Edith Cowan University

Research Online

Theses : Honours

Theses

1993

An analysis of errors made both on computer administered tests and paper and pencil tests and a comparison between them

Gary Fullarton
Edith Cowan University

Follow this and additional works at: https://ro.ecu.edu.au/theses_hons

 Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation

Fullarton, G. (1993). *An analysis of errors made both on computer administered tests and paper and pencil tests and a comparison between them*. Edith Cowan University. https://ro.ecu.edu.au/theses_hons/595

This Thesis is posted at Research Online.
https://ro.ecu.edu.au/theses_hons/595

AN ANALYSIS OF ERRORS MADE BOTH ON COMPUTER ADMINISTERED TESTS AND PAPER AND PENCIL TESTS AND A COMPARISON BETWEEN THEM

BY

GARY FULLARTON BA

A Thesis Submitted in Partial Fulfilment of the Requirements for the Award of Bachelor of Education (Honours)

Faculty of Education

Edith Cowan University

Submitted 17th December 1993

ABSTRACT

Sound assessment is essential in mathematics education. Computer Administered Testing is one measuring device being researched for the assessment of mathematics. A concern with the advent of Computer Administered Testing is the possibility that the mode of testing influences how students perform on tests (Bugbee & Bernt 1990; Ring 1993). This study applied Newman Error Analysis Interviews to investigate if first year Bachelor of Arts Primary Education students exhibit different error patterns for mathematics test items in a Computer Administered System compared with those in an equivalent Paper and Pencil Test.

The implementation of the Computer Administered Test did not appear to significantly affect the students' ability to read the question, understand what the question was asking them to do, transform from the words of the question to an appropriate mathematical strategy, perform the mathematical operations necessary and express the solution in an acceptable form, in comparison to the equivalent Paper and Pencil Test.

There was a significant difference in the number of Careless Errors made by the students. It was reasoned that non-cognitive variables contributed to the Careless Errors. Non-cognitive variables identified as possibly contributing to the difference of Careless Errors between the two modes of testing were the differences in the time to complete the tests, Computer Anxiety and Intimidation.

DECLARATION

I certify that this thesis does not incorporate, without acknowledgment, any material previously submitted for a degree or diploma in any institution of higher education; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in text.

Signature:

Date: February 24, 1994

ACKNOWLEDGMENT

The author wishes to thank Don Smallman, Professor Nerida Ellerton, Doctor Jack Bana, Doctor Geoff Ring, Rod Ellis, Michael Bentley and Leonie Ramondt for their professional guidance and support throughout the study. The author would also like to thank those students who gave up their time to participate in this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

BACKGROUND OF THE STUDY

Educators are always looking for better ways of assessing mathematics.

Students in first year Bachelor of Arts Primary Education are required to take a Bridging Mathematics, Number/ Space/ Measurement, Course "MPE 0101" where a mastery test is given in which the student needs 75 percent or better to pass the unit. One measuring device being researched for this unit is Computer Administered Testing.

A concern with the advent of Computer Administered Testing is the possibility that the mode of testing influences how students perform on tests (Bugbee & Bernt 1990; Ring 1993). Completion of computerised testing has been found to enhance (Bugbee & Bernt, 1990; Liefeld & Herrman, 1990;), hinder (Lee, Moreno & Sympson, 1986) and not affect performance (Olsen, Maynes, Slawson & Ho, 1989; Eaves & Smith, 1986). Research has not provided an understanding of why situational factors affect assessment in the way they do. Such research is essential for a sound theory of assessment (Webb 1992).

PURPOSE OF THE STUDY

While most studies have investigated score differences (Bugbee & Bernt, 19900; Lee, 1986; Olsen, Maynes, Slawson & Ho, 1989; Eaves & Smith, 1986; Spray Ackerman Reckase & Carlson, 1989) this study, through the use of interviews, will investigate the error patterns the students exhibit in both the Computer Administered and the Paper and Pencil Tests. This research aims to complete an error analysis on both the Computer Administered and Paper and Pencil Tests which will indicate at what stage in problem solving are errors occurring.

PROBLEM STATEMENT

Do students exhibit different error patterns for mathematics test items in a computer administered system compared with those in an equivalent paper and pencil test?

RESEARCH QUESTIONS

Are there differences in the number of reading, comprehension, transformation, process skills, encoding or careless errors made in a mathematics competency test taken by first year student primary teachers, between the computer administered test and the paper and pencil test?

SIGNIFICANCE

This study is significant in that, in addition to previous studies, it should offer teachers/students a better understanding of the comparability of Computer Administered Testing and Paper and Pencil Testing, and it should also provide an avenue for further research into the use of Computer Administered Testing. It is essential that state-of-the-art Computer Administered Testing systems be thoroughly evaluated so that the full potential of the computer-based medium for testing can be realised (Ring, 1993). This study is part of a larger evaluation of the Computer Administered System which aims to investigate:

1. student attitudes and anxiety with respect to mathematics, computers and tests (both paper and computer-based);
2. student performance on the Computer Administered Test;
3. the attitudes of instructors and mathematics education experts to Computer Administered Testing;
4. the relationship between these variables and those of student age, gender, academic ability, mathematical ability, and computer experience.

Whereas previous studies have investigated score differences this study specifically identifies error patterns between Computer Administered Testing and Paper and Pencil Testing and thus contributes to a more comprehensive evaluation of the two testing modes. Piaget was struck by the observation that the character of errors made by students held as much importance as the nature of their correct answers (Shulman, 1970 in Radatz, 1979).

CHAPTER TWO

RELATED LITERATURE

ASSESSMENT

Mathematical assessment refers to the comprehensive accounting of an individual's or group's functioning within mathematics or in the application of mathematics (Webb 1992). Properly constructed assessment constitutes an important source of feedback to the learner, and this is intimately bound up with the learning process itself (Gagne 1971 p 339).

The purpose of assessment, according to Webb (1992) is to:

1. be used as a tool by teachers to provide evidence and feedback on what students know and are able to do;
2. express what is valued regarding what students are to know, do or believe;
3. provide information to decision makers and thus be used as a tool to impose on teachers and schools a direct measurable accountability for their effectiveness;
4. provide information on the effectiveness of the Education System as a whole.

One instrument for measuring mathematical assessment is testing. "Tests are systematic procedures of observing behaviour and describing it with the aid of numerical scales or fixed categories."

(Cronbach 1989 in Webb, 1992) When administering tests attention must be paid to the form of each item the test contains. Tests must be designed to measure the objective specifically, not in a general sense. Tests must pose questions that reflect directly the defined objectives of learning (Gagne 1971 p 340 -342).

COMPUTER ADMINISTERED TESTING

The Computer Administered Test is one form of testing being used more commonly in university settings. "The term 'Computer Administered Testing' implies a completely automated environment whereby the test is constructed using the computer, the students complete the test at the computer and the student responses are concurrently scored by the computer" (Ring 1993).

Anderson and Trollip (1982) suggest the following three principles are the basis for designing a computer administered testing system.

1. Ensure easy access to needed information. (Instructors should have access to test results and students should be able to obtain directions on how to use the testing system).

2. Maximise user control. (Students should be able to change answers).

3. Install safety barriers and nets. (Safety barriers prevent or impede undesirable action and safety nets enable recovery from undesirable action).

These apply equally to instructors and students.

The advantages of Computer Administered Testing are:

1. more standardised test administration conditions;
2. improved test security;
3. individually administered tests;
4. immediate test scoring and feedback;
5. increased variety of testing formats;
6. ability to collect test item latency information.

(Olsen, Maynes, Slawson & Ho; 1989; Wise & Plake, 1990)

Moe and Johnson (1988) reported students considered the most positive aspect of the computerised test as being ease of answering with the most negative aspect being glare.

POSSIBLE INFLUENCING FACTORS

Studies have investigated different factors that may impinge on student achievement in computer administered testing. Lee (1986) reported that past computer experience significantly affected performance on the computerised test; however there was not a significant difference between "low experience" and "high experience" persons indicating that minimal work with computers may be sufficient to prepare a person for computerised testing. This was contradictory to findings by Eaves and Smith (1986), Wayne Plumley & Ray (1989) and Moe and Johnson (1988) where results implied microcomputers *do not lead to poorer results* among students with little or no computer experience.

Test flexibility is another important factor in Computer Administered Testing. Spray, Ackerman, Reckase & Carlson (1989) found there was no significant difference when the item presentation software was written to mimic as closely as possible the flexibility of the paper and pencil format. "If score equivalence between item presentation media is required then test taking flexibility under both conditions needs to be equivalent" (Spray et al 1989). Lee, Moreno and Sympson (1986), Ronau & Battista (1988) studied the effect of item presentation medium where examinees were only able to refer to previous items or to change an answer once the answer was recorded in the paper and pencil test. Both studies found the paper and pencil group scored significantly higher than the computerised group. However, Eaves and Smith (1986) consciously had a similar situation where flexibility was only possible in the Paper and Pencil Test and not in the Computer Administered Test but there was no significant difference in the scores between the tests.

According to Wise and Plake (1990) the most common findings of studies concerned with ability and achievement testing are : (a) the reliabilities of computer based and conventional tests are very similar and (b) computer based testing yields scores lower than those of conventional testing, though score differences are typically non-significant.

Research has investigated many factors which may influence student performance in Computer Administered Testing, however few studies (Ronau & Battista, 1988; McDonald, Beal & Ayers, 1992) have compared error patterns. No one, to the author's knowledge,

has applied interviews to compare the error patterns between Computer Administered and Paper and Pencil Testing.

ERROR ANALYSIS AND THEORETICAL FRAMEWORKS

The model used for this study is The Newman Hierarchy of Error Causes. The following section examines various other studies on the analysis of students' errors and how these relate to Newman's model.

Errors provide rich insights to the nature of students' mathematical thinking and the school mathematics learning environment (Siemon 1988). Errors in the learning of mathematics are not simply the absence of correct answers or results of unfortunate accidents. They are consequences of definite processes which must be discovered (Radatz, 1979; Geiger, 1990; Perso, 1992). An individual error pattern can throw considerable light on why the individual makes mistakes on mathematical tasks (Clements 1980) and is of benefit to educators in the quest to improve children's' mathematical performance (Bainbridge, 1981). Analysis of errors offers a variety of points of departure for researching the processes by which students learn mathematics (Radatz 1979).

Roberts (1968) compiled the following four error categories as a means of discovering pupils computational skill deficiencies. They were;

1. Wrong operation: pupil attempts to solve the problem with an inappropriate operation;

2. Obvious computational error: pupil attempts to solve the problem using an erroneous basic number fact;
3. Defective algorithm: pupil attempts to solve a problem employing other than basic number fact errors or inappropriate operation errors;
4. Random responses: pupil attempts to solve a problem in a way showing no discernible relationship to the given problem.

Englehardt (1977) extended the work by Roberts (1968) which led to the identification of eight error types namely:-

1. Basic Fact Error: The pupil responds with a computation involving an error in recalling basic number facts.

2. Defective Algorithm: The pupil responds by executing a systematic (but erroneous) procedure.

3. Grouping Error: The pupil's computation is characterised by a lack of attention to the positional nature of our number system.

4. Inappropriate Inversion: The pupil responds with a computation involving the reversal of some critical aspects of the solution procedure. Computations classified as inappropriate inversions displayed reversals of steps in algorithms which often appeared to promote faster responses.

5. Incorrect Operation: The pupil performs an operation other than the appropriate one.

6. Incomplete Algorithm: The pupil initiates the appropriate computational procedure, but aborts it or omits critical steps.

7. Identity Errors: The pupil computes problems containing O's and 1's in ways suggesting confusion of operation identities, eg; 5 x 1 = 1.

8. Zero Errors: The pupil computes problems containing O's in ways suggesting difficulty with the concept of zero.

Cox (1975a; 1975b) classified errors as systematic, random and careless. Systematic errors are those computational errors that occur in at least three out of five problems for a specific algorithmic computation. They show a pattern of incorrect responses. The student will likely make the same error when encountering similar computational problems. Random errors occur in at least three out of the five problems but contain no discernible pattern, and therefore they are difficult to remediate. Careless errors occur in one or two out of five problems for a specific algorithmic computation. The child basically knows how to perform the correct computation but due to distractions, boredom or a lapse in attention he or she makes careless errors. This classification represents the regularity of a particular error but does not explain why the error occurred.

Cox's (1975a; 1975b), Roberts'(1968) and Englehardt's (1977) classifications of errors relied solely on students written work and were limited to arithmetic problems. A more general approach was that of Radatz (1978) who had the following five categories in his classification of errors;

1 Language problems.
2 Errors due to difficulties in obtaining spatial information.
3 Deficient mastery of prerequisite skills, facts and concepts.
4 Incorrect association or rigidity of thinking.
5 Application of irrelevant rules or strategies.

Bainbridge (1981) offered the following more detailed practical classification for identifying individual differences:

1. inability to read the question;
2. inability to interpret the question;
3. not understanding a particular mathematical term;
4. unfamiliarity with a particular process, eg. long division;
5. difficulty with computing accurately;
6. carelessness;
7. basic lack of conceptual understanding, eg. of a "fraction",
8. inability to apply or transfer learning to different contexts.

Radatz (1978) believed that a close interaction among causes could result in the same problem giving rise to errors from different sources. For this reason he believed a definite classification and hierarchy of error causes seemed impossible to achieve.

Clements, (1980) believed that despite Radatz's pessimism, Newman's Error hierarchy (1977) and Casey's (1978) extension and refinement of the hierarchy have provided data of a kind not to be found in earlier error analysis research. Casey (1978) classified his errors as follows;

1 Question form.
2 Question reading.
3 Question comprehension.
4 Strategy selection
5 Skill selection.
6 Skill manipulation.
7 Answer presentation.
8 Unknown block.
9 Known block

This study is based on the Newman hierarchy of error causes. This hierarchy was established as a criterion for error causes, which would be applicable to most situations where pupils were attempting to solve a variety of mathematical problems which were presented in written form (Newman 1977).

Newman's hierarchy of error causes as reported in Clements (1980, p4) is shown in figure 1 on the following page.

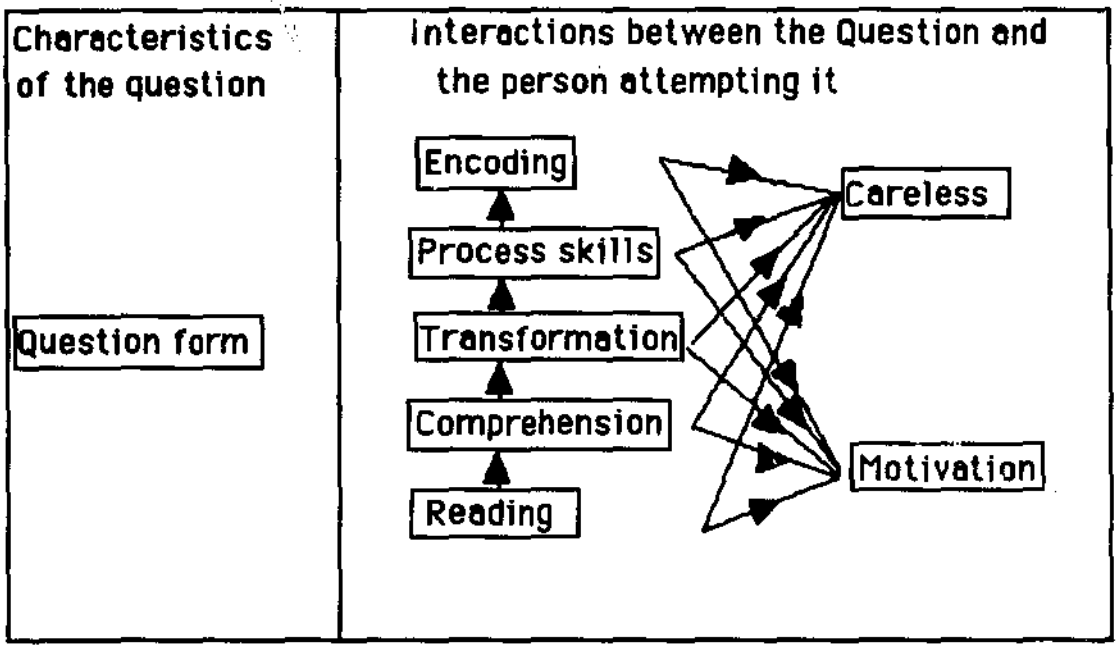

Figure 1. The Newman hierarchy of error causes (from Clements, 1980, p. 4).

Errors due to the form of the question are different from errors in other categories because the fault lies with the person constructing the question rather than the person attempting it (Clements 1980). Casey (1978) included the question form in his hierarchy.

Two categories, Carelessness and Motivation, have also been shown as separate from the hierarchy as these types of errors can occur at any stage of the problem solving process. A Careless error, for example, could occur at the reading or comprehension level. Likewise, having read, comprehended and worked out an appropriate strategy for solving a problem, lack of motivation may prevent the student from proceeding further in the hierarchy (Ellerton & Clements, 1993).

According to Newman (1977), in order to solve a mathematical problem the student must;

1 read the problem;
2 comprehend what is read;
3 carry out a mental transformation from words of the Question to the selection of an appropriate mathematical strategy;
4 apply the process skills demanded by the selected strategy;
5 encode the answer in an acceptable written or typed form.

An error is classified as reading if the student is unable to read a key word or symbol in the written problem to the extent that this prevented him/her from proceeding further along the appropriate problem solving path. Comprehension errors occur when the student is able to read all the words in the question, but cannot grasp the overall meaning of the words and, therefore, is unable to proceed further along an appropriate problem solving path. Transformation errors occur when the student, while understanding the question, is unable to identify the operation, or sequence of operations needed to solve the problem. A Process skill error occurs when the student identifies the appropriate operation but does not know the procedures necessary to carry out these operations accurately. Encoding errors occur when the student correctly solves a problem, but is unable to express this solution in an acceptable form (Newman, 1977; Marinas & Clements, 1990).

Newman used the word 'hierarchy' because she reasoned that failure at any level of the sequence prevents problem solvers from

obtaining satisfactory solutions unless they arrive at the correct solution by faulty reasoning (Ellerton & Clements 1993). Newman's use of the word 'hierarchy' is different from the way it is used in literature, such as Gagne (1971) on learning hierarchy (Clements, 1980; Ellerton & Clements, 1993). Newman's framework for the analysis of errors was not created as a rigid information processing model of problem solving. It was designed to complement rather than challenge descriptions of problem solving processes such as those offered by Polya (1973). With the Newman approach the researcher is attempting to stand back and observe an individual's problem solving efforts from a coordinated perspective, whereas, Polya (1973) focused on the richness of the Comprehension and Transformation levels, as defined by Newman (Ellerton & Clements 1993).

Newman's method of analysing errors differed from those such as Cox (1975a; 1975b), Roberts (1968) and Englehardt (1977) in that interviews were implemented to identify the type of error. A well documented method of analysing errors in mathematics is through interviews. Lankford (1974) stated that interviews determine a person's pattern of thinking as he/she computes. Through individual oral interviews a teacher will soon become aware of a wide variation of computational strategies employed by his/her pupils. Inferences about a student's thinking drawn from his/her written response alone represents little more than guesswork on the part of the researcher. Structured interviews, where the student attempts to verbalise his/her thinking, must be conducted before consistent error patterns can be determined with any degree

of certainty (Clements, 1980; Casey, 1978; Englehardt, 1977; Dickson, Brown & Gibson, 1984).

PREVIOUS STUDIES

The Newman error analysis interviews have been used extensively since their introduction in 1977 as a means of analysing students' errors. The following table displays the percentage of errors in each category in a number of studies using this technique.

TABLE 1. Studies Using The Newman Technique

Type of Error	Newman (1977) 6th grade	Clements (1980) 7th grade	Clarkson (1980) 10th grade	Clarkson (1980) 11th grade	Ellerton & Clements (1993)
Reading Errors	13	2	1	0	7
Comprehension Errors	22	9	3	7	28
Transformation Errors	12	27	10	8	26
Process Skill Errors	26	26	36	37	22
Encoding Errors	2	1	2	4	2
Careless Errors	25	35	49	44	15

Clarkson (1980) suggested that errors tend to concentrate in both process skills and carelessness categories in higher year levels. These studies have not looked at errors at the undergraduate level. A study by Faulkner in Ellerton & Clarkson (1992) used Newman

techniques in research investigating errors made by nurses undergoing a calculation audit. She found that the majority of errors were of the comprehension and transformation category.

Studies have examined the use of computers for diagnosing students errors (McDonald, Beal & Ayers, 1988,1992; Travis, 1984; Ronau & Battista, 1988). McDonald et al (1988) used the work of Cox (1975a), Roberts (1968) and Englehardt (1977) to classify errors made on the computer. The table below shows the break up of errors for students in Years Two to Five:-

TABLE 2. Arithmetic Computer Errors

Error category	Frequency	Percent of errors
Operational	1	0-1
Inversion	37	7
Algorithm	244	44
Basic Fact	94	17
Unidentified	178	32
TOTAL ERRORS	554	

McDonald et al (1988) also tried to identify errors which are related to the students use of microcomputers. The types of errors included:

1. The key held down for an amount of time repeats the key input.

2. Inversion errors could be explained by the computer input routine of entering answers from left to right.
3. Transfer errors from paper to computer.
4. Keyboard error where students missed the desired key.

The computer errors may have related to 29% of errors however McDonald et al (1988) concluded it was not possible to confidently determine the actual amount without a study that compares error patterns in a computer environment and a traditional paper and pencil environment.

Ronau & Battista, (1988) used a diagnostic computer package to compare ratio and proportion error patterns on eighth grade students between the microcomputer test and a paper and pencil test. Students' responses in both tests were classified as follows:-

1. Unclassified: student response did not fit into one of the categories below.
2. Correct.
3. Omitted: student skipped the item.
4. Inverted: student inverted the ratio.
5. Improper Addition: student used addition to form an incorrect numerator or denominator in ratio.
6. Simplify: correct, but student did not simplify the ratio.
7. Subtraction: student employed subtraction to solve a proportion.
8. Multiplication: student incorrectly used multiplication to solve a proportion.

9. Algebra: student formed the correct proportional equation but solved it incorrectly.
10. Incorrect only: student chose the incorrect response to YES/NO item.
11. Inverted Addition: combination of errors 4 & 5 for ratio.
12. Procedure Error: multiple classifications possible: errors 7, 8 and 9.

Distributions of errors differed significantly on both data sets for Unclassified Omitted, Simplify and Subtraction. There was however no flexibility in the tests which could effect the number of Omitted errors.

McDonald et al (1992) also, through computer diagnosis, compared errors in whole number computation skills of Grade Three students between the computer and paper administered versions of the test. Contrary to Ronau & Battista (1988), McDonald et al (1992) found there was no significant difference in errors in all categories namely Reversal of Answer, Operational Error, Algorithmic Error, Transfer Error, Basic Fact Error, Blank Response and Unexplained Error.

McDonald et al (1992) also examined the Systematic, Random and Careless Errors, as defined by Cox (1975a; 1975b) and found there was not a significant difference in the means of errors between the paper and computer test.

CHAPTER THREE
METHODS

The following section describes the design of the research, the sample participating in the research and the instruments that were used to collect the data. It also provides an outline of Data Collection procedures, followed by a justification of the methodology of the study.

DESIGN

Fifty students, from two classes, randomly chosen from the five MPE 0101 unit of study were randomly assigned to one of two groups, within each class. One group in each class sat the Computer Administered Test first, and four or eight days later sat the paper and pencil test. The other group sat the paper and pencil test first and then the computer administered test four or eight days later. McDonald et al (1992) applied a similar design when conducting their research. Both classes participated in a familiarisation session of one hour duration with the computer, approximately a week before they sat the first test. In this session the students were guided through sample questions and given explanations of how the system operated.

Counterbalancing in the design allowed for control of the order effect on student performance (Olsen, Maynes, Slawson and Ho 1989). Students were told that their best result would go towards

their result for the unit. This encouraged effort in both settings, as well as ensuring no one was disadvantaged.

SAMPLE

Twenty one students, consisting of seventeen females and four males, who sat the Paper and Pencil Test first, were subjected to the Newman error analysis interviews. Twenty two students, consisting of nineteen females and three males, who sat the Computer Administered Test first, were subjected to the Newman error analysis interviews. The age of the students ranged from approximately seventeen to forty years. The time of each interview depended on the number of errors the student made. The students were asked to sign a form giving consent for the two tests and the interview.

INSTRUMENTS

The computerised test employed a state of the art Computer Administered Test called "Skillmath".

The "Skillmath" test was characterised by powerful navigational tools, elements of multimedia, variety of question formats, high level of learner control and procedural help.

The system used conventional test strategies and was based on three fundamental principles as seen in Ring (1993):-

1 Students should not be disadvantaged in comparison with non computer based testing.

2 Full use should be made of the computer based medium.

3 Instructor control over the testing environment should be maximised.

The test consisted of 71 questions comprising of 42 multiple choice type questions requiring students to either click/touch (using a mouse to identify areas), "move objects" (directly manipulating objects on the screen using a mouse), or "pull down menus" (selecting from items in a temporary list that overlays the screen on demand). There were 27 questions which required text entries and 2 questions requesting students to mark a given position on a number line.

The use of open ended questions had not previously been used for this unit, however open ended questions appeared to be superior in describing skills that students' possess and could be better diagnosed with respect to bugs or sources of misconception underlying the response pattern (Birenbaum & Tatsuoka, 1987b; Bridgeman, 1992).

The system allowed students options such as previewing, marking questions for review and sample question practice as well as allowing students to change responses or not respond at all. Student and system files were updated after each response to guard against loss of data if the test is terminated due to machine or power failure. The system ran through a network on Apple Macintosh VXII machines.

The split half reliability coefficient, using the odd and even question numbers, for the students who sat the Computer Administered test first, was 0.860 which indicated good internal consistency of the test.

The Paper and Pencil Test was written to mimic as close as possible the question format and objectives of the Computer Administered Test. Students were required to answer on the test booklet next to the question as this is closer to the computer situation and research shows that students' fared better when work space was provided adjacent to the test items (Hembree, 1986).

The split half reliability coefficient, using the odd and even question numbers, for the who sat the Paper and Pencil test first, was 0.885 which also indicated good internal consistency of the test.

An Alternate Forms Reliability Test using Edstats (Knibb, 1993) was conducted on the two tests. The subjects who sat the Paper and Pencil Test first, then the Computer Administered Test produced a coefficient of equivalence of 0.934. This indicates good alternate forms reliability. However, the subjects who sat the Computer Administered Test first, then the Paper and Pencil Test produced a coefficient of equivalence of 0.639. This group were given feedback from the computer at the completion of the first test. The more informative the feedback with regard to the correct answer, the more likely the second test results are to be correct (Birenbaum & Tatsuoka, 1987a). This appears to be the reason for the difference between the two equivalence coefficients.

The instrument used for error analysis was the 'Newman Error Analysis Interview'. In order to classify the errors the students were asked to redo the erroneous questions and then requested to:

1 Please read the question to me (Reading)

2 Tell me what the question is asking you to do. (Comprehension)

3 Tell me a method you can use to find the answer to the question. (Transformation)

4 Show me how you worked out the answer to the question. Explain to me what you are doing as you are doing it. (Process Skills)

5 Now write down your answer to the question. (Encoding)

If the student answered the questions correctly in the interview they were asked if they had answered the question differently in the test and, if so, for what reasons. If the student was unaware of why they answered the question differently it was assumed that the error was a careless one.

A study by Clarkson (1986) showed that the types of errors were consistent and thus it was a sound assumption that the types of errors made in the test were the same type as in the interview.

Audio tapes of the students' responses to the requests were kept and a sample of 31 questions from three students were checked by an independent assessor to ensure that the type of errors recorded are reliable. Of the 31 there were 24 that agreed. In view of the fact that the independent assessor had access only to the tapes of

the interviews and the students' responses, and was not present at any of the sessions, it is perhaps not surprising that most of these errors were classed as Transformation by the independent assessor, while the author tended to differentiate these into either Process Skill or Careless categories.

DATA COLLECTION

The Newman error analysis interviews were conducted two to nine days after the second test by the author and two research assistants, all of whom participated in a training session in Newman Error Analysis Interviews given by Professor Nerida Ellerton. Tables of each student's errors were constructed with their original response in the exam recorded. Even after an incorrect response was given the remainder of the five requests were made in order to verify that the initial error was the one that first caused the incorrect answer. After each question the type of error was recorded. Where time permitted the solutions to the questions were discussed with the student once the interview was compieted.

It was decided that each interview would only be conducted on those questions viewed by the student. It was observed that all students had the opportunity to view all the questions in the paper test. However in the Computer Administered Test some students had used their allotted ninety minutes with up to fourteen questions remaining. Therefore, these students were deemed not to have erred on these questions.

Once the interview was complete the responses to the questions were checked against the type of error first indicated, to make sure

they matched. The questions were administered identically in the interview as in the test. The interviews were conducted on the errors of the sample from each group's first test, as the errors made in the second test may have been influenced by the first test.

JUSTIFICATION OF METHODOLOGY

While this study uses a test format to determine where errors occur on test questions, the methodology employed is one in which students are interviewed individually and asked specific tasks relating to questions on which they have erred during the test, to determine what type of error was made and hence enable the author to compare error patterns between a Computer Administered Test and a Paper and Pencil Test.

Previous studies (Ranou & Battista, 1988, McDonald et al, 1992) compared errors between Computer Administered Testing and Paper and Pencil Testing using written work only. The computer package employed to diagnose errors of McDonald et al (1992) was adapted from Englehardt (1977). Englehardt, as a limitation of his study, said :-

> " Examining written performance without the opportunity to investigate a given error further greatly increased the possibility of misjudging a pupil's erroneous approach. Additional studies need to be conducted in which the inference of the pupils' approaches to incorrect responses is based upon clinical interviews".

It seems that the Diagnostic Interview Technique offers a more valid method for assessing pupils' abilities and difficulties, since with this method the pupil has the chance to either verbalise or demonstrate his or her approach to any particular mathematical task. Further the interviewer is able to observe first hand the pupils' working pattern (Newman, 1977). Orey and Burton (1992) concluded that the interview was a more comprehensive view of students' knowledge of mathematics. An error pattern using computer diagnostic systems may be wrong from the start.

Without the application of interviews it is very difficult to identify language errors. Pinchback (1991) declared it a mistake not to include language type errors, as they represent the language of correctly translating the meaning of the mathematical symbolism. Language and mathematics are intimately connected and it is wrong to try to separate one from the other (Ellerton & Clements, 1991 p7) Watson (1980) used the Newman technique on a similar age group to that of McDonald et al (1992) and found that over fifty per cent of the errors were language based. Therefore, for a comprehensive comparison of error patterns between the Computer Administered Test and the Paper and Pencil Test interviews need to be employed.

CHAPTER FOUR

RESULTS

From the interviews the following data was gathered on errors concerned with the first test students' attempted.

The majority of reading errors occurred when the student added or deleted the word 'not'.

Examples of comprehension errors identified through the interviews indicated that some students believed;

1. a question asking them to find the length of a fence around a garden was asking them to find the area of the garden,
2. a question asking for the length of Western Australia's Eastern boarder required them to estimate the length of Australia's Eastern boarder,
3. a question asking "How many numbers are there between 5.36 and 5.37"?, meant "How many whole numbers are there between 5.36 and 5.37"?

In all of these examples, the students did not comprehend what the question was asking them to do.

Examples of transformation errors found in the interviews indicated that some students did not know;

1. how to convert metric units such as cubic metres into cubic centimetres and grams to kilograms,
2. the procedure for finding the area of a triangle or circumference of a circle.

Examples of process skill errors found in the interviews indicated that some students exhibited errors in estimating time and distance, as well as multiplying 6 to the power of 8 when the question asked for 6 to the power of 7. In these types of errors the student knew which mathematical strategy to employ but could not work the strategy appropriately.

Examples of encoding errors found in the interviews indicated that some students:-

1. responded in dollars when the question requested the answer in cents,
2. rounded answers to whole numbers when they were not requested to do so,
3. one student could not locate the pull down menu in which the answer was to be marked.

Some students had difficulties moving objects such as rulers on the computer screen and this was the initial cause of the error. For example one student left out all the questions with rulers because

of the difficulty she first encountered with them. For this reason a category of 'Test Difficulty' was established for such students.

Through the interviews, it was noted that students learnt information concerning some questions between the tests and before the interviews. Where possible in these instances the type of error originally committed was recorded, however this was not always possible and thus an extra category of 'Learning' was created. This category does not identify the error but explains what happened between the test and the interview.

Interviews were conducted on a total of 378 errors which occurred in the Pencil and Paper Test. The distribution of the errors are displayed in the table below.

TABLE 3. PAPER & PENCIL ERROR FREQUENCY & PERCENTAGE

Type of error	Frequency	Percentage
Reading errors	1	0.26
Comprehension errors	45	11.9
Transformation errors	134	35.45
Process Skills errors	91	24.07
Encoding errors	7	1.85
Careless errors	100	26.46
Learning	1	0.26
TOTAL ERRORS	378	

Interviews were conducted on 493 errors which occurred in the Computer Administered Test. The distribution of errors are displayed in the table below.

TABLE 4 Computer Administered Test Error Frequency & Percentage

Type of error	Frequency	Percentage
Reading errors	3	0.61
Comprehension errors	40	8.11
Transformation errors	146	29.61
Process Skills errors	100	20.28
Encoding errors	7	1.42
Careless errors	185	37.53
Learning	4	0.81
Computer Difficulty	8	1.62
TOTAL ERRORS	493	

The table below summarises the distribution of the error categories among the subjects for the paper test.

TABLE 5. PAPER & PENCIL TEST ERROR SUMMARY

Type of error	Minimum Number	Maximum Number	Mean	Standard Deviation
Reading	0	1	0	0.2
Comprehension	0	6	2.1	1.7
Transformation	1	25	6.5	5.2
Process Skills	0	13	4.3	3.9
Encoding	0	1	0.3	0.5
Careless	0	9	4.8	2.6
Test difficulty	0	0	0	0
Total errors per student	5	46	18.0	11.1

The table below summarises the distribution of the error categories among the subjects for the Computer test.

TABLE 6. COMPUTER ADMINISTERED TEST ERROR SUMMARY

Type of Error	Minimum Number	Maximum Number	Mean	Standard Deviation
Reading	0	1	0.1	0.4
comprehension	0	4	1.8	1.2
Transforn.ation	2	19	6.6	3.9
Process Skills	0	11	4.5	2.6
Encoding	0	1	0.3	0.5
Careless	2	21	8.4	4.7
Test difficulty	0	5	0.4	1.1
Total errors per student	11	49	22.4	8.4

The Reading, Comprehension, Transformation, Process Skills, Encoding, Careless and Total Errors in the Computer and Paper test were compared using an independent samples t-test with the "EdStats" Computer Program (Knibb, 1993). The statistic was also employed by McDonald et al (1992) when comparing errors from the Paper and Computer Test. The following table represents the Computer test versus the Paper test.

TABLE 7. ERROR COMPARISON

Type of Error	Mean Difference	Standard Deviation Difference	T	Probability	Significance
Reading	0.1	0.4	1.000	0.329	NS
Comprehension	-0.3	2.2	- 0.585	0.565	NS
Transformation	0.1	-1.3	0.091	0.890	NS
Process Skills	0.2	-1.3	0.250	0.792	NS
Encoding	0.05	-0.02	- 0.295	0.763	NS
Careless	3.6	2.1	2.8	0.010	S
Total	4.4	15.4	1.314	0.204	NS

CHAPTER FIVE

DISCUSSIONS

READING

Heppner, Anderson & Weiderman (1985) reported that reading performance on a standardised test was better when text is displayed in print rather than on a computer display screen. Heppner et al concluded the difference in performance scores is probably due to the reduction in reading speed associated with using a terminal.

In this study the total Reading Errors were very low, with three errors on the Computer Administered Test and only one error on the Paper and Pencil Test. Nearly all the Reading Errors were made by either leaving out or introducing the word 'not'. The only other reading error was made when one person read 40 - 49 as 45. It was reported that at one stage during the test the colour combination on the screen, white text on bluey-green background, prevented a student from reading a question. Assistance was given to this student during the test and hence this difficulty did not result in an error. It was noted through the interviews that one student with poor eyesight had difficulty reading the questions due to the size of the text. This student did not view the final 14 questions, probably as a result of this. However, as the student did not attempt those questions they were not considered as errors and therefore did not influence the results of the study The students

possessed the reading skills required to answer these questions and, with the exception of this one student, it is the author's view that the computer had no effect on the students' ability to read the question.

COMPREHENSION

Most of the Newman studies reported the highest frequency of errors occurred in the Comprehension Category and it is clear that this category is language based (Ellerton & Clarkson, 1992). Clarkson (1991) suggested that language variables are important for distinguishing between those students who make a high proportion of Comprehension Errors and those who make a few.

The studies of Ronau & Battista (1988) and McDonald et al (1992) could not examine this kind of error. Wright & Lickorish (1983) suggested that the computer had a negative effect on the comprehension of text. This was contrary to the findings of Muter, Latremouille, Treurniet and Bean in Wright & Lickorish (1983). However, these two studies did not look at the comprehension of mathematics on the computer screen.

In this study there were 40 Comprehension Errors, 8.11% of the total errors, with a mean score of 1.8 for each student in the Computer Administered Test, compared with a total of 45 Comprehension Errors, 11.9% of the total errors, with a mean score of 2.1 for each student on the Paper and Pencil Test.

The medium of the computer delivering the question does not appear to have an affect on the way a student understands a given

problem. Although it was reported that students enjoyed the presentation of the questions on the computer screen, it did not significantly improve their understanding of the question.

This study reported a low proportion of Comprehension Errors compared to Newman (1977) and Ellerton & Clements (1993) and thus indicates that students found it easier to comprehend the question compared to these populations and tests. However this low percentage of comprehension errors were also present in the studies of Clarkson (1980) and Clements (1980). This result was not surprising because of the age of the sample and seems to support the claim that students in seventh grade and older do not find it as difficult to understand the meaning of the question.

TRANSFORMATION ERRORS

Ronau & Battista (1988) did not have a classification that could be interpreted as a Transformation Error. McDonald's et al (1992) operational error could however be loosely linked to transformation errors. The operational error was defined as use of the wrong mathematical operation to solve the problem. McDonald's comparisons of this type of error were very similar, with the computer mean being 0.42 and the paper mean being 0.43. This mean is reasonably low as the question was presented in the form of a sum such as; 27 + 5 = ? It can be argued that the appropriate mathematical strategy has already been selected for the student.

This study revealed 134 transformation errors, 35.45% of total errors, with the mean for each student being 6.5 on the Paper and Pencil Test, compared with 146 transformation errors, 29.61% of total errors, with the mean for each student being 6.6 in the

Computer Administered Test. These results together with those of McDonald et al seem to indicate that the use of the computer does not effect the carrying out of mental transformation of words of the question to the selection of an appropriate mathematical strategy.

PROCESS SKILL ERRORS

In the study by McDonald et al (1992) the error classifications of Algorithmic Error and Basic Facts Error could be associated with Process Skill Errors. McDonald et al (1992) found the mean number of Algorithmic Errors for the computer test was 4.23 compared with 4.31 for the paper test. The mean number of Basic Facts Errors for the computer test was 0.8 compared to 0.79 for the paper test. These errors accounted for over 50 per cent of the total errors with the comparisons between the modes being very similar.

The sample used for McDonald et al's (1992) study were Year Three students. Due to the age of the sample and the using of a classification based on written work alone it is difficult to compare the results from McDonald et al to this study.

In this study there were 91 process skill errors, 24.07% of total errors, with a mean for each student of 4.3 for the paper and pencil test. There were 100 process skill errors, 20.28% of total errors, with a mean for each student of 4.5 in the Computer Administered Test. The students ability to perform the mathematical operations necessary for the questions appear to be the same for the Paper and Pencil Test as it was for the Computer Administered test.

ENCODING ERRORS

Encoding Errors occur when the student correctly works the solution to the problem but is unable to express this solution in an acceptable form. McDonald et al (1988) examined errors that may be due to the use of the computer. These errors, studied by McDonald et al, could be, by definition, classified as Encoding Errors. An examination of the errors made by students on the Computer Administered Test did not seem to indicate that the types of errors, as mentioned by McDonald et al (1988) were occurring. McDonald et al (1992) reported 3% of the errors were due to transferring the calculations from paper to the computer screen. Although, in this study students were supplied with scrap paper for the computer test, it was intended that the questions would be performed mentally and thus in most cases eliminating the need for transfer of the answer from paper to screen.

This study revealed 7 Encoding Errors, 1.85% of total errors, with a mean score of 0.3 for each student in the Paper and Pencil Test. There were 7 Encoding Errors, 1.42% of total errors, with a mean score of 0.3 for each student in the Computer Administered Test. Only one of these errors was due to the Computer in that the student could not locate the pull down menu in which the answer was to be marked.

There does not seem to be a difference in the students' ability to give the required answer in an acceptable form between the Computer Administered Test and the Paper and Pencil Test.

CARELESS ERRORS

The category of Careless errors has been examined in the implementation of the Newman Error Analysis Interviews (Clarkson, 1980; Clarkson, 1991,1992; Clements, 1982). This type of error does not appear, in some models, investigating students' errors (Knifong & Holton, 1977). Other writers have suggested such errors should be infrequent (Radatz, 1979) or non existent since all errors, in some way, could be classified as systematic (Ginsburg, 1977 in Clarkson 1991).

This study supported Clarkson (1980; 1991) and Clements (1980) in that a large proportion of the total errors were careless, with 26.46% of the total errors in the Paper and Pencil Test and 37.53% of the total errors in the Computer Test.

Casey (1978) replaced the Careless category with another called Unknown Block Errors. Clarkson (1991) also adopts the term Unknown Block Errors in favour of Careless Errors in the belief that teachers often believe students are at fault for carelessness, however this does not capture the underlying definition of these errors. Results from Clarkson (1980), suggest that Careless Errors may be independent of question type, topic area and difficulty. This enhances Casey and Clarkson's claim that the cause of these errors may have a psychological, rather than cognitive, basis. The general ability of students does not seem to interact with the frequency of Unknown Errors (Clarkson, 1992).

The definition the author has adopted for Careless/Unknown Errors is; psychological self-correcting errors that occur in spite of the student possessing the cognitive ability to solve the problem.

Newman (1977) referred to Careless Errors as careless slips, because of nervousness when completing items at the beginning of the test, tiredness when attempting at the end of the test and hesitancy when less familiar items were dealt with.

Casey's Unknown Errors had the characteristic of self-correction. He suggested that temporary mental malfunctions and external factors, such as momentary distractions and slips when copying, may be the base of some Unknown Errors. Casey (1978) and Clarkson (1980) concluded that this type of error may be linked more with non-cognitive variables.

Another group of non-cognitive variables that may have some relation to this type of error, are the perceived reasons students give for their success and failures in academic situations. Students who attributed their success to be derived from ability tended to make fewer Careless /Unknown errors, but those who linked success with effort made more unknown errors (Clarkson, 1992).

As there were significant differences in the Careless Errors it is important to identify non-cognitive variables which may have caused a difference between the two modes of testing. One non-cognitive variable, which may be associated with a significantly higher amount of careless errors in the Computer Test compared with the Paper and Pencil Test, could be the time taken to complete

the test. The test was a maximum of 90 minutes duration. The average time for the completion of the Computer test was 87 minutes. The approximate average time of the Paper and Pencil Test was 70 minutes. Three students left out a total of 28 questions between them, due to the test time expiring.

An independent samples T-test using the Ed-Stats Program (Knibb, 1993) showed there was a significant difference (less than 0.01), between the time taken to complete the Computer Test and the time taken to complete the Paper and Pencil Test. This was supported by McDonald et al (1992) but contrary to the findings of Wise and Plake (1989; 1990) who stated that computer based tests typically take less time to complete that conventional tests.

It was estimated that approximately 7-11 minutes of test time were lost due to screen changes between questions in the Computer Administered Test. This was a major factor in the time differences between the Paper and Pencil Test and the Computer Administered Test.

McDonald, Beal & Ayers (1992) considered the following three factors which appear to contribute to differences in the amount of time to complete computer and paper versions of a test:

1. It is suggested the computer takes longer to read than printed materials (Bugbee & Bernt, 1990). Contributing to this 'slower' reading pace would be the screen format, screen resolution and fonts.

2. The complexity of the keyboard entry. As the majority of keyboard entries in this study involved either clicking in a designated space or two to three digit numbers the writer does not believe this had a significant influence on the time to complete the test.

3. The transfer of questions from paper to the screen. In this study there was not a great deal of transfer needed in the test. It was intended that students would be able to do the calculations mentally and for difficult calculations the computer would supply the student with a calculator that could be quickly called up on the screen.

From observations of students completing the test, it could be seen that students had difficulty in moving objects on the screen. Students took more time to measure distances in the computer test, using these objects, compared with the matched question in the Paper and Pencil Test.

If students rushed their work this may lead to an increase in the frequency of Careless Errors, (Clarkson, 1992). Some students commented in the interviews that they had to rush the final questions in the Computer Administered Test. An examination of the last 10 questions revealed there were 51 errors with 15 (29.4%) of those errors being Careless in the Paper and Pencil Test. There were 83 errors of which 40 (48.2%) were Careless in the Computer Administered Test. Students made a greater percentage of Careless Errors in the last ten questions compared to the whole test in the Computer Administered Test, however this was not the

case in the Paper and Pencil Test. This would appear to support the claim that students were rushed to finish the Computer Administered Test and thus made a greater number of Careless Errors in the final questions. One possible cause for the rush was that the time clock displayed on the screen lacked precision. It was also noted in the memo of the observations from the first Computer Administered Test Session that some students needed a more precise indication of time remaining.

Due to the increase in time taken to complete the Computer Test, the Computer test may have become a "Speeded Test", where as the Paper and Pencil test was not. Criterion-reference tests, as was used in this study, are legitimately claimed as a 'speeded test' if the time limit decreases the examinees' scores (Lin, 1986). Hembree (1986) stated that standardised tests administered under a power condition as opposed to those with time restrictions found that 30 of the 32 comparisons favoured the power condition.

Computer Administered Tests should only consider the time that the question is on the screen (Bugbee & Bernt, 1990). The term 'question' should also encompass maps or checking screens. Eliminating the wait time between screens is a start towards making the test times more equivalent. Future studies comparing Paper and Pencil tests with Computer Administered tests should endeavour to achieve equivalence in test taking time.

Careless Errors made by students during a test may be realised when the answers are checked. It was noted in the interviews that students were reluctant to re-check answers in the computer test.

One student was quoted as saying "In the computer test I couldn't be bothered waiting for each question to click over to check it". It was observed that this person worked through the Paper and Pencil Test at least twice and had no Careless/Unknown Errors in this particular test.

Wright & Lickorish (1983) suggested that there were fewer error detections when the questions were on the computer screen compared with the printed paper. An examination of the questions that students' returned to, during the test, revealed that 249 questions were returned to in the Computer Administered Test. From these questions 91 errors were made, of which 33 were careless (36.26 %). This was close to the percentage of Careless Errors in the entire Computer Administered Test (37.53 %).

This suggests that the psychological reason for the student's initial error was still present when the student returned to the question. This may also agree with Wright & Lickorish (1983), that students are less likely to pick up errors in the Computer Administered Test. Further research is required, comparing the questions checked in the Paper and Pencil test with the Computer Administered Test, before any conclusions can be made.

Although students were able to flag questions, allowing quicker access to view only these questions, for the purpose of redoing/checking, this facility would generally only be used if the student found some difficulty in attempting the question. However, when Careless Errors are made, the student is unlikely to have knowledge of such errors, and therefore would find it unnecessary

to use the flagging facility. Clements (1982) states that the mathematically confident students tended to make a greater proportion of Careless Errors than other students. Students may be too confident during the test and hence do not pay enough attention to the detail of process. However, when required to rework the question in the interview, this, in itself, may be enough to ensure due attention to detail is given (Clarkson, 1992). This suggests it would be unlikely for a student to use the flagging facility for a Careless Error, due to his/her confidence in the answer given.

The Paper and Pencil Test provided easy access to answers for the checking of Careless Errors. To recheck all answers in the Computer Administered Test 71 screens would need to be viewed, compared with just 19 pages in the Paper and Pencil Test, a tedious and lengthy process. Perhaps this problem could be overcome if the Computer Administered Test provided a special checking mode, whereby the student could view approximately 4 questions/answers per screen, click on any Question with an error found, and return to that question in the test mode for redoing.

Other possible reasons for greater number of Careless Errors in the Computer Administered Test compared to the Paper and Pencil Test could be related to computer anxiety or lack of computer experience. Computer anxiety refers to a general negative/aversive attitude about computers (Bernt, Bugbee & Arceo, 1990). The negative attitude may affect how a student answers questions in the test and thus the student may make errors on questions that he/she possesses the cognitive ability to answer. Studies (Bernt et al ,1990; Koohang, 1987) revealed a significant correlation between

the negative attitude towards computer testing and computer experience. In Koohang (1987) female pre-service teachers exhibited a higher degree of computer anxiety than male pre-service teachers. As there was a high proportion of female students in this study (84%) computer anxiety could possibly contribute to an even greater proportion of careless errors.

During the Computer Administered test situations arose that may have contributed to a higher degree of computer anxiety. Such situations included difficulties in moving objects and computer problems, where the computer hung, rendering the computer inactive. The computer hung twice in each of the Computer Administered Testing sessions and the student required professional assistance to rectify the problem. Future research needs to compare the levels of computer anxiety to the number of careless errors made.

It was noted in the memo from the first Computer Administered Testing Session that "The supervisors presence in the room may have been a distraction/intimidation." To ensure that the Computer Administered Test ran smoothly four supervisors were present to assist in any difficulties students may have with the computer. One student mentioned in the interviews that it was intimidating having all the supervisors in the room knowing they could see the answers and any mistakes made. This feeling of intimidation could be another non-cognitive variable that contributed to a significantly higher number of Careless Errors in the Computer Administered test compared to the Paper and Pencil test.

It may be argued that since the Unknown/Careless Errors are by definition an error which can be self corrected when a student has the opportunity to rework the item, there is no need for particular concern in this area. That would be true if indeed students had the opportunity to rework items. When Unknown Errors have been found to consistently account for 25-35% of errors their importance cannot be underrated (Clarkson, 1992).

It cannot be ruled out that with errors classified as Careless, incidental learning may have occurred between the tests and the interview, due to the repeated measures of the design, although attempts during the interviews tried to identify whether any learning may have occurred. An examination of Careless Errors showed that of the 185 Careless Errors in the Computer Administered Test, 26 were also errors when these students attempted the Paper and Pencil test. Of 95 Careless errors examined in the Paper and Pencil Test, 36 were also errors when these students sat the Computer Administered Test.

Future research should conduct the interviews between the two tests. This was, however, not possible in this study due to the time limitations and the influence the interviews would have had on other research being conducted on the Computer Administered Testing System.

TEST DIFFICULTY

All the errors in this category occurred due to the student using the computer. One person had five errors due to difficulty in moving an object in the first question. This person, after encountering difficulty in the first question, did not possess the motivation to attempt other similar questions requiring the movement of objects.

Other students who made this type of error also indicated frustration with moving objects. These types of errors could also be linked to Newman's Motivation category as the student did not persist at trying to move objects after some initial difficulty.

LIMITATIONS

It should not be imagined that the pattern of errors made by one set of students on a series of mathematical tasks will be similar to the pattern made by the same pupils on a different series of mathematical tasks, or even, to the pattern made by a different set of pupils on the same tasks (Newman, 1977). For this reason any generalisations would be limited to First Year Bachelor of Arts Primary Education students and to this particular Mathematical Competency Test.

The high proportion of female students in this sample (84%) would also limit the generalisations of the study.

In the Computer Administered Test students were issued with rulers and calculators only for specific questions, where the examiner wished to have them present. In the Paper and Pencil

Test students had access to a calculator and ruler for all questions. Thirteen questions were identified as being possibly made easier to answer by the presence of calculators and rulers in the Paper and Pencil Test, where they were not present in the Computer Administered Test.

If the calculator and ruler had an impact on the questions it should result in less Process Skill Errors as all steps up to this level would remain the same with or without these instruments. In the Computer Administered Test there was a total of 77 errors, on these 13 questions, of which 21 (27%) were Process Skill Errors. In the Paper and Pencil Test there were 52 errors, on these 13 questions, of which 15 (28%) were Process Skill Errors. The differences of Process Skill Errors, between the two modes of testing, for these 13 questions, ranged from -1 to 3.

The presence of rulers and calculators could have a psychological effect on the students' performance on these questions and ,hence, there may be differences in the number of Careless Errors. It was noted that 42 (55%) Careless Errors were made on these 13 questions, in the Computer Administered Test, with only 21 (40%) Careless Errors made on these 13 question, in the Paper and Pencil Test.

The differences between the administration procedures of the two testing modes were exemplified by one question in which the students were required to estimate distance within a limited time frame. For the matching question in the Pencil and Paper Test students had access to a ruler and no time constraints. For this

question there were five Process Skill Errors and six Careless Errors, in the Computer Administered Test, compared with three Process Skill Errors and one Careless Error in the Paper and Pencil Test.

CHAPTER SIX

CONCLUSIONS

Through the use of interviews it was concluded that the implementation of the Computer Administered Test did not appear to significantly affect the students' ability to read the question, understand what the question is asking them to do, transform the words of the question to an appropriate mathematical strategy, perform the mathematical operations necessary and express the solution in an acceptable form, in comparison to the equivalent Paper and Pencil Test.

There was, however, a significant difference in the number of Careless Errors made by the students. It was reasoned that non-cognitive variables contributed to the Careless Errors. One of the major non-cognitive variables identified as contributing to the difference of Careless Errors between the two testing modes was the difference in time taken to complete the tests. It appears that some students may have rushed their answers toward the end of the Computer Administered Test. This was realised by the greater proportion of Careless Errors made on the last ten questions in the Computer Administered test.

Another possible non-cognitive variable associated with a greater proportion of Careless Errors on the Computer Administered Test

was Computer Anxiety. This could be related to lack of computer experience, and difficulties with the computer during the test.

Test flexibility also needs to be considered in relation to Careless Unknown Errors. Studies where students do not have the flexibility to change answers may produce significant differences in the scores due to an inability to check possible Careless Errors. All research comparing Computer Administered Testing to Paper and Pencil Testing needs to have equivalent test administration procedures.

The test was designed to assess students' mathematical competency. As Careless/Unknown Errors accounted for 26.46% of total errors in the Paper and Pencil test and 37.53% in the Computer Administered test in this study, they may, for some students, be the difference between passing and failing. Research needs to explore ways to eradicate Careless Errors. There may be a need to develop different test strategies for the Computer Administered Test to reduce these errors.

As an advantage over the Paper and Pencil Test, the Computer Administered Test provides more standardised test administration, immediate test scoring and feedback, and the ability to collect item latency information. It is the author's belief that the amount of Careless Errors made would be similar if the differences in time needed to complete the test and situations which could lead to an increase in Computer Anxiety were eliminated. If this could be achieved the Computer Administered Test would be an efficient new strategy for the testing of mathematical competency.

REFERENCES

Anderson, R. I. & Trollip, S. P. (1982) A Computer Based Private Pilot (Airplane) Certification Examination: A First Step Towards Nation-wide Computer Administration of FAA Certification Exams. *Journal of Computer-Based Instruction, 8* (3), 65-70

Bainbridge, R. (1981) To err is human: Towards a more positive approach to young children's mistakes in arithmetic. *Mathematics in Schools, November*, 10-13

Bernt, F. M., Bugbee, A. C. & Arceo, R. D. (1990). Factors influencing student resistance to computer administered testing. *Journal of Research on Computing in Education, 22* (3), 265-2

Birenbaum, M. & Tatsuoka, K. K. (1987a). Effects of 'on line' test feedback on the seriousness of subsequent errors. *Journal of Educational Measurement, 24* (2), 145-155

Birenbaum, M. & Tatsuoka, K. K. (1987b) Open-ended versus multiple-choice response formats - it does make a difference for diagnostic purposes. *Applied Psychological Measurement, 11*, (4), 385-395

Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. *Journal of Educational Measurement, 29,* (3), 253-271

Bugbee, A. C. & Bernt, F. M. (1990) Testing by Computer: Findings in Six Years of Use 1982 - 1988. *Journal of Research on Computing in Education , 23* (1), 87-100

Casey, D. (1978). Failing Students: A strategy of error analysis. In P Costello (Ed), *Aspects of Motivation* (p 295-306). Melbourne. Mathematical Association of Victoria.

Clarkson, P. C. (1980). The Newman Error analysis - Some extensions. In B. A.. Foster (Ed) *Research in mathematics education in Australia* 1980 Vol 1 p11-22

Clarkson, P. C. (1983). Types of errors made by Papau New Guinean students. *Educational Studies in Mathematics, 14* 355-367

Clarkson, P. C. (1986). On Consistency of Classifying Errors. *Research in Mathematics Education in Australia.* December 1986, 22-24

Clarkson, P. C. (1991). Language Comprehension Errors: A further investigation. *Mathematics Education Research Journal, 3* (2) 24-33

Clarkson, P. C. (1992). Unknown/Careless errors: Some Implications for traditional test procedures. *Focus on Learning Problems in Mathematics, 14* (4), 3-16

Clements, M. A. (1980) Analysing children's errors on written mathematical tasks. *Educational Studies in Mathematics, 11* (1), 1-21

Clements, M. A. (1982). Careless errors made by sixth-grade children on written mathematical tasks. *Journal for Research in Mathematical Education, 13* (2), 136-144

Cox, L. S. (1975a). Diagnosing and remediating systematic errors in addition and subtraction computations. *The Arithmetic Teacher, 22*, 151-157

Cox, L. S. (1975b). Systematic errors in the four vertical algorithms in normal and handicapped populations. *Journal for Research in Mathematics Education, 6* , 202-220

Dickson, L. ,Brown, M. & Gibson, O. (1984). *Children learning mathematics: A teachers guide to recent research.* Oxford: Schools Council.

Eaves, R. C. & Smith, E. (1986) The Effect of Media and Amount of Microcomputer Experience on Examination Scores. *Journal of Experimental Education, 55,* 23-26

Ellerton, N. F. & Clarkson, P. C. (1992). Language factors in mathematics learning. In B. Atweh & J. Watson (Eds), *Research in Mathematics Education in Australasia 1988 - 1991*. Kelvin Grove Queensland: Mathematics Education Research Group of Australia

Ellerton, N. F. & Clements M. A. (1991). *Mathematics in Language: A Review of Language Factors in Mathematics*. Geelong: Deakin University.

Ellerton, N. F. & Clements M. A. (1993) Implications of Newman Research for the issue of "What is basic in school mathematics?" Unpublished paper Deakin University

Englehardt, J. M. (1977). Analysis of children's computational errors: A qualitative approach. *British Journal of Educational Psychology, 47,* 149-154

Gagne, R. M. (1971). *The Conditions of Learning* second edition. New York: Holt Rinehart and Winston.

Geiger, V. S. (1990). Serendipity and Students Errors. *Teaching Mathematics, 15* (3), 7-10

Hembree, R. (1986). Effeects of non-content variables on mathematics test performance. *Journal for Research in Mathematics Education, 18* (3), 197-214

Heppner, F. H., Anderson, J. G. T., Farstrup, A. E. & Weiderman, N. H. (1985). Reading from a standardised text is better from print than from computer display. *Journal of Reading, 28* (4), 321-325

Hollander, S. K. (1978). A literature review: Thought processes employed in the solution of verbal arithmetic problems. *School Science & Mathematics, 78*, 327-335

Knibb, K. (1993). EdStats (computer package), unpublished, Edith Cowan University: Perth

Knifong, J. D. & Holton, B. D. (1977). A search for reading difficulties among erred word problems. *Journal for Research in Mathematics Education, 8* (3), 227-230

Koohang, A. A. (1987). A Study of the Attitudes of Pre-service Teachers Toward the use of computers: *Educational Communications and Technology Journal, 35* (3) , 145-149

Lankford, F. G. (1974). What can a teacher learn about pupils thinking through oral interviews? *Arithmetic Teacher, 21*, 26-32

Lee, J. A. (1986). The Effects of past computer Experience on Computerised Aptitude Test Performance: *Educational and Psychological Measurement, 46* (3) 727-733

Lee, J. A., Moreno, K. E. & Sympson, J. B. (1986). The effects of mode of test administration on test performance. *Educational and Psychological Measurement, 46* (2), 467-474

Liefeld, J. P. & Herrman, T. F. (1990) Learning consequences for university Students using Computerised Mastery Testing. *Educational - Technology Research and Development, 38* (2), 19-25

Lin, M. H. (1986). *The impact of time limits on test behaviours.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Marinas, B. & Clements, M. A. (1990). Understanding the problem: A Prerequisite to problem solving in mathematics. *Journal of Science and Mathematics Education, 13* (1), 14-20

McDonald, J., Beal, J. & Ayers, F. (1988). Computer Administered Testing: Diagnosis of addition computational skills in children. *Journal of Computers in Mathematics and Science Teaching, 7* (1-2), 38-43

McDonald, J., Beal, J. & Ayers, F. (1992). Details of performance on computer and paper administered versions of a test of whole number computational skills. *Focus on learning Problems in Mathematics, 14* (3), 15- 27

Moe, K. C. & Johnson, M. F. (1988) Participant Reactions to Computerised Testing: *Journal of Educational Computing Research, 4* (1) p 79-86

Newman, M. A. (1977) An analysis of sixth grade pupils on written mathematical tasks. In M. A.. Clements & J. Foyster (Eds) *Research in Mathematics in Australia* , Vol 1, 239-258

Olsen, J. B., Maynes, D. D. ,Slawson, D. & Ho, K. (1989) Comparisons of Paper-Administered, Computer - Administered and Computerised Adaptive Achievement Tests: *Journal of Educational Computing Research, 5* (3) 311-326

Orey, M. A. and Burton, J. K. (1992). The trouble with error patterns. *Journal of Research on Computing in Education, 25* (1), 1-14

Perso, T. (1992) Making the most of errors. *The Australian Mathematics Teacher, 48* (2), 12-14

Pinchback, C. L. (1991). Types of errors exhibited in a remedial mathematics course. *Focus on Learning Problems in Mathematics, 13* (2), 53- 62

Polya, G. (1973). *How to solve it: A new aspect of mathematical method.* Princeton, NJ: Princeton University Press.

Radatz, H. (1979) Error analysis in mathematics education. *Journal for Research in Mathematics Education,* 10, 163-172

Ronau, R. N. & Battista, M. A. (1988). Microcomputer versus paper and pencil testing of students errors in ratio and proportion. *Journal of Computers in Mathematics and Science Teaching, 7* (3), 33-38

Ring, G. (1993) Research and Development into Computer - Administered Testing in Multimedia /Hypermedia Environment. Unpublished Conference Paper: Edith Cowan University Perth

Roberts, G. H. (1968). The failure of strategies of third grade arithmetic pupils. *The Arithmetic Teacher 15*, 442-446

Siemon, D. E. (1988) Errors and the learning process. In D Firth (Ed) *Maths Counts - Who Cares?* (p 128-133) Melbourne. Mathematics Association of Victoria.

Spray, J. A. ,Ackerman, T. A. ,Reckase, M. D. & Carlson, J. E. (1989) Effect of the medium of item presentation on Examinee Performance and Item Characteristics. *Journal of Educational Measurement, 26* (3) 261-271

Travis, B. (1991). Using computers in collegiate remedial mathematics. *Collegiate Microcomputer, 9* (1), 43-47

Watson, I. (1980) Investigating Errors of Beginning Mathematics. *Educational Studies in Mathematics, 11* (3) , 319-329

Wayne Plumley, J. & Ray, N. H. (1989) Computer Administered testing in a Classroom Setting. An Alternative. *Journal of Research on Computing in Education,* 22 (1), 261-271

Webb, N. L. (1992). Assessment of students Knowledge of Mathematics: Steps towards a theory: In D. A. Grouws (Ed) *Handbook of Research on Mathematics Teaching and Learning.* New York. MacMillan publishing company

Wise, S. L. & Plake, B. S. (1989). Research on the effects of administering tests via computers. *Educational Measurement and Evaluation: Issues and Practices,* 8 (3), 5-10

Wise, S. L. & Plake, B. S. (1990). Computer-based testing in higher education. *Measurement and Evaluation in Counselling Development,* 23, (1), 3-9

Wright, P. & Lickorish, A. (1983). Proof reading text on screen and paper. *Behaviour and Information Technology,* 2 (3), 227-235

APPENDIX 1

Memo

EDITH COWAN UNIVERSITY
PERTH WESTERN AUSTRALIA
MOUNT LAWLEY CAMPUS

To : Geoff, Jack, Don, Mike, Gary
From : Rod Ellis
About : Observations from the first CAT session (1P4, Wednesday 25/8/93)
Date : 25/8/93

The following points were noted or actions taken during the testing period. The order of points made below is of no consequence.

1. One student had not done the awareness session with Geoff. She chose to do the test and Carl Delides assisted her for about 5 minutes prior to her commencing the preview.

2. One computer temporarily 'hung' twice during the test. It was reinstated with Chas Bray's assistance. He used 'Command >' to solve what he called a memory problem. The computer was No.2 on the right side (as you face the back of the room).

3. Question 38 has the same distracter for alternatives (a) and (e)

4. For question 47 there was no square root function available on the calculator.

5. Because sound was turned off the 'Cycling Man' stopped without a sound signal to indicate time was up.

6. Gary was asked to record finishing times for all pen & paper candidates.

7. Some graphics may well distract rather than help (my observation only)

8. The Time Clock lacked precision and some students needed more precise indication of time remaining.

9. Jack forgot to issue rulers (but we remembered in time!) Sorry Jack!

10. Size of text could be bigger. It was certainly a problem for one student! This one student was quite frustrated and was particularly slow (completed about 50 questions only).

11. Some colour combinations could be changed. eg. white text on bluey-green! One student could not read this at all.

12. One student used the scrap paper to assist with the estimation in question 64.

13. One student only used the preview for about 30 seconds then started the test.

14. The supervisors presence in the room may have been a distraction/intimidation. We will need to be less conspicuous at future sessions.

15. As far as possible we separated the candidates in the room but proximity to neighbouring screens is an issue that in future should be controlled better.

16. One student could not enter an answer to question 36. We had her answer the question on paper (see attachment). She scored 76% anyway so she has effectively passed the test.

17. As far as we (the supervisors) could tell, nobody chose to use the map until very late in the test and probably not before they had seen or tried the 71 questions.

18. The student who did not do the awareness session started after the others and completed the test first. She also scored 80%. *Carl Delides is taking all the credit for this!*

19. Four or five people had some trouble with the slow moving rulers & grids.

20. The pen & paper test candidates finished sooner (noticeably!) than the CAT people. We have access to all finishing times to compare this more objectively.

21. There were 12 CAT candidates and 10 pen & paper test candidates. We will need to adjust the next group split to make sure we have balanced totals for analysis.

22. Mike has the completed questionnaires and Gary has the completed Written test papers.

23. Mike, Chas Bray and Carl stayed throughout to supervise. Leonie reviewed the test as a pseudo student under test conditions.

24. Almost nobody chose to use Help or make a comment using the Comments option.

Overall the test went smoothly and student questions during test time were infrequent.

APPENDIX 2

Memo

EDITH COWAN UNIVERSITY
PERTH WESTERN AUSTRALIA
MOUNT LAWLEY CAMPUS

To : Geoff, Jack, Don, Gary, Mike
From : Rod Ellis
Subject : Notes about CAT session 27/8/93 (1P5 9-11am)
Date : 30/8/93

The following points are things noted, or actions taken, during the CAT session with 1P5 on Friday 27/8/93.

- Ther were 11 CAT candidates. All of these had participated in the 'awareness session'.
- There were two left handed students but both were happy to use the mouse as set up for the machine that they were using.
- Two machines 'hung' at different times during the CAT when a move object question was being attempted. Both were successfully reinstated with <Command/Period>
- The move object questions bought the most queries. The time delay before an object begins to move is proving to be an issue of concern.
- It is worth noting that CAT candidates who find they are running out of time cannot quickly guess the answers to any remaining unanswered questions. *Mike's observation!*
- One student could not do the 3 estimation questions (19/25/68) because she must have tried them in the preview. Her answers were taken on paper and given to Gary who will mark both the CAT test & the P & P test..
- P & P candidates had a square root key on their calculator but the CAT candidates did not.
- In the P & P test, one student did not understand question 16. Perhaps need to include 'tile' as well as 'tessellate'.
- Having calculators and rulers available in the P & P test meant students could calculate or measure when required to estimate.
- One student in the computer test used her own ruler (but not where pull down rulers were available) to assist with some questions. For example, this was done in response to the number line question.
- One student in the CAT test had forgotten how to use the green arrows to go from one question to the next. This was quickly solved with intervention from a supervisor.
- Once again, in general, the P & P test candidates finished sooner than the CAT candidates.
- Once again the CAT candidates seemed to only use the 'map' option after working their way through the 71 questions.

APPENDIX 3

EDITH COWAN UNIVERSITY
Declaration of Willingness to Participate in a Research Project.

I ..... agree to participate in the following:

1 Familiarisation session in preparation for the computer administered test of approximately 45 minutes.

2 An attitude scale on tests, computers and mathematics of approximately 45 minutes.

3 A computer based test in MPE 0101 which will take approximately 90 minutes.

4 A paper and pencil test in MPE 0101 which will take approximately 90 minutes.

5 A one to one interview of about thirty minutes regarding errors made in the above tests.

I have read the above information and any question I have asked has been answered to my satisfaction.

I am aware that the best mark from these two tests will be used for my grade in MPE 0101 and all other data will be used for research purposes only. I am also aware that I may take the MPE 0101 test at the end of semester two 1993 if necessary. I agree to participate in this project realising that I may withdraw at any time. I agree that the research data gathered for this study may be published provided the participant is not identifiable.

Signature of participant .....

Research Project Director: Dr Geoff Ring

Directors Signature ..... Date .....