

THE UNDERSTANDING OF STATISTICAL TERMS BY
POSTGRADUATE STUDENTS OF EDUCATION

by

DONALD REX EMSLIE

Submitted in partial fulfilment of the requirements for the Degree of Master of Education (Educational Psychology) in the Department of Educational Psychology, University of Natal, Pietermaritzburg.

Pietermaritzburg: University of Natal

July 1992

ABSTRACT

Postgraduate students of education are expected to be able to critically evaluate articles written in research or experimental form. Such articles require the reader to have an understanding of statistical terms. The purpose of this study was to investigate the level of comprehension of statistical terminology in an intact sample of 155 postgraduate education students from an English medium university in Natal. From a review of literature it was predicted that the comprehension would be low and it was hypothesised that those students who had done a course in statistics would perform better than those students without such a course. An instrument, devised by McLaughlin, Hinojosa and Trlica (1973), consisting of 23 statistical terms was used to measure the subjects' understanding of statistical terms. The responses were scored on a 3 point scale by two raters. The inter-rater reliability was high ($r = 0,99$). Results of the predetermined groups indicated a low level of understanding of statistical terms with the best performing group having a mean of 8,13 (possible max. is 46 points). A one-way analysis of variance test of statistical significance showed that the hypothesis was supported by the results of two of the three groups sampled. These results suggest that the subjects sampled would be unable to critically evaluate much of the published research and that further training in statistics is needed if this is expected of them. It was suggested that various statistical courses (as well as research design and psychometric courses) for various needs should be offered at undergraduate and postgraduate level to facilitate the improvement of measurement literacy. Replication of this study at other universities and among serving educators would establish the extent of this low level of statistical comprehension. Further research is needed to investigate other aspects of the ability to critically evaluate published research in education.

2.4	RESEARCH STUDIES OF STATISTICAL COMPREHENSION	24
<u>CHAPTER 3</u>		36
3.0	METHODOLOGY	36
3.1	RATIONALE FOR THE INVESTIGATION	36
3.2	RESEARCH DESIGN	38
3.3	MEASURING INSTRUMENT	39
3.4	PROCEDURE	42
<u>CHAPTER 4</u>		44
4.0	ANALYSIS OF RESULTS	44
4.1	SCORING THE RESPONSES AND ORGANISING THE DATA	44
4.2	ANALYSIS OF THE DATA	45
	4.2.1 TEST OF THE HYPOTHESIS	45
	4.2.2 INVESTIGATION OF INTER-ITEM PERFORMANCE	51
<u>CHAPTER 5</u>		52
5.0	DISCUSSION	52
<u>CHAPTER 6</u>		57
6.0	CONCLUSIONS	57
6.1	IMPLICATIONS FOR THEORY AND PRACTICE	57
6.2	LIMITATIONS OF THIS STUDY	60

6.3	FUTURE RESEARCH	62
	REFERENCES	67
	APPENDICES	70

ACKNOWLEDGEMENTS

I wish to acknowledge the help of all those who have assisted me in undertaking and concluding this research study. I am particularly grateful to the following:

Robin Farman, my supervisor, for his valuable advice and suggestions in the completion of this study;

Dave Weekes for his help in the marking of the scripts;

my wife, Caro and sons Stefan and Brent, for their unconditional understanding, love and support;

Kathryn Buchanan for her typing of the manuscript and Peter Levey for his computer expertise; and

finally, a special acknowledgement to my friends and parents for their encouragement.

P R E F A C E

Unless specifically indicated to the contrary in the text, this thesis is my own original work.

SIGNED: 

D.R. EMSLIE

LIST OF TABLES AND FIGURES

TABLE		PAGE
1	Burkhardt, Casady and Forsyths' (1971): Table of Specifications.	29
2	Understanding of Statistical Terms.	41
3A	Comparison of Group Mean Scores.	47
3B	Analysis of Variance Summary Table.	47
4	Pairwise Comparisons between Groups using a one-way ANOVA test.	48
5	Comparison of Groups With and Without a Statistical Course.	49
6	Item Scores for Statistical Terms.	50
FIGURE		
1	The Logical Structure of the Quantitative Research Process.	11

CHAPTER 1

1.0 INTRODUCTION

Professionals and postgraduate students in the field of education are expected to be au-fait with recent and relevant research in their particular field. The assumption that postgraduate students are able to critically evaluate research raises the question : Are postgraduate students asked to do what they cannot do? A failure to evaluate research will lead to an unconditional acceptance of findings or an avoidance of research altogether, which defeats the objective of published research, which is to disseminate knowledge through reading. In quantitative research, the objects and behaviours being observed are measured, the data are collected and subjected to statistical analyses and then included in the published report. Thus to be able to read the published report effectively, a course in statistics and research design is a pre-requisite. The aim of the present dissertation is to determine the level of understanding of statistical terms by postgraduate students of education.

At the present time in South Africa, post graduate students of education (who are often older part-time students) are already in positions of influence (for example, inspectors, principals and heads of departments) or will be influential decision makers in South African educational terms. Thus, with the present reorganisation of education and the great challenges ahead, there is a need to avoid costly mistakes caused by wrong decisions. This highly qualified minority, through its influence of leadership, decision making and teaching, need to be aware of theory and practice in published educational research.

Post graduate students disseminate knowledge by sharing new ideas with colleagues, make decisions based on acquired information and ultimately test these new trends in the classroom. The success or failure of research then stimulates further research. However, many teachers and students alike view educational research as an expendable academic appendage and only a few are aware of the studies of competent investigators - especially those published in educational journals (Farman, 1980). This makes one sceptical about research ever influencing or informing practice.

Research writers, academics in educational faculties and members of professional boards assume that once students successfully complete a course in statistical techniques they will be able to read quantitative research. However, Lambert (1991) has noted that students in education and psychology in the United States of America, are in a state of crisis as far as their low standard of measurement literacy is concerned. In training institutions there has been a lack of supply of measurement specialists and a poor demand by students for such courses. Eysenck and Kline (1991), also voiced their concern about the low level of psychometrics and attributed this to a lack of psychometricians in training institutions, in Britain, which in turn affects the quality of lecturing. Davison, Damarin and Drasgow (1986) have also noted a drop in the number of courses in research design and statistics offered at American universities. A resulting effect of low standards means that postgraduate students feel disempowered in that they have limited access to research and find it difficult to evaluate the methodology and statistical analysis of research articles. A further effect would be difficulty experienced in interpreting standardised tests in the classroom. In 1980 - 1981, Lambert, R.F. reported the results of a survey (in the

opinions of teacher representatives) of teacher attitudes toward tests and testing. Most reported that teachers saw the tests in the following light "...negative, afraid of results, suspicious, a threat to job security" (p. 14). Ruddell's (1985) investigation of field educators (teachers, principals and district office personnel) and legislators' knowledge and attitudes towards testing in the California district, revealed similar findings in that teachers had difficulty interpreting tests and relied on other experts. Ruddell's study negates the assumption that teachers and educational professionals in California using tests and making decisions based on test data, are knowledgeable about the specific types of tests, test scores and measurement concepts.

It is the present writer's impression that the teachers' negative attitude towards testing is representative of most students' attitude when confronted with a compulsory statistical module as part of their degree course. Thus, in the reading of research and experimental articles, postgraduate students are most likely to view the research design and statistical manipulations as being tedious, anxiety-provoking and complex.

A literature search revealed that no research has been conducted in South Africa on the understanding of statistical terms. Although lecturers are aware of the difficulty students' experience with research design (Steinberg, 1983) and statistics (Heyns, 1983), the reason for this lack of research may be due to:

- i) a lack of interest in this field of research by masters students;
- ii) the consequences and pervasiveness of a lack of statistical comprehension not being evident;
- iii) a lack of experts in the measurement field in psychology and educational faculties,

and

- iv) the study of statistics being considered a minor part and hence the 'less' important part of degree courses.

Limited research has been conducted elsewhere in the English-speaking world. Research by Hammill and Blumberg (1967) and McLaughlin, Hinojosa and Trlica (1973) in the United States of America showed that samples of special education students had a low level of understanding of the statistical terminology used in most literature of special educational journals. The McLaughlin et al. (1973) study found that those special education students who had undertaken at least one course of statistics performed better than those that had not. Both groups of researchers suggest that there is a need for further training in statistics and research design.

Statistics and research design are presently taught in South Africa as part of some degree courses at different levels of sophistication and depth. Thus postgraduate students of education vary in their knowledge of statistics and research design, from those with no training, to those who have done a number of courses in research design and statistics. If the low level of measurement literacy (that is, psychometrics, research design and statistics) in the United States of America is an indication of students' competence, as Lambert (1991) suggests, then one questions whether postgraduate students in South Africa are capable of understanding statistical terminology in research articles.

This dissertation focuses on one aspect of measurement literacy, that is, the ability of postgraduate students of education to understand statistical terms found in research journals of education. It is also hypothesised that those students who have done a course

in statistics will have a better comprehension of statistical terms than those who have not undertaken a course. The population to be sampled will be education students at one English language university in South Africa. The objectives of this study will be:

- i) to replicate the study of McLaughlin et al. (1973) so as to establish the level of understanding of statistical terms. McLaughlin et al's. study is replicated as it is economical, the subjects are easily accessible and it provides a comparative base;
- ii) to confirm or refute the trend in the USA;
- iii) to consider the implications for the practice of reading research to diffuse ideas as students who have difficulty understanding statistics cannot criticise research paradigms, and
- iv) to inform future research in measurement literacy.

The level of comprehension of statistical terminology will be ascertained by presenting McLaughlin et al's. (1973), list of 23 items of statistical terms from special education journals to a sample of students. They will be required to define such terms and these will then be scored for accuracy.

To provide a basis for the present dissertation and to show how the research conducted will add to knowledge, review of relevant literature will be presented in Chapter 2. The methodology which will include the research design, measurement instrument and statistics used will be presented in Chapter 3. Chapter 4 will present the results and these will be discussed in Chapter 5. Finally, the implications for theory, practice and further research will be indicated in Chapter 6.

CHAPTER 2

2.0 LITERATURE REVIEW

The following presentation will be adopted in this chapter. Firstly, quantitative research and its theoretical underpinnings as applied to social research will be discussed. Secondly, the types of statistics and terminology used in educational research will be mentioned. Thirdly, the most recent research in measurement literacy and its implications for training will be discussed so as to reflect the crisis that exists. Fourthly, statistics and its importance in research design, textbooks, psychometrics and computers will highlight the importance of understanding statistics. Lastly, research relating directly to the comprehension of statistics will be critically evaluated.

The above mentioned review will show that the study of statistics is important and relevant for the reading of research, for conducting research, testing, interpretation of tests and decision making. At this present time in South Africa important decisions and challenges lie ahead and postgraduate students need to be aware of research and the understanding of statistical terms important to critically evaluate research. No previous research, on the understanding of statistical terms, has been conducted in South Africa. This field needs to be tested.

2.1 RESEARCH

Investigations in which the results are analysed by statistical procedures and include

design issues are referred to as quantitative research and occur in varying proportions in educational journals. It is necessary to investigate the theoretical underpinnings of quantitative research so as to provide a context within which the comprehension of statistics can be discussed.

2.1.1. QUANTITATIVE RESEARCH AND POSITIVISM

Quantitative research is underpinned by the natural science model as used in such subjects as physics. This means that research has the underlying assumptions, rules of logic and procedures borrowed from the model. The reason for using the natural science model is mainly two-fold. It has been very successful at facilitating understanding through its scientific investigation and there has been a need to secure government funding to carry out such research (Bryman, 1988). It seems that government funding is partial to the scientific model as it is objective in its approach and the concepts studied are measurable and statistically tested.

The notion of positivism underlies the scientific method that underlies quantitative research.

2.1.1.1 THE POSITIVISTIC POSITION

The term 'positivism' has a range of meanings which "...rarely agree on its essential components" (Bryman, 1988 p. 14). However, positivism is comprised of the following principles:

- i) The principle of methodological naturalism. This is a belief that the methods and

procedures of natural sciences are appropriate to social sciences. This implies that the objects of social sciences, although having qualities not normally held by the objects of natural sciences, are not an obstacle to the implementation of the scientific method. For example, such concepts as power or feelings are measured in social science research.

- ii) The doctrine of phenomenalism or empiricism. This entails the belief that only phenomena observable by the human senses can be collected and validly warranted as knowledge. This doctrine rules out the possibility of incorporating subjective experience or feelings as scientific knowledge, unless they are observable. By implication this means that observations are the final tests of theoretical disputes, whereas theoretical reasoning is subjugated to a relatively minor role. This aspect is further supported by operationalism which seeks to remove the ambiguity in the concepts that are typically embedded in scientific theories, by specifying the operations by which they are measured.

Philosophers of science have criticised the notion that observations are independent of theory and hence are neutral and uncontaminated by the scientists' theoretical or personal predilections. Kuhn (1970; cited in Bryman, 1988 p. 16) suggests a view that "observations are theory dependent". For example, this is seen in the history of science that as one model is eclipsed by another the scientist's image of the world also changes and observations are interpreted within a different context. An example is the changing model of the atom from one of a single sphere of solid matter (Dalton's model) to one consisting primarily of space, with various orbiting particles in mathematical wave functions about the nuclei (classical wave

model). Harré (1972; cited in Bryman, 1988) has also criticised positivism for its insistence on directly observable phenomena when, as a result, it has failed to give adequate recognition to the role in many scientific theories of hypothetical entities which may not be directly observable. These entities are used to facilitate the understanding of the causal mechanisms which underpin the phenomena being observed. Critics argue that this use of hypothetical entities does not help positivism to support the logic of the natural sciences.

- iii) The doctrine of inductivism. This occurs when "...scientific knowledge is arrived at through the accumulation of verified facts" (Bryman, 1988, p. 15). These facts are then developed into a theory which is a reflection of these accumulated findings or laws.
- iv) The doctrine of deductivism. This occurs when hypotheses are derived from scientific theories - usually in the form of postulated causal connections - between entities which are then submitted to empirical tests. If the hypothesis is rejected after thorough empirical testing then the theory must be revised. This implies that there is a separation between fact and theory.
- v) Positivism also takes a particular stance in relation to values. Firstly, it tries to eliminate the scientist's subjectivity which may alter his or her objectivity and so undermine the validity of knowledge especially in the field with moral or political implications.

The second aspect on values is that positivism draws a sharp distinction between scientific issues and normative ones. The reason is that a normative position can be investigated, but cannot verify or falsify the position itself.

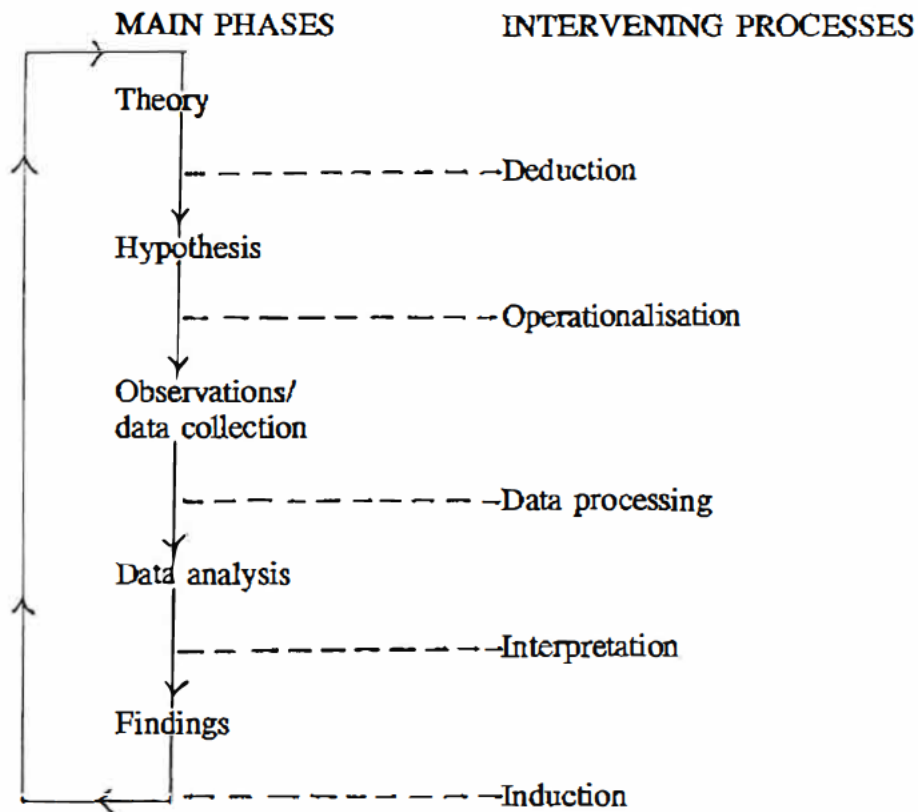
Bryman (1988, p. 16) points out that "...no single treatment of positivism which entails all of these principles and not all positivists (living or dead) would subscribe to all of them." Although positivism does not adequately describe the nature of natural sciences, there is no superior apologetics for quantitative research. Therefore quantitative research is conventionally believed to be positivistic in conception and orientation.

2.1.1.2 THE INFLUENCE OF POSITIVISM ON QUANTITATIVE RESEARCH

The impact of the scientific method and, in particular, positivism, on quantitative research has created a number of preoccupations which are prevalent in both reports of investigations and various writings on matters of method. These preoccupations are, namely, concepts and their measurement, causality, generalisation, replication and individualism. Refer to Bryman (1988) for further elaborations of these preoccupations.

Quantitative research then, can be seen as linked partly to positivism and partly to a diffuse and general commitment to the practices of the natural scientist.

Fig 1 THE LOGICAL STRUCTURE OF THE QUANTITATIVE RESEARCH PROCESS (Source Bryman 1988 p. 20):



The present dissertation is positivistic and reductionistic. It aims at fulfilling the logical procedures and processes of quantitative research. The main focus is the concept of the understanding of statistical terms. The hypothesis is directly related to the concept (deduction process) in that it is proposed that postgraduate students of education who have done a course in statistics have a better comprehension of statistics than those students who have not done such a course. The hypothesis is operationalised and is measured by using a measuring instrument consisting of 23 terms of statistics. The subjects complete

the questions, the data is then collected and processed by statistical techniques (descriptive and inferential) to analyse the data. Interpretation takes place and the findings are used to verify or negate the hypothesis. Implications for theory and practice are then discussed. The dissertation is positivistic and reductionistic in that the structure of quantitative research is followed and the statistical terms used can only be right or wrong.

2.2 STATISTICAL TECHNIQUES USED IN QUANTITATIVE RESEARCH

This dissertation focuses on the ability of postgraduate students of education to comprehend statistical terms. Statistical techniques in quantitative research are used to process and analyse collected data as well as to investigate relationships between variables so as to verify or reject the hypothesis.

Statistical methods can be divided into two groups, namely, descriptive and inferential statistics. Descriptive statistics include those of measures of central tendency, variability and relationships which describe and organise data which have been collected, for example, a table of ages. Inferential statistics, however, do not just describe data but also allow inference and conclusions to be made from the descriptive data as well as enabling researchers to test hypotheses.

2.2.1 DESCRIPTIVE AND INFERENTIAL STATISTICS

Descriptive data can be presented in various forms such as in tables and graphs and enables consumers of research to have easy access to the data. Descriptive statistics have had a poor image owing to their misuse by commerce. For example, statistical graphs have been used to support sales figures without displaying the overall picture, that is the

axes of the graph. Terminology used under descriptive statistics are the mean, mode, median (measure of central tendency); standard deviation, range, variance (measures of dispersion); cumulative frequency distribution, skewness, normal distribution (measure of distribution); percentile scores, quartile scores, interquartile range, standard scores and T scores (scores); and correlation, regression line, contingency tables and scattergrams (measures of relationships). In contrast, inferential statistics have been criticised for the number of ways producers of research have used them (Derrick, 1976). If statistics are not understood then the ability to detect faulty experimental logic and findings is impaired. By implication then, research findings will be unquestionably accepted or avoided altogether and will eventually result in a deterioration of research and lowered standards in measurement literacy. Terminology used in inferential statistics are the following; null hypothesis, statistical significance, level of significance, $p < 0,1$, statistical tests, t-test, F test, one and two-tailed test, ANOVA, ANCOVA, chi squared test and others. In publications inferential statistics have a reputation of enhancing conclusions more so than in the use of descriptive statistics. When solving the problem in principles of human behaviour, care has to be taken that statistics do not become the central focus of research, at the expense of the research design.

2.3 STATISTICS IN MEASUREMENT LITERACY

Research articles investigating the comprehension of statistics, measurement in tests and experimental logic all mention the need for further teaching of statistics, research design and measurement at graduate level because of what Lambert (1991) refers to as the 'measurement literacy crisis'.

The crisis has arisen because of the assumption that courses in individual testing, program evaluation or statistics satisfy the measurement standards, which they do not (Lambert, 1991; Hammill & Blumberg, 1967; McLaughlin et al., 1973). Lambert sees the crisis in economic terms, that is, one of supply and demand. The crisis is brought about by a lack of supply of psychometric academic staff and measurement courses, and the lack of demand by graduate students for such training.

Eysenck and Kline (1991) are alarmed at a similar trend in Britain as there is a lack of psychometricians at British universities. This lack of qualified staff has a ripple effect on psychometric training in theory and practice, research design and statistical techniques. This ultimately leads to lowered standards among psychological and educational professionals.

Lambert (1991) stresses the need to raise standards of measurement for education and psychology. She sets a minimal literacy level which involves the acquisition of four types of knowledge as set out in Appendix 1. The emphasis is mainly on test, test theory and testing which all teachers are involved in. Psychologists and psychometricians would be required to have a greater in-depth knowledge of (a) theory and norms behind intelligence, personality and guidance tests (b) research design, and (c) more advanced statistical techniques. Post graduate teachers, depending on their field of study, would be found on a continuum between teachers and psychologists. Although Lambert's use of the words 'measurement literacy' refers mainly to tests (standardised and psychometric) and testing, there is an overlap with statistics in that concepts such as validity, variability and measures of central tendency and others are used.

However, Lambert (1991) highlights a number of pragmatic problems in raising levels of competency, particularly associated with training. Although very little research is used to verify these problems, Lambert expresses her opinion based on her experience as head of an educational psychology faculty, and having served on a number of boards to establish measurement standards. These problems are the following:

- i) A poor supply of psychometricians on the faculty staff. When staff retire they are not replaced because of the low student demand for those particular courses.
- ii) In the past, the profile of courses offered in measurement have been low as a result of poor student numbers and a lack of emphasis and explanation of content by staff.
- iii) Financial concerns have resulted in university administrators being reluctant to supply staff after considering i) and ii) above. They suggest using staff from other departments such as the department of statistics.
- iv) Psychometricians' or experts' qualifications in measurement literacy are expected to be of a hybrid nature. They should be psychometrists who have a background in personality theory, developmental psychology, research design and statistics. The availability of such people is limited.
- v) The availability of psychometricians is poor because of competition in terms of remuneration offered by publishing and research companies.

Lambert (1991) sees the answer to these problems as a promotional one on the part of training institutions, professional boards, publishers of tests, psychologists and professional educators. Suggestions made to rectify these problems are the following:

- i) To promote measurement literacy the faculty and courses offered should have a high profile. This should be done by measurement specialists, heads of, psychology, education and educational psychology faculties. To recruit students measurement literacy should be offered as a primary field of study using incentives such as bursaries and in turn for qualified persons to spend at least one year teaching in the faculty. The importance of measurement instruction and how it can be widely applied to program evaluation, teacher preparation, school psychology and post graduate students of education who conduct research, should be stressed. This should increase student numbers which will result in the improvement of faculty positions.
- ii) Professional boards, professionals in practice and publishers of tests should set high standards for accreditation as professionals and users of tests. The curriculum content should also be improved.

Lambert suggests that classical and modern test theory and their applications to common psychological problems, he included in courses that presently only offer practical courses in psychometrics. Courses should be offered for various purposes namely:-

- a) For training in statistics so that articles are understood. McLaughlin et al. (1973) suggests that practical methods be used such as reading through research articles and critically analysing them with an instructor. The practical implementation of statistics in basic research to improve the comprehension of statistics is also suggested.
- b) For those postgraduate students involved in research, statistics should be

taught in conjunction with research design.

- c) For all test users a course in psychometrics with basic statistics should be offered. In a survey Lambert (1991), found that there is a demand for measurement literacy courses for teachers (in the USA) but that the number of courses offered is minimal.
 - d) For the psychologist standards are higher. Humphreys (1987, p. 413) comments "A one year graduate course following a brief inadequate undergraduate exposure to quantitative methodology is not a sufficient basis for a career in psychological research or teaching." He suggests a psychologist should have an undergraduate major or minor in mathematics including matrix algebra and computer science, and that in graduate school there should be at least two full years in methodology courses or the equivalent in independent study.
- iii) Financial institutions should be approached to finance chairs in measurement literacy so that further staff may be employed.

Thus the promotion of measurement literacy and the increased standards will create a need and increase the demand for measurement instruction which will in turn put pressure on the administration to secure more faculty members for measurement instruction, research and theory development (Lambert, 1991).

Statistical terminology or techniques are part of measurement literacy but are not confined only to research articles; they are found in the textbooks, the field of tests and testing, research design, and are available in computer software. These areas will be briefly

considered to reinforce the argument that postgraduate students need to understand statistics.

2.3.1 STATISTICS IN EDUCATION AND PSYCHOLOGY

It is not only writers and readers of research articles who require a knowledge of statistics, but, according to writers of textbooks in psychology and education, students should also understand statistical concepts. Boneau conducted a survey "...to determine which terms and concepts in psychology's basic subfields are judged to be of sufficient importance that they should be of general knowledge within the psychological community, especially students" (1990, p. 891). A questionnaire containing 200 to 250 terms and concepts from each of the 10 subfields were sent to authors of textbooks (n=249) in those subfields, to rate the terms and concepts. Of the top 100 concepts listed, 42 were from the subgroup 'Method and Statistics'. The total response rate of all the authors was 64 percent on the first return, of which statistics and methods author's response was 72 percent on the first return (100 percent on the second). Within the subgroup of 'Method and Statistics' the top 10 rated terms (rating max. of 5) and concepts were the following:-

Control group; correlation coefficient; dependent variable; experimental group; hypothesis testing; independent variable; normal distribution; sample; significance level; significant difference.

Boneau (1990, p. 893), comments that "...finding 10 or more psychologists who are unanimous about anything clearly represents a major achievement, but it happened in the statistics subfield!" The reason, proposed by Boneau for this finding, is that every

psychology major is subjected to a statistics course and one assumes that he/she should have, through experience mastered the terms.

This survey is an exploratory one from the perspective of textbook authors. The survey serves to highlight the status of statistics and the necessity to understand methods and statistics.

A perusal of Sprinthall and Sprinthall's (1981) textbook on educational psychology used in the training of postgraduate teachers in South Africa, gives an indication of the depth of knowledge and understanding of statistics required to understand research articles.

Sprinthall and Sprinthall (1981) devote two chapters to statistics. In the first chapter, reliability, validity, graphs and measures of central tendency and variability are dealt with on a superficial level. The approach is that of defining the statistic, explaining it, listing the different types of statistics used followed by an example of each type. The correlational technique is given more exposure than the other statistics because of its more frequent use in research. The text is written in a style that is 'reader friendly' for the uninitiated (teachers) in the subject of statistics.

The other chapter focuses on the reading and understanding of research. Four of the most common statistical tests used to evaluate chance factors are discussed: these being the 't' test, the F ratio or analysis of variance, the Pearson correlation coefficient and the chi squared test. These tests are discussed at a level for understanding but not for use in experimental design, in that only the function of the tests is explained with no examples, formulae or calculations given. The underlying assumptions on which these tests are

based, are dealt with at a superficial level. The two types of research design briefly mentioned are those of the associative and causative type. The authors suggest a follow up to the text with one semester course in statistics and research design. In South Africa, there seem to be few opportunities for the teacher to undertake the further study recommended by South African teacher societies.

Although Sprinthall and Sprinthall's textbook would allow teachers to read most journal articles, the teacher still needs practical handling of the statistics as in the textbook they are not seen in the journal context. The level of instruction would not allow the user to conduct his/her own research.

2.3.2 STATISTICS IN THE FIELD OF TESTING

In the educational field, teachers and educational professionals are constantly using tests, interpreting test data and making decisions based on the outcome of the tests. This activity requires a general sophistication in measurement concepts if the tests are to be interpreted accurately. Ruddell (1985) in one of the aspects of his study on the knowledge and attitudes toward testing by field educators and legislators, focused on the individuals interpretation of the scale scores and the standard error of measurement (SEM). Subjects were asked to compare scores reported from two academic years and interpret the differences. Findings indicated that the ability to interpret the scores was low in that only 11 percent of teachers and 17 percent of principals understood the concept of SEM which is necessary to interpret many standardised tests. Of the office personnel, 63 percent understood the concept while only 17 percent of legislators did so. It was noted that interpretation of test scores varied greatly, depending on the role of the

interpreter.

The generalisation of these findings are limited because of the following reasons:

- i) geographical location: Ruddell's (1985) survey was limited to seven school districts in the state of California in the United States of America (USA). Although there is a cross-section of school districts, the findings may be too general in that they do not apply to any specific school district.
- ii) statistical techniques: The survey uses descriptive statistics in the form of percentages which makes the study accessible to readers with limited knowledge of statistics. However, the reader is unaware if the findings are significant.
- iii) methodology: The sampling techniques are not mentioned nor does the educational level of subjects sampled indicate whether they had any training in statistics and measurement. For example, one subject - a legislator with an engineering background - who was able to correctly interpret the SEM was mentioned. Thus the reliability is affected because of the lack of information.

Ruddell's study is narrow in the aspect of asking the subjects to interpret only one type of data (or statistic), that is standard scores with the SEM's. Thus the lack of inferential statistics to analyze the collected data, the sample size and sampling techniques limit generalisation to a wider population.

However, one might hypothesise that this trend may be representative of the whole state and other school districts in the USA and therefore needs to be replicated.

Another aspect of the study was to investigate whether the subjects or their peers would have difficulty understanding certain tests scores. These scores included total test score,

raw score and percentile score for each subtest on a larger test. Most teachers (89%) indicated they did not understand, or had some difficulty understanding such test scores, and would need assistance in interpreting them. Overall findings suggested that most educators emphasized the need for interpretation from someone familiar with measurement concepts.

A definite need is expressed by Ruddell (1985) for in-service and pre-service training for educators which supports Lambert's (1991) call for improved measurement literacy standards.

In South Africa, it seems that very little use is made by the classroom teacher of standardised tests developed by the Human Science Research Council (HSRC). It is usually left to the remedial teacher, counsellor or educational psychologist to administer such tests. Most tests or examinations used by a school are developed internally and little cognizance is given to validity and reliability of such tests. However at primary school level, group intelligence tests are administered in Standards 2 and 5 by teachers who are accredited testers. The type of information and statistics needed to be known is set out in Appendix 2.

In my experience of accrediting teachers to use the group intelligence test, it was found that most teachers have a limited knowledge of the statistics used, do not know how to apply them, and their retention of statistics is short term. However, the emphasis in this paper is on the reading of research. It is also required of some teachers, educators and psychologists to use statistics in research design and analysis.

2.3.3 STATISTICS AND RESEARCH DESIGN

Another aspect of measurement literacy is experimental logic or research design. It is not the present focus of this dissertation and therefore will only be commented on.

There are various misuses of statistical techniques of which the correlational technique is one. The misuse of the correlation technique in published research will be referred to so as to highlight the importance of statistical knowledge and comprehension.

Often there is the erroneous labelling of correlational studies as experiments, as in Rhue and Lynn's report on 'Hypnosis and Related Phenomena' (1988: cited in Bringham, 1989). The article, according to Bringham, introduces the text in the right perspective but later presents correlational findings as if causal relations have been demonstrated. However, alternative explanations are not proposed. Bringham cites the error of labelling as being due to the independent variable not being manipulated. Dominowski (1989) draws attention to the fact that it is not the manipulations of an independent variable or random sampling that prove causality (although these methods help improve causality), but rather the elimination of equally plausible and contradictory interpretations.

A postgraduate student who has limited or no understanding of statistical techniques and research design, will not be able to critically evaluate such an article. Even students who have had introductory courses in statistics may miss erroneous labelling in research articles. In Hickson, Christie and Schumkler's (1990) article, in the South African Journal of Psychology, it was stipulated that a multivariate analysis of variance was carried out between the three groups, whereas it was a two-way analysis of variance. The

detection of such an error requires an in-depth knowledge of statistics.

Often, postgraduate students in the research design of their dissertation will include a statistical analysis knowing that statistics is a powerful tool, without considering whether the analysis will improve the understanding of the data. It may in fact even lower the quality of their research (Heyns, 1983 p. 25) if the statistics used are incorrect or inappropriate.

2.3.4 STATISTICS AND COMPUTERS

Computer hardware and software have enabled the researcher to compute a large amount of data in a short time. An over reliance, by the undiscerning researcher on the computer to do his/her statistical analysis can lead to problems. Glover and Ronning (1987, p. 100) make the point that "...some software programs continue to have errors and bugs of various sorts that unsuspecting users may not recognize." Computer programs are only as good as the persons who use them. Unfortunately the statistically illiterate researcher may misinterpret the data if he relies on the computer (Humphreys, 1987). A need for further training in statistics is needed.

2.4 RESEARCH STUDIES OF STATISTICAL COMPREHENSION

A search of literature revealed very few previous studies. The lack of statistical proficiency of all the postgraduate students and professionals researched in the reviewed literature is well documented (Aylward & Verhulst, 1991; Hammill & Blumberg, 1963 and McLaughlin, Hinojosa & Trlica, 1973). This lack of statistical proficiency has particular implications for all disciplines which are consumers of quantitative research.

Emerson and Colditz (1983; cited in Aylward & Verhulst, 1991) in a review of four volumes of the *New England Journal of Medicine* (from 1978 to 1979) made an analysis of the readability of statistics used in articles. Findings showed that no knowledge of statistical techniques would allow the reader access to 58 percent of the articles reviewed. Further knowledge of statistical techniques meant more access. For example, the familiarity with three techniques (that is, descriptive statistics, 't' tests and analysis of contingency tables) allowed statistical access to 73 percent of the articles. Aylward and Verhulst, (1991) reviewed the *Journal of Developmental and Behavioural Pediatrics* (1988 through 1990) and found that there were more correlational techniques used than in the Emerson and Colditz (1983) review. Familiarity with "correlations, chi squared, and 't' tests was found to allow the reader statistical access to 60 percent of articles; inclusion of ANOVA/ANCOVA and regression increased access to 88 percent (1991, 370). Friedman and Phillip's (1981; cited in Aylward & Verhulst, 1991 p. 373) survey of a sample of almost 700 paediatric residents, revealed that less than 20 percent of the sample understood correlation coefficients and just over 50 percent of the sample could interpret probability values. Not only does the understanding of statistics increase one's access to information but also the ability to evaluate that information.

Aylward and Verhulst (1991) found that the readability of published articles "is compounded further because of significant problems found in statistics ... from medical and related literature." (p. 370). These problems were sampling issues (for example, small sample size, inappropriate control groups, failure to randomise), statistics (violating assumptions underlying statistical techniques, errors in interpretation in tests of significance, analysis of repeated observations as if they were independent), and other

conceptual flaws (such as a tendency to not consider clinical importance and statistical significance simultaneously).

Aylward and Verhulst (1991), in their article proceed to discuss various statistical concepts and design issues used in medical journals, explaining underlying assumptions, pitfalls and interpretations using medical studies as examples so as to benefit investigators engaged in research as well as those reading research articles. The readers of medical research are provided with a mental "checklist" of questions to be used when reading reports. These are as follows and are of benefit for all disciplines:

- "i) are the statistical procedures described adequately?
- ii) are the statistical analyses appropriate?
- iii) is the statistical material presented satisfactorily?
- iv) are confidence limits provided? and
- v) are the conclusions drawn from the statistical analyses justified?"

(Squires, 1990: cited in Aylward & Verhulst, 1991 p. 375)

Thus one of the functions of Aylward and Verhulst's (1991) article, is for readers to use it as a reference.

As the present study focuses on postgraduate students of education, the researcher reviewed a number of educational research journals that students consult, to establish the percentage of research articles using statistics and statistical techniques. The following percentages were established for the 1990/1991 journal articles:

The Reading Teacher	2%
Journal of Learning Disabilities	65%

Academic Therapy	7%
SA Journal of Education	24%
Review of Educational Research	36%
Educational Research	60%
International Journal of Disability, Development and Education	70%
Journal of Educational Psychology	80%
Developmental Psychology	60%
Mental Retardation	50%

The implication of the above finding is that the understanding of statistics would allow the student and professional greater access to published research.

A literature search revealed few research studies specifically focusing on the comprehension of statistical terms. Of the four located, an article sampling paediatric residents has been mentioned (Aylward & Verhulst, 1991). A second article was specifically aimed at constructing a test of statistical comprehension, for use by graduate students in physical education (Burkhardt, Casady & Forsyth 1971, p. 235). The other two articles focus on the comprehension of statistical terms by teachers (Hammill & Blumberg, 1967) and students of special education (McLaughlin et al., 1973). The present dissertation focuses on the comprehension level of statistics.

The article of Burkhardt et al. (1971), on the development of a statistical comprehension test and norms, is very specific and structured. It is specific in that its target population

is graduate students of physical education and the statistical terms are based on statistical techniques and methods that frequently appear in the Research Quarterly journal from 1962 to 1966. These statistical concepts mentioned in the article are listed in Table 1 with the mean percentage of questions to be drawn from each statistic as suggested by the Research Council of the American Association for Health, Physical Education and Recreation.

The structure is set by the development of the scale, the administration of the test, the statistical calculations establishing the validity, reliability and parallel forms of the scale and finally the development of norms. As this study did not set out to establish the comprehension level of students, it is limited in its application to the present dissertation.

However, it must be realised that statistics are used in various disciplines so that all graduate students would profit by understanding statistical terms and thus certain aspects of Burkhardt et al's. (1971) study are discussed.

TABLE 1

BURKHARDT, CASADY AND FORSYTH's (1971) : TABLE OF SPECIFICATIONS

CONTENT AREA	MEAN PERCENTAGE
Descriptive Statistics	
Graphical procedures	4.8
Percentiles and percentile ranks	2.4
Other	<u>4.1</u>
	11.3
Concepts in hypothesis testing	7.8
Techniques of hypothesis testing for means, variances, and proportions	
t-test	12.6
Analysis of variance	11.7
Follow-up to analysis of variance	4.8
Analysis of co-variance	3.9
Other	<u>3.1</u>
	36.1
Correlation and regression	
Linear correlation and regression	9.2
Multiple correlation and regression	4.0
Generality and specificity	2.1
Other	<u>5.6</u>
	20.9
Factor analysis	4.3
Nonparametric tests	
Chi-square	3.9
Other	<u>3.6</u>
	7.5
Reliability	
Test-retest	4.2
Split-halves	2.3
Spearman-Brown prophecy formula	1.9
Other	<u>3.6</u>
	12.0

Source: Burkhardt, Casady and Forsyth (1971, p. 237)

Although the statistical test developed ($n = 1013$ from 51 colleges of the USA) proved to be reliable and valid, its norms apply only to physical education graduates in the USA and were found to be reliable only for master's and doctoral students with five or more semester hours in statistics and measurement. Thus the generality of the scale is limited and was not used in this study. A finding of interest in the development of the scale was that mean performance scores on the scale increased with the amount of formal instruction in statistics and/or measurement courses.

Hammill and Blumberg's (1967) study on the comprehension of statistical terms by teachers of special education, although dated, is still relevant as evidenced by Lambert's (1991) address to the American Psychological Association about the crisis in measurement literacy, and Eysenck and Kline's (1991) concern about the lack of psychometric experts at British universities.

Hammill and Blumberg found that while there was an "... increased statistical orientation in the literature, there is apparently no corresponding trend toward increased teacher familiarity with statistical symbols." (1967, p. 325) The investigation of the understanding of statistics sampled 76 special education teachers at Temple University in the Philadelphia district of the USA. Demographic information included degrees obtained and previous universities studied at. The comparison between those with at least one course in statistics ($n=29$) and those without a course in statistics ($n=47$) showed that those with statistics performed better in that they recognized more of the six symbols (SD; X; $p < 0,01$; r ; t ; F) than those without a course. Nearly fifty percent of the statistics group and eighty-five percent without statistics recognised two or less statistics (mainly

descriptive statistics which they define as from non-statistical articles). Only ten percent of the statistics group could recognise five or six symbols whereas none of the group without statistics could recognize more than four symbols. Of note was the fact that forty-five percent (n=13) of the teachers who had taken a statistics course failed to recognise more than two symbols; of the thirteen, twelve of these teachers did not have a masters degree. Hammill and Blumberg (1967, p.324) give a number of possible reasons for the poor performance of this predominantly undergraduate group, namely:-

- i) Most teachers had forgotten their statistical knowledge.
- ii) Inadequate teaching occurred.
- iii) There was adequate teaching but practical application may have been unrealised by the student and thereafter the subject matter was forgotten.
- iv) The course in statistics or measurement literacy is viewed with apprehension, in that the techniques are difficult to understand. This approach becomes a self-fulfilling prophecy, and most students pass and thereafter try to forget the somewhat unpleasant experience. Further training for students and in-service training for teachers is suggested.

Criticisms of the research carried out by Hammill and Blumberg (1967) is that it is over twenty-seven years old and only uses a measuring instrument of six statistics from special education journals. The scoring of the students' answers is referred to as being "lenient" without discussing how they were marked. The statistics used in portraying the results are percentages and therefore easily accessible to all readers, yet the comparison between the two groups is not mentioned as significant as no hypothesis was tested. The study is descriptive in nature and is not generalised in its application to the present investigation

and is possibly unreliable owing to the passage of time. Nonetheless, Lambert (1991) and Eysenck and Kline (1991) have voiced their concern that low measurement standards are still in evidence at present.

McLaughlin et al.'s. (1973) study on the comprehension of statistics by special education students showed similar findings to Hammill and Blumberg's (1967), and Aylward and Verhulst's (1989) studies. Postgraduate students have a low comprehension of statistical terms. McLaughlin et al. in their study sampled special education students which included certification, undergraduate and postgraduate students. It was also found that students who had at least one course in statistics did better than those who had no course in statistics. Although causality cannot be implied, one can hypothesise that a course in statistics improves one's comprehension of statistics. The assumption that advancement in degree level improved one's performance on the measuring instrument (due to increased research article reading) was not supported. The reason for this is that the group of graduate students with no previous statistical training performed at the same level as doctoral students with no statistical training. The investigation of inter-item performances on the list of terms showed that the subjects were more familiar with descriptive than inferential statistics, thus limiting access to research articles using advanced statistical techniques.

The study by McLaughlin et al.(1973) is to be replicated in the present dissertation. However, there are a number of issues which affect the replicability. They are the following:

- i) though the sample size (n=122) is adequate no referral is made to the method

used in sampling the subjects. If a random sampling procedure was used this would increase the generality of the study.

- ii) the study was conducted approximately 19 years ago and this is bound to affect the reliability and generality of the study. The list of statistical terms used, which were originally from five special educational journals needs to be compared with terms used in the present educational journals.
- iii) the demographic data elicited was minimal. The students supplied their year of study and their major subjects as a means of identification. It appears that the major subjects are used to divide the students into two groups, that is, those with and without statistics.
- iv) the problems experienced by the researchers with the administration of the test were not mentioned. In the scoring of the answer sheets, no mention is made of the number of raters used for the inter-rater reliability.
- v) the list of statistical terms is substantiated by referring to its development, yet no statistical evidence is given for its validity and reliability. The development of the final list (short form) is not discussed.

Although one is aware of the constraints of a journal article, there is a lack of substantial evidence or reference to back-up statements. For example, "...special education students represented a similar proportion of the total population of students within each category." (McLaughlin et al. 1973, p. 410). The number or the proportion are not mentioned. The subjects sampled are from one institution. However they do not elaborate how this limits the study.

Also a level of sophistication in the area of statistics is expected of the readers of McLaughlin's article. For example, the use of a post hoc test (Dunn's multiple comparison test) to give direction is mentioned yet there is no elaboration on the purpose for this test. Inferential statistics are used which enables one to generalise at least to that particular university.

Although the factors above do affect the generalisation, reliability and replicability of the study there is a contribution that the McLaughlin et al. study makes to the topic of understanding statistical terms. Their study does highlight a relevant and pertinent problem that has ramifications for the teachers of statistics, the writers and readers of articles. The low level of understanding among special education students attending Austin Texas University implies (and it is stated) that further training in statistics is necessary, so that the purpose of disseminating research ideas and findings through publishing articles is achieved.

The present dissertation uses the measuring instrument derived by McLaughlin et al. (1973) at one educational institution. It samples only postgraduate students from the faculty of education. The subjects sampled are studying for the following diploma and degrees:

Higher Diploma in Education (HDE)

Bachelor of Education (B.Ed)

Bachelor of Education (Educational Psychology) B.Ed (Ed Psych.)

Masters of Education (Educational Psychology) M.Ed (Ed Psych.)

It can be deduced from the literature review that the standards in measurement literacy,

research design and in particular, statistics is low. The purpose of the study is to evaluate the level of comprehension of statistical terms and to compare the performance of students who have had a course in statistics with those who have had no course in statistics.

McLaughlin's et al. (1973) article is the basis for the hypothesis that postgraduate students, who have had training in statistics, should have a better comprehension of statistical terms than those who have not had a course in statistics.

The following chapter sets out the methodology used in the present dissertation.

CHAPTER 3

3.0 METHODOLOGY

This chapter sets out the four steps to be undertaken in developing the method of this research. Firstly, the rationale for the investigation will be discussed. Secondly, the research design will be discussed. This will be followed by the sampling and instrumentation used in the research. Lastly, the procedure will be documented.

3.1 RATIONALE FOR THE INVESTIGATION

Statistical terms and techniques are used in quantitative research to confirm or refute association, or causality between variables. It is thus necessary for all students, especially postgraduate students, to be able to understand statistical terms so that research dissertations and articles may be evaluated. Once evaluated, rigorous research may then be disseminated and used in teaching, counselling, decision making and in the performance of one's profession.

In chapter two, previous research has been documented and discussed to provide a basis for the present investigation. The necessity to understand statistical terminology to gain access to, and evaluate research, was mentioned by Aylward and Verhulst (1991). The low level of understanding of statistics by postgraduate students was documented in research by Friedman and Phillips (1981: cited in Aylward & Verhulst, 1991), Hammill and Blumberg (1967) and McLaughlin, Hinojosa and Trlica, (1973).

The first issue this study investigates is whether students at postgraduate level, in the

faculty of education, of one university in the Province of Natal, are able to comprehend statistical terms so that they may critically evaluate research articles.

Expectations are that those students who have had a course in statistics should have a better comprehension of statistical terms. Burkhardt, Casady and Forsyth (1971, p. 239) in devising a test of statistical comprehension for graduate physical education students, demonstrated that a "relationship existed between the scores achieved by the examinees on the test and their grade point averages for all graduate courses in statistics and/or measurement. Correlation coefficients ranged from 0,25 to 0,42 and all were significant at the 0,1 level." For masters and doctoral candidates "the mean scores increased with increasing amounts of formal instruction in statistics and/or measurement courses." (p. 240). This implies that attending the course per se does not enhance one's comprehension of statistical terms, but rather that the understanding of the course contents improves test performance.

McLaughlin et al. (1973) found that taking a statistics course was helpful in understanding statistical terms when compared with special education students without such a course. However, relative to the testing situation they still performed poorly. For example, masters students with statistics had a mean of 26,7 which, when compared with the maximum possible 46 points, indicated that just over half the terms were understood which was poor. This leads us to the second issue of the present study; whether students identified above would differ in comprehending statistical terms on the basis of whether or not they had previously studied at least one course in statistics.

Hypothesis: It was hypothesised that postgraduate students of education who had previously taken a course in statistics would do better than those without a course in statistics.

3.2 RESEARCH DESIGN

An attempt was made at a direct replication of McLaughlin et al's. study (1973). The procedures and instrumentation used in the present study is the same with similar analysis but with different samples.

SAMPLE

It was decided that postgraduate students, on various courses, in the Faculty of Education at the University of Natal, Pietermaritzburg would constitute an accessible and relevant intact group for the replication of McLaughlin et al's. (1973) study.

These postgraduate students are taking diplomas or higher degrees and are involved in the reading of research, with some students conducting their own research, and others testing and making decisions based on findings of research and/or tests.

It would be of interest to have investigated some other groups such as teachers, principals in schools and psychologists in practice; however these groups are not as easily accessible because of the scale and the time constraints of the study.

The groups used were 1991/92 H.D.E., B.Ed., B.Ed. (Ed. Psych.) and M.Ed. (Ed. Psych.) students in the greater Pietermaritzburg area. Intact groups were asked for their

voluntary participation in the study and no attempt was made at random sampling. Given this approach no specific claims can be made that these groups constitute representative samples of any particular population.

The B.Ed. students were sampled in 1991 and 1992, whereas those specialising in the psychology of education, together with HDE students, were sampled in 1991.

Data providing information regarding the number of students sampled in each group, and those with or without a course in statistics, is presented in table 3A and 4 (Chapter 4) respectively. Of note is the fact that students were divided into a group with and a group without statistics based on their undergraduate majors. Those majors, including a module on statistics in their courses, were psychology and geography. Subjects such as sociology included a course in research design in their third year, primarily to help these students produce a research paper. This course included statistical techniques but on a brief and minor scale. It was decided to place sociology students into the group without statistics.

The 155 students who chose to participate did so on a voluntary basis. A decision was taken not to reject any answer scripts, because students often left out many items which they could not answer and thus were given a score of zero for that particular item.

3.3 MEASURING INSTRUMENT

The survey technique was employed using a measuring instrument devised by McLaughlin et al. (1973) to collect the data.

SCALE TO ASSESS STATISTICAL TERMS

The McLaughlin et al. (1973) study focused on the ability of special education students to read and understand statistical terms used in research articles. The present dissertation attempts to replicate their study using the same measuring instrument.

In the development of the scale, a list of 23 statistical terms most frequently used by the researchers in the field of special education was delineated.

Initially a multiple choice test was designed using these terms. However, it was decided after pilot runs of the instruments, to change the design to the present short form of the test. The multiple choice design was not similar to the real experience of reading journal articles, that is, in reading an article a choice of alternative definitions of statistical terms are not given. McLaughlin et al. (1973) recognise that the short form finally used is still out of context but it is more appropriate than the multiple choice design. The frequency of a term found in the journals determined the inclusion of the term in the final list (Table 2). Finally the terms were randomised and presented to the subjects.

McLaughlin et al. (1973) do not include information about the validity and reliability of the instrument. Presumably these matters were considered and addressed in their study but are not reported in the journal article. Thus the accuracy of the scale needs to be confirmed.

TABLE 2 : UNDERSTANDING OF STATISTICAL TERMS

1. Median
2. Standard deviation
3. Range
4. Level of significance
5. $p < .01$; $p < .05$
6. Mode
7. Hypothesis
8. Mean
9. Analysis of Variance
10. N
11. F
12. Correlation
13. Covariance
14. df
15. r
16. Regression
17. Validity
18. r_{rho}
19. Reliability
20. Factor analysis
21. Interaction
22. Chi square test
23. Percentile

Source: McLaughlin, Hinojosa and Trlica (1973, p.409)

Advantages of using this instrument were that it i) allows for a single collection of data; ii) takes a period of about thirty minutes to administer and iii) is straight forward to score. These factors make the instrument economical for the present application.

This measuring instrument could be used to measure the statistical comprehension of subjects from any culture as statistical terms are universal. For this reason the instrument did not require adaption before it could be applied in the present context.

The definitions of all the terms in McLaughlin et al's. (1973) study were extracted from English and English (1965). In the present dissertation the definitions were extracted from English and English (1959) as the 1965 edition was unavailable. These definitions in English and English (1959) (see Appendix 3) were compared with Yaremko, Harari, Harrison and Lynn (1986) (see Appendix 4) so as to consider any changes in meaning or emphasis of statistical terms used. In comparison it was noted that core meanings of the statistical terms have not changed although some definitions do differ in detail and clarity. In the scoring of the instrument, using the definition of terms, various raters may be used allowing for inter-rater reliability to help decrease the researchers subjective bias in the scoring of the responses.

3.4 PROCEDURE

Permission to sample the postgraduate students was granted by the Education Faculty of the University. The measuring instrument was administered by a lecturer in the educational psychology department during 1991 and 1992, in the same venue, for a 45 minute period. The students participated in groups for which they were registered, for

example, B.Ed. degree group. The H.D.E. group was represented by all races in approximately equal proportions. The B.Ed. group was predominantly Zulu with a small number of white and Indian students. The smaller groups B.Ed. (Ed. Psych.) and M.Ed. (Ed. Psych.) consisted of an equal proportion of races.

Instructions were given in english and none of the groups was warned in advance of the testing situation. At the beginning of the testing session, a brief introduction to the study was given and they were asked to define the 23 statistics listed in the instrument. The students' co-operation was sought, anonymity and group performance was stressed, and the option to withdraw made the test voluntary. All students made an attempt, although some showed resistance as evidenced by their restlessness. At the end of the session the scripts were collected and debriefing took place, explaining the purpose of McLaughlin et al's. (1973) article and the significance of the study.

The scoring of the data and analysis of the results will be discussed in chapter 4.

CHAPTER 4

4.0 ANALYSIS OF RESULTS

This fourth chapter will be divided into two sections, the first being the approach used in organising the data, the second being the statistical analysis of the data.

4.1 SCORING THE RESPONSES AND ORGANISING THE DATA

For each group [H.D.E.; B.Ed.; B.Ed. (Ed. Psych.); M.Ed. (Ed. Psych.)] the mean was calculated. This was done in the following manner:-

Each of the 23 items was scored according to a 3 point scale (2 = answer fully meets the criterion; 1 = answer approaches criterion and 0 = answer totally incorrect). The definition given for each item by the subject was compared to the definition given by English and English (1959) in the Appendix 3. For each individual the 23 scores were summed. The total individual scores were then summed for each group and then averaged. Hence a group mean score for each group was obtained (see Table 3A). This procedure was used because the appropriate unit of analysis is the group, not the individual. This allowed for comparisons between groups of different academic levels.

The rating of the responses was carried out by two raters working independently, using the definitions of English and English (1959) as set out in Appendix 3. The inter rater reliability coefficient, using the product moment correlation coefficient

statistic, was 0,99 which is an extremely high correlation. However, differences between raters were evidenced on items scored 1, that is, when the rater had to use his discretion to decide whether the subjects definition approached the criterion.

As discussed in chapter 3, the investigation of the understanding of statistical terms by postgraduate students of education and the hypothesis that is tested in this study were based upon the review of the literature.

The hypothesis tested was as follows:

H: The understanding of statistical terms by postgraduate students of education who have had a course in statistics, would be better than those who have not had a course in statistics.

As previously stated the means for each group were established. The number of groups was doubled (for those groups who were not exclusively psychology graduates) so as to compare those groups who had a course in statistics, and those who had not (Table 5).

4.2 ANALYSIS OF THE DATA

4.2.1 TEST OF THE HYPOTHESIS

In this section, the statistical analyses for the hypothesis is presented. The hypothesis was tested, by the one-way analysis of variance test, for statistical significance at the 0,01 level of confidence.

In analysing the hypothesis, the format used was:

- i) the hypothesis was stated,
- ii) the statistical data were reviewed, and
- iii) the decision to accept or reject the hypothesis was made.

The results will be discussed in Chapter 5.

The results of the present investigation, as illustrated in Table 3A, confirm the expectations that postgraduate students of education would have a low level of understanding of statistical terms. The highest mean attained is 11,16 points by the B.Ed. (Ed. Psych.) group; this is less than a quarter of the possible maximum of 46 points. The mean's range from 3,056 to 11,16 points with the 1991 B.Ed. students achieving the lowest results.

The means and standard deviations measured central tendency of the data and dispersion, respectively. The differences between the groups sampled are indicated by the means, and an ANOVA indicated that there was a significant difference between the groups as shown in Table 3A. The F ratio 7,874 was significant beyond the 0,01 level of significance.

TABLE 3A : COMPARISON OF GROUP MEAN SCORES

GROUP	SAMPLE SIZE	MEAN ¹	SD	F RATIO
1992 H.D.E.	61	4,557	5,005	7,874
1991 B.Ed.	36	3,056	2,672	
1992 B.Ed.	49	3,143	2,614	
1992 B.Ed. (Ed. Psych.)	6	11,166	3,545	
1991 M.Ed. (Ed. Psych.)	3	8,333	0,333	
Group	155	4,090	4,182	

¹From a possible 46 points.*Significant at $\alpha=0,01$ level.**TABLE 3B : ANALYSIS OF VARIANCE SUMMARY TABLE**

SOURCE	SUM OF SQUARES	df	MEAN SQUARES	F
Between Groups	450,29	4	112,572	
Error	2144,45	150	14,296	
Total	2594,74	154		7,874*

* Significant at $\alpha = 0,01$

A pairwise comparison between groups, using an ANOVA as the test of statistical significance, gives an indication of direction on performance on the measuring instrument.

The results of table 4 indicate that students of psychology in education [B.Ed. (Ed. Psych.) and M.Ed. (Ed. Psych.)] have a better understanding of statistical terms than the other postgraduate groups.

For example, B.Ed. (Ed. Psych.)

>92 H.D.E. (F = 9,316)

>92 B.Ed. (F = 46,669)

>91 B.Ed. (F = 14,186)

TABLE 4 : PAIRWISE COMPARISON BETWEEN GROUPS USING A ONE-WAY ANOVA TEST.

GROUP	91 B.Ed.	92 B.Ed.	91 B.Ed. (Ed. Psych.)	91 M.Ed. (Ed. Psych.)
92 H.D.E.	2,619	3,042	9,316*	1,577
92 B.Ed.	0,023		46,669*	11,586*
91 B.Ed.			14,186*	11,391*
91 B.Ed. (Ed. Psych.)				1,770

* Significant F values at the $\alpha = 0,01$ level

Hypothesis: The understanding of statistical terms by postgraduate students of education who have had a course in statistics, would be better than those who have not had a course in statistics.

The hypothesis was supported by two of the groups (H.D.E. 1992; B.Ed. 1992) as presented in Table 5. The two groups consisting of postgraduate psychology students of education; B.Ed. (Ed. Psych.), M.Ed. (Ed. Psych.) have all done a course in statistics and therefore a comparison was not possible. Table 5 gives a resulting significant F value of 23,99 for 1 and 59 degrees of freedom (for the H.D.E. group) and 13,16 for 1 and 47 degrees of freedom (for the 1992 B.Ed. group) beyond the 0,01 level of significance. The F value 0,305 for the B.Ed. 1991 group was not significant at a 0,05 level of significance ($F = 0,305$) when those with a course in statistics were compared to those without.

A decision was made to reject the null hypothesis and accept the hypothesis for the 1992 H.D.E. and B.Ed. group. The null hypothesis was accepted for the 1991 B.Ed. group.

TABLE 5

COMPARISON OF STUDENTS WITH AND WITHOUT A STATISTICAL COURSE

GROUP	Without Statistics			With Statistics			Signif. F values
	Mean	SD	N	Mean	SD	N	
92 H.D.E.	2,333	2,704	36	7,760	5,819	25	23,99
91 B.Ed.	3,166	2,842	30	2,500	1,643	6	0,305NS
92 B.Ed.	2,138	1,865	29	4,600	2,891	20	13,161
91 B.Ed. (Ed. Psych)				11,166	3,545	6	
91 M.Ed. (Ed. Psych)				8,333	0,333	3	
			95			60	

NS - not significant.

TABLE 6
ITEM SCORES FOR STATISTICAL TERMS (N=155)

TERM	SCORE 0		1		2	
	N	Percent	N	Percent	N	Percent
Median	72	46,5	62	40,0	21	13,5
Standard Dev.	111	71,6	33	21,3	11	7,1
Range	100	64,5	33	21,3	22	14,2
level of Sign	141	91,0	11	7,1	3	1,9
p<.01:p<.05	133	85,8	17	11,0	5	3,2
Mode	133	85,8	8	5,2	14	9,0
Hypothesis	133	85,8	21	13,6	1	0,6
Mean	129	83,3	5	3,2	21	13,5
Analysis of Variance	152	98,1	2	1,3	1	0,6
N	125	80,7	29	18,7	1	0,6
F	154	99,4	1	0,6	0	0
Correlation	127	82,0	27	17,4	1	0,6
Covariance	151	97,4	4	2,6	0	0
df	124	80,0	4	2,6	27	17,4
r	146	94,2	9	5,8	0	0
Regression	155	100,0	0	0	0	0
Validity	136	87,7	2	1,3	17	11,0
r _{rho}	151	97,4	4	2,6	0	0
Reliability	128	82,6	20	12,9	7	4,5
Factor Analysis	152	98,1	2	1,3	1	0,6
Interaction	135	87,1	20	12,9	0	0,6
Chi square test	153	98,7	2	1,3	0	0
Percentile	149	96,2	3	1,9	3	1,9

4.2.2 INVESTIGATION OF INTER-ITEM PERFORMANCE (see McLaughlin et al. 1973, p. 410)

An overall view of responses to the statistical terms included in the 23 item measuring instrument is provided in Table 6. This was investigated in an effort to determine if a level of sophistication could be defined. The percentages are based on a population sample of 155 students. The low percentage response presumably indicates the low comprehension of statistics. The highest percentage of an item fulfilling the criterion so that a score of 2 is given, is 17,4 percent for degrees of freedom (df) and the highest approaching the criterion (score = 1) is the median (40%). In general the descriptive statistics (median, mean, mode, range and others) scored higher than the inferential statistics (analysis of variance, chi square test and others). Most statistical terms were not well understood, those terms and symbols which are the least understood (>94 % scoring 0) are the following:-

Analysis of variance; F; Covariance; r; Regression (worst); r_{tho} ; Factor analysis; Chi square test.

Of note was the fact that the definitions given by English and English (1959) were detailed and the students' definitions were very broad.

The overall poor performance on the measuring instrument is indicated by means (table 3A), the percentages of students who failed to score (table 6), and the better performance by students who have had a course in statistics is shown by the support for the hypothesis. These results need to be interpreted in the context of previous research in this field. This will be discussed in Chapter 5.

CHAPTER 5

5.0 DISCUSSION

This chapter will consider the results in the context of previous research conducted in the comprehension of statistical terms.

The aim of this study was to investigate the comprehension of statistical terms by postgraduate students of education and relate it to their ability to read research articles and utilise psychometric tests.

One hypothesis was to be tested, using a statistical analysis of variance. The population sampled was 155 postgraduate students of education in the greater Pietermaritzburg area. The students' comprehension of statistics was sampled by using a measuring instrument requiring definitions of twenty-three statistical terms.

It was expected that postgraduate students of education have a low level of understanding of statistical terms. This view was supported by the mean score of each of the four groups sampled. The ANOVA showed that these results were significant in that there are differences between the means of the groups. Further pairwise comparisons between groups (using the F ratio test) indicated that postgraduate students of psychology in education [B.Ed. (Ed. Psych.), M.Ed. (Ed. Psych.)] perform far better than the other groups. A possible reason for this is that all the students of these two groups have done

one or more courses in statistics.

This result supports the previous findings by Hammill and Blumberg (1967), McLaughlin et al. (1973) and Phillips and Friedman (cited in Aylward and Verhulst, 1991) where teachers, special education students and paediatric residents showed a low level of understanding of the statistical terminology. The present results are based on the same measuring instrument as used by McLaughlin et al. (1973) whereas Hammill and Blumberg (1967) used only six statistics (standard deviation, mean, probability less than 1%, correlation, t test and F test analysis of variance) and Phillips and Friedman (1981: cited in Aylward & Verhulst, 1991) used statistics from medical journals. The very low level of performance is possibly due to statistical illiteracy, a lack of use of statistics, the passage of time since a statistics course was undertaken (this would not apply to H.D.E.'s) and the false assumption that students of education actually read the method and results section of research articles. If postgraduate students of education did read and understand research, one would expect, as the level of study increased from H.D.E. to B.Ed., an improvement in the understanding of statistical terms. However, the results of table 5, of students without statistics, show that the level of understanding is low and does not increase with degree level, for example H.D.E. (1992) has a mean of 2,333 and B.Ed. 1992 has a mean of 2,138.

The fact that the acquisition of knowledge of statistical terms did not improve beyond the H.D.E. level is supported by the McLaughlin et al. study (1973). The generalisability, as in McLaughlin et al's. study, is limited as the data was collected at one training institution.

Secondly, it was hypothesised that the understanding of statistical terms by those who have studied statistics will be better than those who have not. The results of the mean values in table 5 indicate that taking a statistics course is helpful. Within each educational level, postgraduate students who had taken at least one course in statistics, performed better than those who had not, for example, the 1992 H.D.E. and B.Ed. groups. Of note is the fact that the results of the 1991 B.Ed. group (with and without statistics) were non-significant.

The highest scoring group with statistics, B.Ed. (Ed. Psych.) performed poorly (mean = 11,16) relative to the testing situation where a maximum of 46 points was possible. This indicated that they understood less than a quarter of the terms. Of note is the fact that, of the students who had a course in statistics, the B.Ed. groups performed the worst (means of 2,5 and 4,6). A possible reason is that B.Ed. students have been working in the field (teaching) and there has been a fall off effect with the passage of time. The H.D.E., B.Ed. (Ed. Psych.) and M.Ed. (Ed. Psych.) have recently been on a course of statistics assuming that the M.Ed. (Ed. Psych.) students have recently completed the B.Ed. (Ed. Psych.) degree.

These findings imply that with or without statistics, the possibility of understanding the results of data analysis in research articles is minimal. This supports McLaughlin et al's findings that students cannot critically evaluate research articles and there is a need for further in-service training.

In the comparison of 23 items of the measurement instrument in table 6, it was revealed

that descriptive statistics were better understood than inferential statistics. For example scores on the median, standard deviation, range and mean were greater than scores of the co-variance, factor analysis and chi square test items. Descriptive statistics are easier to understand, and are used more often in textbooks in other subjects such as physics, geography and sociology. The findings in table 6 indicate that access of most students to quantitative research will be limited to articles with descriptive techniques.

What is cause for concern is the percentage of 155 students who failed to score on individual items. Except for the terms median (46% failed to score) and range (64,5% failed to score), all other terms had a 70 or above percent of students who failed to score. A surprising result was the fact that although 60 students had a course in statistics only 9 of the 155 students could recognise the symbol r for correlation coefficient while 28 students ($n = 155$) could define the term correlation (they scored 1 or 2). The term 'correlation coefficient' was one of the top 10 rated terms by authors of textbooks on methods and statistics in Boneau's study (1990) in the USA, yet 94,2 percent of students in the present study, failed to recognise its symbol (r). Another term rated in the top 10 terms by Boneau's (1990) study was 'significance level' and in the present study only 14 ($n = 155$) students could define the term whereas 91 percent of students failed to define it.

Sprinthall and Sprinthall's (1981) textbook on educational psychology gives a great deal of exposure to correlation and measures of central tendency, which is an introductory text to statistics so that students are able to read educational journals. However, this study shows that most (80%) postgraduate students are unable to comprehend statistical terms.

Hammill and Blumberg (1967) included the symbols F and r in their measuring instrument. In their study 14 percent of students tested recognised the symbol r and 3 percent recognised the symbol F , whereas in the present study 5,8 percent recognised r and 0,6 percent recognized F . Clearly the results of the present study appear to show a lower level of understanding of statistical terms.

In comparing the definitions of the postgraduate students to the model definitions (English & English, 1959), it was found that the textbook often went into detail whereas students gave very general definitions. However, definitions of symbols by English and English (1959) were brief, for example df was defined as "symbol for degree(s) of freedom" without explaining the meaning of degrees of freedom. It is speculated that with the passage of time, most students who had done a course in statistics (especially B.Ed.) will have forgotten the precise definitions of statistical terms.

In general, these findings substantiate the findings of Hammill and Blumberg (1967) and McLaughlin et al's. (1973) study that special education teachers and students do not demonstrate a high degree of understanding of statistical terminology. Therefore, the assumption that potential readers of published educational articles understand and critically evaluate all the material is false. These findings support Lambert's (1991) notion that there is a crisis in measurement literacy in that the standards are low. This then in turn has implications for training institutions as voiced by Eysenck and Klein (1991) in Britain and Lambert in the USA. Eysenck and Klein (1991) are concerned about the long term effects as indicated by poor training of students because of a lack of a supply of psychometricians.

CHAPTER 6

6.0 CONCLUSIONS

This chapter concludes the dissertation by considering three areas. Firstly, the implications of the findings for theory and practice are considered. Secondly, the limitations and flaws of the present dissertation are considered. Thirdly, issues and considerations for further research of this topic are identified.

6.1 IMPLICATIONS FOR THEORY AND PRACTICE

The findings of the present study have no direct implications for theory, as the notion of positivism and scientific enquiry are not affected by these results. However, part of the process of quantitative research is not understood by postgraduate students of education, and this is borne out by the poor comprehension of statistical terms. It is in the practice of the teaching of research design, and specifically statistics and the presentation of research findings, that the present findings have implications.

To rectify these low standards an overall policy needs to be developed by policy makers. Those students who are electing to specialise in education should be required to have done a basic research design and statistics course at undergraduate level.

At graduate level there should be various courses offered for various needs. For example, a refresher course may be offered to those who have done a course in statistics and need to sharpen their memory to enable them to read research, whereas a basic course in statistics and in context (in research articles) is needed for those who have had no

exposure to statistics. Those being trained to be educational psychologists however, need training in all aspects of measurement literacy. All students training to be teachers (H.D.E.) need a brief course in psychometrics so as to be able to develop tests and use test results effectively when making decisions about pupils. South Africa is a developing country where the majority of people have a low level of education. Those postgraduate students (B.Ed student) are part of an elite group (in terms of education) who will be giving direction to others, helping to form policy and making decisions in the future that lies ahead. It is imperative that this elite group has a solid grounding in statistics so that they may continually keep abreast of research.

Higher degrees, requiring research papers as part of the degree requirements, should provide a compulsory course in research design and statistical techniques. As part of the course, postgraduate students should be required to carry out some straightforward but rigorous research of their own, based on some problem they have experienced in the field or to replicate some previous research.

Those postgraduate students of psychology in education [B.Ed. (Ed.Psych.) and M.Ed. (Ed. Psych.)] should be required to be competent in research design and statistical techniques, so as to carry out quantitative research. They should also be able to interpret psychometric tests, for example, to know what IQ tests measure, interpret test scores when the difference between verbal and non-verbal scores are significant, standard error of measurement and others. Classical and modern theory behind psychometric tests should be taught so that test strengths, weaknesses and purposes of usage are understood.

As previously stated, standards in statistical literacy need to be raised; otherwise research findings will either be incorporated into a reader's knowledge without evaluation, or the findings will be eliminated because the reader is unable to evaluate them. As most students face statistics with trepidation, and some hope soon to forget the unpleasant experience, it is necessary for them to be convinced of the necessity to understand statistics and their application to practical problems and research, so that they will not avoid reading research. Baker (1984) and Eisner (1984) suggest most teachers do not read research and it is possible that their lack of statistical knowledge may be one of the reasons for this.

The findings of the present dissertation also have implications for writers of research articles. If the writer uses complex statistics, the accessibility of the article to most (if not all) postgraduate students is limited. Writers who restrict themselves to measures of central tendency, percentages and frequencies have a wider readership as articles are more easily understood in their entirety. Such a journal article is Hammill and Blumberg's (1967) descriptive study on 'The Special Education Teacher's Understanding of Statistics', as the data are displayed and discussed in percentages. University academics are encouraged to submit written research to the most prestigious journals which often include fairly complex statistics, but which are often unable to be fully understood by the target audience. Thus more thought should be given to the reader of the article when a paper is written. However, the standard of the articles should not suffer as a result of a less complex portrayal of statistics. One way of coping with the above situation is for editorial boards (of journals) to allow journal articles to be lengthier so that more in-depth explanations of research design and statistics may occur. In many instances, the case for

the selection of a statistical procedure is not argued.

6.2 LIMITATIONS OF THIS STUDY

The present study replicates McLaughlin et al's. study (1973). The population sampled was limited to postgraduate students of education in the Pietermaritzburg area. Most primary school teachers in the Republic of South Africa are not graduate students and thus the findings cannot be applied to them. These primary school teachers may or may not have adequate training in statistics to understand and evaluate journal articles. Furthermore, training institutions in other geographical areas may place an added (or less of an) emphasis on statistical techniques.

Another aspect limiting generalisability was the fact that the population of the educational institution sampled was intact, therefore random sampling was not employed. This lack of random sampling and the sampling of 1991/2 subjects affects the inferences drawn from the findings, such as representativeness.

The statistical terms used in the measuring instrument were originally from five special education journals and are not representative of a cross-section of educational journals. However, statistical terms tend to be universal and it is considered that the students' performance is an indication of their comprehension of a variety of statistical terms that are frequently encountered in other education journals. The validity of the instrument used was not well established hence its generalisability is affected.

The demographic information was sparse, obtaining information about educational levels

of study and major subjects only. Using this information was limiting, as a postgraduate student may come from another university where statistics is required in a certain undergraduate course. Hence, dividing the group into those with or without statistics was arbitrary and may not fully represent the situation. More specific information about the number and date of statistical courses taken would have enabled the researcher to ascertain if the passage of time had an effect on a group or person's statistical comprehension. Trying to obtain extra demographic information is limited by the students themselves who often protest that it is discriminatory.

The sample size of the postgraduate psychology students of education is small compared with the other groups. However, the number represents the total intake of the educational psychologists in training. Ways of compensating for sample size would have been to sample other institutions or successive years of psychology students of education.

The total sample consisted of students of various cultural groups, including some for whom English was their second language. However, these students' knowledge of the English language was sufficient to cope with the demands of a postgraduate course. Nonetheless, their knowledge of English at the time they undertook a statistics course (those with a course in statistics) is unknown, and could have affected their understanding of statistics.

A decision was made not to reject any scripts that had missing answers to the statistical terms presented, as the subjects often left out definitions they did not know. Hence, it was difficult to detect and reject spoilt papers.

6.3 FUTURE RESEARCH

The findings of the present research support the trend found in the USA that paediatric residents, teachers and special education students have a low level of comprehension of statistical terms. However, the research undertaken on this topic is sparse and further replication is needed to verify the pattern and the importance of statistics so as to remediate and prevent long term effects. Such replication of research will benefit training institutions, educational and psychology faculties, education departments, publishers of tests such as the Human Science Research Council (HSRC) and research itself.

Direct replication of the study needs to be conducted on a national basis sampling various educational faculties or teacher training institutions in various geographical areas, so as to improve representativeness and generalisability and to ascertain the extent of statistical illiteracy.

Systematic replication using different groups such as psychologists, social workers, health workers could be studied as they are also required to keep abreast of research and use the findings in their treatment of their clients.

Other types of studies could include:-

- i) a survey of measurement staff so as to determine the problems associated with the teaching of statistical techniques
- ii) perceptions that students have of statistics before and after a course in statistics
- iii) a study to assess whether it is possible to raise the level of statistics
- iv) reasons for the low level of statistics

- v) the methods used in teaching statistics
- vi) studying students' understanding of statistics in the context of a research problem.
- vii) interviewing journal editors to assess the level of statistical knowledge required in research articles.

Suggestions for practice have been made from the findings of this study, but a further issue needs to be considered, namely whether these findings will inform or influence educational practice.

The low level of statistical understanding by postgraduate students of education suggests that they omit the method and results section when reading journal articles in which statistical techniques are used. However, a number of researchers suggest that teachers or practitioners for various reasons do not even read research (Eisner, 1984; Baker, 1984; Tizard, 1990). Critics of researchers claim that researchers are out of touch with practitioners (Eisner, 1984; Baker, 1984 and Tizard, 1990). Note needs to be taken of the criticisms so that relevant, practical ways of disseminating research are formulated whereby teachers are enabled to read research. Researchers and practitioners appear to be involved in two different worlds in that researchers are producing findings which are often contradictory (Chandler, 1981), idealised, bounded and complex, using academic jargon (Baker, 1984; Tizard, 1990). Practitioners are unconcerned with research findings as in their daily tasks other issues, such as making sure there is enough equipment for a lesson, sport fixtures or other concerns take preference. Thus their decision making will not be guided by research, but rather by their own experience which may include research findings that are distorted to fit their own experience. If this is so,

why should teachers or practitioners be concerned about statistical terms if they are not going to use them? Thus published research appears to be an academic exercise where the audience consists of colleagues and researchers, and the aim is to impress them so as to gain qualifications, promotion (within the ranks of the academia) and status.

Among researchers there are sceptics, pragmatists and those who believe in actively helping research to inform practice. Eisner is a sceptic of research in its present form, as he suggests "much of what we do in the schools is influenced in some way by such beliefs, models, images, metaphors. But when it is, can we legitimately claim that it is educational research that is informing practice? Rhetoric might be informing practice - but research? I wonder" (1984, p. 449). Whereas Baker (1984) disagrees with Eisner (1984) and opts for a "messier model that more closely approximates how the world works" as suggested by Guba and Clark (1974: cited in Baker 1984, p. 453). Here there is an interaction between practitioner and researcher to solve problems and adapt the situation to the classroom.

Baker writes "Academics don't know how to apply their knowledge" (1984, p. 454) and "The way in which knowledge is transmitted is often as important as the substance of the knowledge itself" (p. 453). Thus factors that influence the impact of research, the structures that disseminate research and those who control the structures or 'gateways' need to be considered if research is going to inform practice (Tizard, 1990).

The findings of the present dissertation, that postgraduate students' understanding of statistical terms is low, suggests that standards of statistical understanding need to be

raised so that in practice the postgraduate students are able to critically evaluate research. The structures that disseminate research are journals, educational supplements such as teacher magazines, contents of initial and in-service training as well as higher degree courses, and lastly the popular media. The factors that will influence the impact of the research through these structures will be the characteristics of research, the language used, its complexity and who wrote the research, and whether the policy makers are a highly centralised body.

Tizard (1990) suggests researchers approach the controllers of these structures and actively write articles or programmes for them. Thus an academic style will be used for academic journals whereas a practical style will be used for teacher journals, and a more lively writing style for the popular media. Heads of educational faculties, teacher trainers and policy makers need to be approached so that they are aware of the research findings and the need to raise statistical understanding. In this way different problems in training and various courses to be offered for different needs will be discussed, so that researchers and practitioners communicate and problems are solved. This relationship will enable researchers to research the problems perceived and articulated by the practitioner.

Tizard (1990) warns of researcher bias that accompanies such involvement and that policy makers and practitioners are also working within certain paradigms and often will only hear what they want to hear. To overcome this aspect Tizard (1990) suggests there should be a match between research and the 'guardians of the gateways' so that dialogue continues. Therefore for the standards of the understanding of statistical terms to be raised, there needs to be an ongoing liaison between policy makers, practitioners and

researchers, or otherwise research will not inform practice. If research is read with an understanding of its complete contents, then challenges in education may be faced with greater confidence and more informed decisions may be made.

REFERENCES

- Aylward, G.P. & Verhulst, S.J.** (1991). Data analysis techniques in behavioral pediatrics : Some practical advice. Developmental and Behavioral Pediatrics, 12, (6), 370 - 377.
- Baker, E.L.** (1984). Can educational research inform educational practice? Yes! Phi Delta Kappan, 65, 453 - 455.
- Boneau, A.C.** (1990). Psychology literacy : A first approximation. American Psychologist, 45, 891 - 900.
- Brigham, T.A.** (1989). On the importance of recognising the difference between experiments and correlational studies. American Psychologist, 44, 1077 - 1078.
- Bryman, A.** (1988). Quantity and quality in social research. London : Unwin Hyman.
- Burkhardt, E., Casady, D.R. & Forsyth, R.A.** (1971). Statistical comprehension for graduate students in physical education : Test and norms. Research Quarterly, 42 (3), 236 - 243.
- Chandler, H.N.** (1981). Research and teachers : The interface remains fracture. Journal of Learning Disabilities, 14 (10), 604 - 607.
- Cohen, J.** (1990). Things I have learned (so far). American Psychologist, 45 (12), 1304 - 1312.
- Davison, M.L., Damarin, F. & Drasgow, F.** (1986). Psychometrics and graduate education. American Psychologist, 42, 584 - 586.
- Derrick, T.** (1976). The criticism of inferential statistics. Educational Research, 19

(1), 35 - 40.

Dominowski, R.L. (1989). Method, theory and drawing inferences. American Psychologist, 44, 1078.

Eisner, E.W. (1984). Can educational research inform educational practice? Phi Delta Kappan, 65, 447 - 452.

English, H.B. & English, A.C. (1959). A comprehensive dictionary of psychological and psychoanalytical terms. New York : McKay.

Eysenck, H.J. & Kline, P. (1991). Measuring the low level of psychometrics. The Psychologist, 4, 107.

Farman, R.H. (1980). Evaluation of educational research. Unpublished monograph. Birmingham : University of Birmingham.

Hammill, D.D. & Blumberg, A. (1967). The special education teacher's understanding of statistics : A study. The Journal of Special Education, 1 (3), 321 - 325.

Heyns, I. de V. (1983). Some cautions in the use of statistical methods in educational research. In Steinberg, M.B. & Philcox, S.E. (Eds.). Research methods for higher degrees. Papers presented at a lecture arranged by the Higher Degrees Committee in the Faculty of Education in February 1983. Cape Town: University of Cape Town.

Hickson, J., Christie, G. & Schmukler, D. (1990). A pilot study of world view of black and white South African adolescent pupils : Implications for cross - cultural counselling. South African Journal of Psychology, 20 (3), 170 - 177.

Humphreys, L.G. (1987). Quantitative methodology. Then, now and the future. In

- Glover, J.A. & Ronning, R.R. (Eds.). Historical foundations of educational psychology. New York : Plenum Press.
- Lambert, N.M.** (1991). The crisis in measurement literacy in psychology and education. Educational Psychologist, 26 (1), 23 - 35.
- Lambert, R.F.** (1980-1981). Teacher attitudes on testing: A multiple perspective. College Board Review, 118, 13, 14, 29, 30.
- McLaughlin, J.A., Hinojosa, V. & Trlica, J.** (1973). Comprehension of statistical terms by special education students. Exceptional Children, 39, 408 - 412.
- Ruddell, R.B.** (1985). Knowledge and attitude toward testing : Field educators and legislators. The Reading Teacher, 38, 538 - 543.
- Sprinthall, R.C. & Sprinthall, N.A.** (1981). Educational psychology : A developmental approach (3rd ed.). New York : Random House.
- Steinberg, M.B.** (1983). Editorial comment. In Steinberg, M.B. & Philcox, S.E. (Eds.). Research methods for higher degrees. Papers presented at a lecture arranged by the Higher Degrees Committee in the Faculty of Education in February 1983. Cape Town: University of Cape Town.
- Tizard, B.** (1990). Research and policy : Is there a link? The Psychologist, 3, 435 - 440.
- Yaremko, R.M., Harari, H., Harrison, R.C. & Lynn, E.** (1986). Handbook of research and quantitative method in psychology for students and professionals. New Jersey : Lawrence Erlbaum Association.

APPENDIX 1 : WHAT IS MEASUREMENT LITERACY

To have achieved minimal literacy in measurement as a psychologist, an educator, a personnel manager, a counselor, or other human service practitioner, one would have acquired at least four types of knowledge:

1. Knowledge of basic assumptions that underlie rendering or quantification of observations, assigning objects or events to classes, ordering of units of observations from greatest to smallest, or transforming of the number of right and wrong test answers on a formal or informal test.

2. Familiarity with the general rules by which observations, rank orders, item scores, and individual difference data are translated into measurement units, such as frequency counts, probability estimates, measures of central tendency, and measures of variability.

3. Implied familiarity with concepts of validity and reliability and the ability to utilise these concepts in selecting, using, or interpreting numbers that are derived from the several approaches to educational and psychological measurement.

4. Knowledge of sources of error and the ability to apply appropriate standard errors of measurement in making psychological or educational diagnoses, classifications, inferences, or predictions.

SOURCE: LAMBERT, N.M., 1991 p. 24.

APPENDIX 2: RATING SCALE FOR ACCREDITING TEACHERS AS TESTERS

Surname & Initials of Teacher

(Block Letters) : Mr/Mrs/Miss: _____

Name of School: _____

PREPARING FOR TESTING

- 1.1 Use of alternate test booklets (where applicable)
- 1.2 Adequate spacing of pupils
- 1.3 Appropriate size of group
- 1.4 Suitable venue

COMPLETION OF BIOGRAPHICAL INFORMATION

- 2.1 Biographical information checked for correctness
- 2.2 Pupils age checked for correctness

INTRODUCING THE TEST

- 3.1 Appropriate explanation of purpose of test
- 3.2 Adequate rapport with and motivation of pupils

CODING AND USE OF ANSWER SHEETS

- 4.1 Effective demonstration of coding procedure
- 4.2 Pupils' understanding of coding checked

TESTING

- 5.1 Standardised instructions followed
- 5.2 Pupils' understanding of practice examples checked
- 5.3 Strict adherence to set time limits

SCORING AND RECORDING

- 6.1 Attention to accuracy when scoring and recording
- 6.2 Correct and accurate use of norm tables
- 6.3 Correct recording of IQ on Cumulative Record Cards
- 6.4 Correct recording of scores on form E.D.P. 26

GENERAL

- 7.1 Understands the meaning of IQ and normal distribution of IQ scores
- 7.2 Understands meaning of discrepancies between Verbal and Non-verbal IQ's
- 7.3 Appreciates need for confidentiality w.r.t. IQ scores
- 7.4 Understands the concept "Reliability"
- 7.5 Understands the concept "Validity"
- 7.6 Understands the concept "Standard Error of Measurement"
- 7.7 Understands "Percentile rankings"
- 7.8 Understands "Stanine rankings"
- 7.9 Appreciates need for, and attention to, secure storage of test material

SOURCE: REVISED GUIDE AND RATING SCALE FOR THE ACCREDITATION OF TEACHERS AS TESTERS - NATAL EDUCATION DEPARTMENT.

**APPENDIX 3 : DEFINITION OF TERMS TAKEN FROM ENGLISH &
ENGLISH (1959)**

1. **MEDIAN** - That score in a ranked distribution which has exactly half of the cases below it and half (or half minus one, when N is an even number) above it.
2. **STANDARD DEVIATION** - A measure of the dispersion or variability of a whole distribution (that provides an indication of the average amount by which scores deviate from the mean of the distribution).
3. **RANGE** - The distance from the highest to the lowest score or value in a distribution. (Sometimes the range is taken as the distance from highest to lowest, plus one unit. This is a crude measure of dispersion).
4. **LEVEL OF SIGNIFICANCE** - i) The value of alpha (α) selected for a significance test, sometimes expressed as a percentage. The accepted value of α under most circumstances is .05; thus most significance tests are carried out using the .05 or 5% level of significance.
ii) The smallest value of α at which the observed value of a test statistic would be significant.
5. **$p < 0.01$; $p < 0.05$** - The probability of obtaining a value of a test statistic that is less than 0.01 (1%).
- The probability of obtaining a value of a test statistic that is less than 0.05 (5%).
6. **MODE** - The score or category that has the greatest frequency of occurrence in a distribution.
7. **HYPOTHESIS** - An explanation of a complex set of data, admittedly tentative and not yet proved. Cf. speculation, for which supporting facts or arguments are very few and inadequate).

8. **MEAN (AVERAGE)** - A measure of central tendency calculated by dividing the sum of all the values of the number of cases in a statistical series: by the number of cases or observations.
9. **ANALYSIS OF VARIANCE (ANOVA)** - A method for determining whether the differences (expressed as a variance) found in a dependent variable, when it is exposed to the influence of one or more experimental variables, exceed what may be expected by chance.
10. **N** - The number of instances of whatever sort in the total population.
11. **F** - A measure of the probability of the beyond-chance difference (in inferential statistics) or a statistic used in estimating the chance probability of equalling or exceeding a given difference between the variances of two samples.
12. **CORRELATION** - The tendency to concomitant variation; the degree to which two (or more) variables vary together.
13. **COVARIANCE** - The tendency of two variables to change together, as measured by the mean of the products of the paired deviations of the two variables, taken from their respective means.
14. **df** - Symbol for degree(s) of freedom.
15. **r** - product - moment correlation coefficient.
16. **REGRESSION** - The fact that the predicted score or value of a dependent variable is closer to the mean of the sample than is the predictor score or value of the independent variable.
17. **VALIDITY** - A property of the whole measuring or testing process, but especially of the test instrument, that insures that the obtained test scores correctly measure the variable they are supposed to measure; the property of the measuring process

that makes the obtained scores useful in predicting a given variable.

18. **rho** - The correlation coefficient for data which are pairs of ranks (squared rank differences).
19. **RELIABILITY** - The complex property of a series of observations, of a measurement instrument or the entire measuring process, that makes possible the obtaining of a similar result upon repetition (the degree to which such similar results may be predicted - free from random influence).
20. **FACTOR ANALYSIS** - A statistical method for interpreting scores and correlations of scores from a number of tests.
21. **INTERACTION** - A mutual or reciprocal influence between two or more systems.
22. **CHI SQUARE TEST** - A formula by which to estimate the probability that a given, set of data conforms to what would be expected to be observed if a certain law or cause were in operation.
23. **PERCENTILE** - One of the 99 point scores that divide a ranked distribution into groups or parts, each of which contains 1/100 of the scores or persons.

**APPENDIX 4 : DEFINITION OF TERMS TAKEN FROM YAREMKO,
HARARI & HARRISON, (1986).**

1. **MEDIAN** - The middle value of a set of numbers arranged in order of magnitude (ie 50%tile). For small sample, the median usually is taken as the middle value of an odd number of scores, and halfway between the two middle values of an even number of scores.
2. **STANDARD DEVIATION** - A measure of the variability or dispersion in a set of scores that provides an indication of the average amount by which the scores deviate from the mean of the distribution. It is the square root of the variance, so the symbols and the general definitional formulas for the standard deviation are readily derived from those for the variance.
3. **RANGE** - A measure of spread or variability in a set of scores, defined as the largest score (L) minus the smallest score (s), but this is correct only if the scores are integer measures of a continuous variable.
4. **LEVEL OF SIGNIFICANCE** - i) The value of alpha (α) (ie. probability of a Type 1 error) selected for a significance test, sometimes expressed as a percentage. The accepted value of α under most circumstances is .05; thus, most significance tests are carried out using the 0.05 or 5%, level of significance.
ii) The smallest value of α at which the observed value of a test statistic would be significant.
5. **$p < 0.01$; $p < 0.05$** - The probability of obtaining a value of a test statistic that is less than 0.01 (1%).
- The probability of obtaining a value of a test statistic that is less than 0.05 (5%).

6. **MODE** - The score or category that has the greatest frequency of occurrence in a distribution.
7. **HYPOTHESIS** A statement about the relationship among events.
8. **MEAN** Any of several measures of central tendency, for a distribution of scores. Usually refers to the arithmetic mean, AM - found by dividing the sum of the measurements by the number of measurements.
9. **ANALYSIS OF VARIANCE** - A set of procedures for testing the significance of differences among means. The total variability of the observation from an experiment is divided into those portions that are attributable to chance factors or experimental error.
10. **N** - Not defined in Yaremko et al. (1986).
11. **F** - Not defined in Yaremko et al. (1986), but F ratio is defined.
12. **CORRELATION** - A general term describing the relationship between two variables that are associated or co-vary.
13. **COVARIANCE** A measure of relationship between two variables. It is defined as the average (for a sample) or expected value (for a bivariate population) of the products of deviations of paired scores for their respective means.
14. **df** - Symbol for degrees of freedom A statistical concept associated with many tests of significance (eg t, F, χ^2) and with estimates of variability. In general, it is the number of observations on which the statistic is based, minus the number of restrictions placed on the freedom of these observations to vary.
15. **r** - Symbol, any index of the degree of relationship between two variables, specifically, the correlation coefficient refers to the Pearson product - moment correlation, an index of the degree of linear relationship between two variables

- that are measured on an interval or ratio scale.
16. **REGRESSION (ANALYSIS)** - A general term for the statistical procedure concerned with fitting a regression or prediction equation to data, estimating parameters, and testing hypotheses about the true relationship between the dependent or criterion variable and the independent or predictor variable(s).
 17. **VALIDITY** - A general term denoting correctness of a measure (measures what it is supposed to measure).
 18. **r_{rbo}** - The correlation coefficient for data which are pairs of ranks.
 19. **RELIABILITY** - A general term denoting consistency of measurements derived from repeated observations on the same subject under the same circumstances.
 20. **FACTOR ANALYSIS** - Multivariate statistical methods for describing the relationships among a large number of variables in a correlation matrix by reducing them to a few relatively independent but conceptually meaningful composite variables called factors.
 21. **INTERACTION** - In experiments with two or more independent variables, the condition that exists when the effect of one independent variable differs at different levels of the other independent variables.
 22. **CHI SQUARE TEST** - Any of several tests statistics that have a chi-square distribution.
 23. **PERCENTILE** - One of the 99 score points that divide a distribution into hundredths. The percentiles are labelled P1 through P99.