of estimation as "...the square root of the sum of the squared errors divided by the degrees of freedom. It represents an estimate of the standard deviation of the actual dependent values around the regression line; that is, it is a measure of variation around the regression line."

10.4 Levels of significance

McMillan and Schumacher (1992:341) state that since the basis of inferential statistics is the probability of estimation, the acception or rejection of the null hypothesis is also related to probability of chance. The **level of significance** indicates what the chance is that the researcher is wrong in rejecting the null hypothesis. In this study a level of significance of 5% was accepted. That implicates that assumptions made should be proven correct in 95 out of 100 instances. According to De la Rey (1976:172) researchers, ultimately aim at significance on the 1% and especially the 0.1% level, but all three of the levels mentioned are accepted in scientific research.

10.5 Steps taken in statistical analysis

The statistical analysis in this study involved the following steps:

- The mean and standard deviation for all predictive and criteria variables

were calculated. A frequency distribution for each variable was drawn. Attention was given to the median, as well as the skewness and kurtosis of each distribution.

- A multiple regression analysis on the two sets of data was performed.

- The results of the senior students were analysed, again using multiple regression, in the cross validation study.

- The level of significance of the results was calculated and discussed.

- A battery for the selection of engineering technician students at Technikon Pretoria was then suggested.

10.6 Conclusion

Quantitative research always implies statistical analysis. In this chapter the statistical techniques that played an important role in the data processing procedure followed in this study were discussed. Attention was specifically given to multiple regression analysis, as this technique is important in predictive studies. The results obtained in the analysis will be discussed in Chapter 12.

Chapter 11
Determining the predictor variables

11.1 Introduction

From the literature overview reported in Chapter 6, it is clear that a large number of possible predictors for academic performance exist. All of them cannot be tested in the practical process of selection and not all will contribute to the fairness and predictive validity of the selection battery. As the main aim of this study is to find the best combination of practically measurable predictors to include in a selection battery for Engineering Technology students at Technikon Pretoria, a specific instrument was used to identify some of these predictors. The results obtained by means of this instrument were combined with some of the variables found to be valid in similar studies. The determination of the predictor variables will be discussed in detail in this chapter.

11.2 The Comprehensive Structured Interviewing for Potential System

As described in Chapter 8, the actual performance of students in the programme was analysed in order to determine the competencies important for success in an Engineering Technology training course. As a starting point in this process the Comprehensive Structured Interviewing for Potential system (CSIP) was completed by 12 lecturers of the Faculty of Engineering, who are all directly involved in the training of Engineering Technology students.

The questionnaires asked of the respondents to select the eight competencies

from a list of 65 which are indispensable for the success of an Engineering Technology student. Subsequently, the respondents were asked to rate these eight competencies in order of significance. One of the essential areas in this process was to ensure that the competencies referred to requirements set for the incoming student and not for the required student level at the end of the training programme.

These responses were then computer processed and the results indicated to the researcher which predictive variables were thought to be important for the success of students by academic staff concerned with the training of engineering technician students. The competencies indicated by the respondents were:

- Creativity
- Reading Comprehension
- Mental Alertness
- Vocabulary
- Numerical Ability
- Composition of Wholes
- Spatial Reasoning
- Perception

Chapter 7 gives a detailed description of each of these competencies, as defined by the programme developers.

11.3 Academic criteria

At present the admission requirement for technikon education is a Senior Certificate. In the case of Engineering Technology courses at Technikon Pretoria the Senior Certificate should include Mathematics and Physical Science as

subjects. A prospective student should furthermore obtain at least a C symbol on the Standard Grade or an E symbol on the Higher Grade to be eligible for admission.

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 at a severe disadvantage if this is used as sole selection criteria. As the majority of the applicants for the Engineering Technology 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 academic performance will again be looked at.

11.4. Conclusion

The competencies indicated by academic staff involved with the training of Engineering Technology students at Technikon Pretoria were hypothesised to be indicative of a potentially successful student. After the identification of these predictor variables the assessment battery to be used in this study was compiled. This was then included in a comprehensive set of data regarding each applicant, together with the required school performance. Chapter 12 will describe the results obtained after a statistical analysis on the total set of data had been performed.