THE UNIVERSITY OF ALBERTA

THE EFFECTS OF OPEN- AND CLOSED- BOOK EXAMINATIONS IN MATHEMATICS - AN EXPERIMENTAL STUDY

by

© SHIRLEYANNE MICHAELS

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA

FALL, 1972

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled, "The Effects of Open- and Closed- Book Examinations in Mathematics - An Experimental Study." submitted by Shirleyanne Michaels, in partial fulfillment of the requirements for the degree of Master of Education.

Thomas E. Kieren
Supervisor

A. T. Olson

J. O. Maguire

Date Nov. 7, 1972

ACKNOWLEDGEMENT

The writer gratefully acknowledges the assistance she received in the planning and execution of this study. First of all, she wishes to thank her adviser, Dr. T. E. Kieren, not only for his skillful direction, but also for his patience and understanding. She also thanks her other committee memebers, Dr. A. T. Olson and Dr. T. O. Maguire, for their constructive criticism and advice.

The writer acknowledges tha assistance of the Department of Education, Edmonton, Alberta in the preparation, administration and compliation of the results of the study.

Thanks is given to the many individuals, particularly Mrs. Edith Remin and Mrs. Margo Cosontos, that assisted in the typing of the manuscript.

Finally, the writer expresses her appreciation to her husband, Geoff, for the support and interest he showed in the thesis.

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF TABLES (CONTINUED)

CHAPTER 1: THE PROBLEM

BACKGROUND FOR THE PROBLEM

Evaluation programs today must use optimum testing procedures and instruments if they hope to be successful. The development of the proper measurement instruments and selection of appropriate settings for many evaluation processes is still lacking. As a result, the relationships between relevant variables are often vague and uncertain.

This study will concentrate on the effects of different test settings. In the past the inappropriate use of or the poor quality of measurement instruments led to the situation where certain parts of the evaluation process were ignored or forgotten and other parts were poorly completed. One of the inappropriate uses was not considering the test setting when measuring a particular skill or concept.

This study has as its focus, group achievement processes, although it recognizes the need and importance of individual measurement techniques as well. It is concerned with group examinations written in two settings- open-book and closed-book with the emphasis on examinations written in an open-book setting. The settings - open-book and closed-book are defined within the context of Departmental Examinations regulations in the Province of Alberta (1966). It is hoped the following discussion will show for which situations the open-book setting is best and where it is not suitable.

STATEMENT OF THE PROBLEM

The following questions define the scope of this study. Answers to these questions will help to define the measurement instruments and the way they are used in evaluation procedures further.

1. Do examinations written in an open-book setting, as contrasted with those written in a closed-book setting, provide a different assessment of the student's abilities, skill and attitudes in a particular field or study? Does his achievement differ between the two settings?

2. What are the implications of a student's attitude to the subject area being tested, his level of anxiety while being tested and his feelings toward the testing process itself?

3. What relationship exists between the total test statistics of variance, validity, and reliability when the two test settings are compared?

THE PURPOSE OF THE STUDY

The purpose of this study is to examine the measurement process of administering group examinations in an open-book setting and to compare this information with similar material for administering group examinations in a closed-book setting.

Taking a close look at examination development procedures illustrates why this type of investigation is important. An examiner confronted with the task of constructing an examination must deal with two problems successfully. First, before he can construct any type of examination he must have specified the behavioral objectives he hopes to measure. Secondly he must find a valid way of measuring them.

The first problem, deriving behavioral objectives has been discussed in detail by others, such as Taylor and Cowley (1972). They list the steps necessary to identify the new patterns of behavior to be acquired

by the student. After the examiner has acquired a list of objectives, he must interpret them correctly and devise means of testing that they have been achieved.

This second job of validly testing the behavioral objectives is important. Only as a result of an accurate measurement can one determine if the objectives are being reached. The examiner must choose the significant concepts, principles and operations involved in the objective being evaluated and then prepare a valid means of measuring them (Sueltz, 1961).

With the problem of measurement the examiner is faced with two additional problems. First, he must decide the type of question to use, for example, an essay question or a multiple-choice item. To perform this task he has much research and related literature to use in reaching his decision. If he wishes to determine the degree of synthesis the student has attained in a particular area, he chooses the essay question. If he wishes to test comprehension or application, he likely chooses the multiple-choice item. This permits him to test his objective in the most efficient manner. The second problem facing the examiner is not so easily resolved. Very little research has been done on the basic problem of evaluating the type of examination best fitting the needs of the course. It has been generally assumed that one type of examination--the traditional closed-book--is best for all situations. However the desired outcomes of all courses might not profit from being measured by a closed-book examination regardless of the type of questions that are found in it. A course oriented to using outside sources of information

such as library or community likely cannot be adequately measured by a closed-book examination where the student is expected to rely totally on his memory to answer the examination. Thus, the examiner must consider the type of test he will use as well as the type of question.

Specifying the setting of examinations has been largely neglected in literature. Yet its importance in establishing valid measurement techniques should not be under-estimated. Some learning processes may best be evaluated by a take-home project or by an open-book examination (use of textbook, notes and references) instead of a traditional closed-book examination.

To examine the two types of examination settings the following research questions will be investigated.

1. What effect does the setting for an examination have on the student's performance on that examination?
2. What is the relationship between student anxiety scores and the examination setting?
3. What is the relationship between student attitude, the setting of the examination and the examination achievement?
4. Which examination setting do students favor?
5. Do students who have a favorable attitude toward the subject of mathematics like examinations set in an open-book setting better than those students who have an unfavorable attitude toward mathematics?
6. Comparing the open and closed-book setting, what is the relationship between total examination statistics such as variance, reliability and validity?
7. Do students respond differently to individual items written under a closed-book setting versus an open-book setting?

THE NEED FOR THE STUDY

The need for general investigation of different measurement processes has already been indicated. This section deals with the specific need of investigating the group examination written in an open-and closed-book setting. The first part of this section describes some of the history of open-book examination to provide background information on the need to examine group examinations tested in an open-book setting more fully.

The group examination written in an open-book setting has been used sporadically over the years in North America. A general definition of an open-book setting is one in which textbooks, notes and additional references may be used by the student as he writes the examination.

Stalnaker and Stalnaker (1935) described a three hour open-book comprehensive examination at the University of Chicago. Thus, the use of examinations written in an open-book setting is not new, but it is relatively unexplored.

However, in the Province of Alberta this is both a new and unexplored experience. The first open-book examination was conducted at the provincial level in 1967. The two hour open-book examination was given in Chemistry 30X, a laboratory oriented course. This examination differed from the traditional examination written in a closed-book setting where the student is required to answer the question from memory and not use any external aids.

In 1971, for the first time, the courses Chemistry 30, Physics 30, Physics 30X and Biology 30 were also assigned examinations to be written in an open-book setting. The rationale behind this action was that open-book examinations permit more flexibility in course design and allow testing of more important course principles. They also encourage the student to make a sounder preparation for the test and provide a testing situation more closely related to the classroom situation.

Using open-book set examinations as measurement devices, raises some important questions. Since at present the examinations are an important part of the evaluation scheme in the province of Alberta, it becomes imperative that this trend to open-book examinations and away from closed-book examinations be based on a sound foundation. As indicated earlier, at present the use of examinations written in an open-book setting has not been extensively investigated. The factual studies completed deal with limited university populations writing open-book examinations in different physical situations and for different purposes than the grade twelve students in the Province of Alberta. An example is the general article on the advantages of examinations written in an open-book setting by Tussing (1951) in which he presents a logical argument for his point of view. He describes some testing situations, states his preference for open-book set examination but does not present empirical research to support his feelings.

Descriptive studies have been done on the effects of types of examinations on students. Furst (1958) states:

> "What students emphasize in their studying undoubtedly depends more upon what they expect in examinations that upon any formal statement of course aims (p 6)."

He further feels that tests consisting of knowledge questions encourage students to memorize isolated facts (cram) in order to pass the examination. On the other hand, examinations of the open-book type encourage students to prepare by arranging the information they had gained so they are able to apply the methods and principles of the course to new situations and problems. However, very little actual research has been done on such problems.

In this era of critical questioning of evaluation in general, it seems that the values and concepts involved in constructing and administering group examinations in an open-book setting should be properly researched. If properly channelled, the use of these group examinations can be multi-purpose. For example, forms of placement examinations for adult students may well be more accurate and yet less stressful for the person writing if written in an open-book setting. Thus there is a great need for this type of research to add to, substantiate and further the investigation of group measurement processes.

DELIMITATIONS

Although the study was designed to cover the field of open-book examining versus closed-book examining, it was found necessary to delimit the study as follows:

1. The sample used for the study was six hundred Mathematics 30 students selected from different areas in the province. To make the most use of the results of the study, it is necessary to generalize from this sample to the Mathematics 30 population and further to all Grade XII courses using examinations written in an open- or closed-book setting. This extensive generalization requires care in interpreting the experimental results in terms of other situations.

2. No attempt was made to control home, school, or community factors that might have affected the study.

3. The examination used was of the form of a standard closed-book Mathematics 30 examination.

CHAPTER 2: REVIEW OF RELATED LITERATURE

INTRODUCTION

In this chapter the literature that has dealt with the problem of examination setting in the past is reviewed. The development of the history of examinations in the province of Alberta provides a framework against which the concepts of writing examinations in an open-book setting can be evaluated. The first section traces some of the changes in form and content through which examinations have passed over the last hundred years. The next section is concerned with the history of examinations written in an open-book setting. Areas in need of further investigation in administering examinations in such a setting are discussed. Similarly, as with examinations in general, these examinations have varied over the years. The term "open-book setting" has meant anything from writing an examination with the aid of a "crib" sheet or slide rule to using a complete set of notes, texts and references. In this study the latter definition will be used.

The relationship of anxiety, attitude, and achievement to test setting are described in this study. Background information from previous studies is reported in this chapter.

HISTORY OF ALBERTA'S DEPARTMENTAL EXAMINATIONS

Evaluation has been an important part of western civilization since its advent. Over the years its form and content have

changed but one major purpose has remained the same, to provide an indication of the success or failure of an individual in a particular course of action. Ebel (1968) sums up the history of the evaluative process as he states:

> "It is safe to predict that changes will come in evaluation as in other aspects of education... A new idea today must be very good indeed to be better than the hosts of good ideas that have preceded it ... But changes do come. The changing social scene brings changes in educational emphasis. (and thus changes in the form and emphasis of evaluation.) (p 33)."

Thus the form of evaluation is a function of the time in which it exists and changes to fit that time.

To understand the place of open-book examinations in the evaluative scheme of the province of Alberta today a short review is given of the history of Alberta's departmental examinations.

External examinations were in effect in Alberta even before the province came into existence in 1905. Since then a consistent policy has been in effect to conduct province-wide departmental examinations at the end of Grade IX and Grade XII. At certain times these examinations have been administered at other grade levels as well. A chronological description of Alberta's examination history is given by Chalmers (1968).

> "In Alberta the departmental examination is older than the province... It applies to standard V to VII, corresponding approximately to Grades VIII to XIII, that is, high school entrance to graduation. Since 1912 when the "grade" system replaced the older "standard"

> classification, provincial examinations have been administered at different times upon completion of every grade from VIII to XII, although for some 30 years they have been limited to Grade IX and high school graduation level, Grade XII (p 91-92)."

Initially the papers were set by scholars who were masters of their disciplines but who had never taught in the high school classroom, e.g., university professors. This often led to the construction of tests that did not measure the objectives of the course. Teachers petitioned for representation on examining bodies to try to prevent such occurrences. At first one teacher representative was allowed on the policy board responsible for deciding the examination content. Later the actual responsibility for examination construction was given over to practising teachers. The number of teachers involved with the construction of an examination grew from a single individual to a committee of three to six teachers today. The time involved changed from one or two weeks of effort by one person to a year long process involving many weeks and individuals.

Originally the examinations were open-ended essay-type papers. By the 1930's, however, according to MacArthur and Hunka (1960) papers contained some questions which could be answered quite briefly by the student and which could be scored objectively by the examiner. This trend continued until 1958 when the departmental examinations were mostly objective-scored questions.

Each of the two types of testing mentioned above have their merits and disadvantages. MacArthur and Hunka (1960, p 387, pp 40-43)

discuss some of these in terms of validity, reliability and practicality of the total examination. Suggestions are also given for improving the forms and quality. Alberta decided to use objective question types in evaluation. Since this decision was made, efforts have been constantly made to up-grade the objective items being produced.

In 1964 a significant change occurred with the initiation of multiple-choice items as a major part of many examinations papers. By 1969 all examinations were multiple-choice papers.

Items on the papers were categorized according to thought level and subject area. Thought level categorization was carried out according to Bloom's taxonomy and recently in the field of mathematics and science according to Avital and Shettleworth's modification of Bloom's taxonomy. (Bloom et al, 1956) Several subject taxonomies have been written over the past years illustrating the relationship between behavioral objectives and questions measuring these objectives.

The modification of Bloom's taxonomy, used in this study, was based upon the hierarchy by Avital and Shettleworth (1968). The hierarchy can be summarized in the following form.

1.00 Knowledge

To answer items at this level the student needs only to recognize or remember materials learned directly from text books or through classroom instruction.

2.00 Comprehension

At this level the student must make a simple transfer or generalization using well comprehended knowledge.

3.00 Application

At this level the student must solve a problem of transfer dealing with an unfamiliar situation and the solution is generally a multi - step procedure.

4.00 Analysis

At this level the student does not have available a set of procedures or method of solution. He must be able to examine the material and derive his own relationships to solve the problem.

5.00 Synthesis

At this level the student must be able to put together given elements in an entirely new way to find the solution.

This classification scheme was used to differentiate between the different thought levels the students were asked to exhibit while writing their examinations. Classification is used in this study as one factor in the determination of validity.

The function of Grade IX examinations changed in 1970. Their coverage moved from being exclusively Grade IX to the entire junior high school program. Also instead of being used as a pass-fail yardstick, the function became one of guidance. With the change in purpose of the Grade IX Examinations (now Junior High Achievement Battery)

only the Grade XII examinations were left. As mentioned earlier a committee of three to six teachers is appointed by the High School and University Matriculation Examinations Board. These examiners construct items for the examination emphasizing the objectives laid down in the respective curriculum guides. The items are pre-tested on a representative sample of Grade XII students. The results from this pre-testing are then analyzed at the Operational Research Branch of the Department of Education. By studying this item-analysis, examiners are able to choose the best items to include in their examinations.

After the examiners have constructed the actual examination, it is given to a second committee - the revisors. The revision committee consists of experts in the subject being tested. Generally each committee is composed of a university professor who has specialized in the subject, a school inspector or superintendent who has also specialized in the subject and two or three teachers who are well qualified in and are presently teaching the subject. The revisors read and check the paper for errors and apparent weaknesses. Questionable items are eliminated or modified.

Finally the paper is sent in its final form to the printers. The galley-proof is proof read by a member of the revision committee and a Department of Education representative and then the paper is printed.

The preparation of examinations is a complex process. As indi-

cated above, the construction process is carefully carried out to ensure the best possible product results. Since the uses of these examinations are many and varied they deserve continuing careful attention. Thus the examination of test setting will aid in developing the best testing situation possible.

In the previous paragraphs a brief sketch of the history and background of examination in the province of Alberta has been given. The case that has been developed for and against the open-book setting for examinations is given in the next section.

HISTORY OF OPEN-BOOK EXAMINATIONS

One of the oldest references to open-book examinations, given by Stalnaker and Stalnaker (1934) was reported earlier. They further state:

> "Examinations to which students are allowed to bring some outside aids are very old. Individual instructors have used them occasionally for many years. Engineers, for example, expect to use slide rules in an examination ... Instructors in other fields have from time to time permitted students to bring whatever notes or books they wished to an examination ... Other instructors have announced the examination questions in advance ... (p.214)."

This was the first time an open-book examination was officially recognized at the university level according to the authors. In the spring of 1934 students at the University of Chicago wrote a three hour open-book examination covering history, religion and science in the morning. In the afternoon they wrote a three hour traditional

closed-book examination on literature, philosophy and art. The complete examination was largely objective, although a one-hour essay questions was included in the open-book and short-answer material in the afternoon.

No systematic survey of student opinions was conducted. It appeared to the authors that students were using their books in the open-book section. Comparing the results of the two sections gave a correlation of .84. In the previous year's set of examinations the correlation between the two examinations was .88. That year both sections of the examination were closed-book. The small difference was not significant. No effort was made to classify the types of questions on each examination beyond noting that the memory exercise questions were found on the afternoon examination.

Cowley (1934) commenting on this open-book examination which involved 500 students stated:

> "This type of examination has been used occasionally by individual instructors in many institutions but never before has it been officially recognized by a university as an acceptable method of testing the knowledge of a large number of students. The program is frankly experimental and constitutes an attempt to measure ability rather than rote memory. (The philosophy behind the examination is that) the student who thoroughly understands the subject is not penalized because he forgets a simple detail and that the student who does not have thorough understanding of the subject-matter cannot pass by hasty perusal of his texts and notes (p 399)."

Thus the results of their study indicated that open-book examinations gave the student no added advantage; however, it was felt by

the authors that better examination questions were formulated for open-book examinations, and the situation presented the student with a more useful and natural setting as he wrote the examination.

Furst (1958) reports further study of the two examination settings by Bloom at the University of Chicago resulted in Bloom concluding:

> "Thus, at one institute it was found that when comprehensive examinations consisted of knowledge questions while the instruction emphasized problem-solving skills, students tended to memorize information (and ignore much of the instruction) in order to pass the examinations (p 10)."

Bloom felt that use of examinations written in an open-book setting was a definite improvement because they tested problem-solving skills.

An extension of this point of view is given by Furst (1958, p 10). He claims that examinations of the open-book type foster the opposite of what is mentioned above. Students preparing for an open-book examination, he states, would seek to apply the methods and principles of the course to new situations and problems rather than cramming.

Tussing (1951) summarized in the following points why his college decided to adopt an open-book system of final examinations.

1. The test can be constructed and used in all the various forms that the traditional test can be used.
2. Much of the fear and emotional block encountered by the student is removed.

3. This system of testing points the course toward a different type of learning. Emphasis is placed on the practical problems and reasoning, and less emphasis is placed upon pure memory of facts and items.

4. Cheating with cribs and other devices is eliminated. A student feels that he has a good a chance to have the right answer as the fellow next to him.

5. This approach is more adaptable to evaluating student attitudes and presenting the question of what action should be taken on social issues (p 602).

In summary, he felt that the open-book final presented a practical means of achieving a valid measure of the work presented in a course. He did not, however, present any statistical evidence that favored open-book examinations over closed-book examinations.

Kalish (1958, pp 200-204) carried on Tussing's work in the following manner. He chose to consider three variables. First, he felt open-book examinations would lead to fewer errors. Second, open-book examinations measured different abilities than were measured by closed-book examinations. Last, there was no correlation between student ratings of the help received from examinations written in this setting and their test scores. The experiment which consisted of two groups, experimental and control, involved 158 students. Both groups had the same closed-book multiple-choice examination administered to them. Six weeks later one group wrote the second multiple-choice examination as an open-book examination and the other group wrote it as a closed-book examination. A replication of the study was run. The results were no significant difference in the number of errors per examination in the open or closed-book groups. A small significant differ-

ence was found in comparing the correlations of scores students received on their first and second examination. The open-book examination appears to be testing different abilities. No significant relationship was found between the attitude exhibited by the student toward open-book examinations and their achievement on the examination. However, this may have been a result of the rather ambiguous way attitude was tested. The student was asked in one question how much open-book examinations helped him. Kalish concluded that more research is needed before open-book examinations can be used in the most efficient fashion.

Based on the work done earlier, Feldhusen (1961, pp 637-645) investigated student attitudes to open-book and closed-book examinations on both objective and essay tests. Ninety students wrote two essay and two objective examinations - one of each type was open-book and one of each type was closed-book. After writing the examinations they recorded their reaction to them on a thirteen item questionnaire. Although the reactions to the questionnaire cannot be generalized to any great extent due to the select group involved in the study, the results of the questionnaire were generally favorable towards open-book examinations. Some points of particular interest are indicated below. The students felt they did equally well on open-book and closed-book examinations. They also felt the tension produced by a traditional closed-book examination was reduced in an open-book examination situation, and in general they preferred open-book examin-

ations. Finally, they felt that preparation methods were approximately the same for the two types of examinations but that open-book examinations reduced memorization of factual material and superficial studying. This descriptive study also had made a good start toward developing an attitude scale which measures a student's attitude toward examinations written in an open-book setting.

The final study concerning open-book examinations that is discussed in this section is a study conducted by Marco (1966). The general purpose of the study was to relate psychological and psychometric correlates of achievement test modes. Four classes of educational psychology students (N = 166) at the University of Illinois served as the subjects of the investigation. Measures were made in the cognitive domain, affective domain and the environmental situation used in the study during a seven week period. Classroom achievement tests, the Openness Discrimination Measure, and selected tests from the Kit of Reference Tests for Cognitive Factors represented the cognitive domain, while the Guilford-Zimmerman Temperament Survey and the Anxiety Differential covered the affective domain. The Openness Indicator was the only situational measure used in the study.

As a result of a carefully planned study a number of conclusions were reached about the relationships of the above variables. First, Marco found that achievement was consistently better on open-book examinations, although differences were small and of no practical importance. Second, knowledge items appeared to be better evaluation

instruments for subjects whose temperament favored the open-book test mode compared to those whose temperament favored the closed-book test mode. Also of interest was his finding that subjects on an open-book examination were less anxious when anxiety was measured on the Anxiety Differential. Findings concerning the test as a whole showed that some test variances and reliabilities were higher under the open-book test mode and there was little difference between the two test modes in regard to validity.

In the study presently being conducted some of the areas of Marco's work will be repeated. Similar comparisons of student achievement, test means, variance, reliability and validity will be made. No work will be done with predictive validity since an anchor test was not used in this study.

This study examines the effect of anxiety levels, setting and achievement. The Anxiety Differential used in this study was also used by Marco. In this investigation student attitudes to testing and mathematics are studied. Marco did not explore attitudes in his study. Marco tried to establish, by the use of factor analysis, an open and a closed-book factor. He was not able to find any single factors that satisfied his requirements. Thus this part of his study has not been repeated.

Since the results of Marco's work can only be generalized to other subject areas and age groups with extreme care because of sample size and composition, it is hoped the present study will generate more universal results. The present study involves over 600 students

selected on a sampling basis from the province of Alberta at the grade XII level. Calculations concerning test variance and reliability in Marco's study must be viewed with reservations since each test form administered contained only twenty multiple choice items, each with four alternatives. These tests he further divided into two sub-tests each composed of ten knowledge and ten application items. Calculations carried out on total test and sub-test values, as Marco pointed out, would be greatly affected by chance and the technical construction of formulas. The comparison of knowledge and application questions would be similarly affected. It is hoped the analysis of data in this present study avoids some of these limitations found in Marco's results. A comparison of his findings and the findings of this study is given in Chapter V to show which results are duplicated by this present study. Marco's study presents to date the best empirical investigation of the possible relationship of psychological and psychometric factors to achievement test written in different settings.

FACTORS RELATED TO THE TESTING OF MATHEMATICS

The three factors related to the testing of mathematics in this study are achievement, attitude and anxiety. Since the two specific test settings investigated are open and closed-book, these three factors are looked at in each of these settings so comparisons can be made. Achievement was measured on two parallel tests, each

administered in a different setting. Attitude to the two different test settings was measured, as well as attitude to the subject matter. Anxiety was measured in a neutral setting and in the two different testing settings.

The achievement score was the number of items a student answered correctly on a particular test. As indicated in earlier studies the number of items a student answered incorrectly did not seem to be related to the type of examination he was writing. Further investigation into other factors, such as aptitude, affecting the achievement of the student were not considered in this study. It is assumed that these factors were randomly distributed among the students and did not affect their achievement in any systematic way.

Factors related to test achievement that also were considered in this study were test variance, reliability and validity. It was necessary to consider these factors if a thorough comparison was to be made between examinations written in an open-book and closed-book setting. The articles included on test variance, reliability and validity consider ways for producing an accurate measuring instrument. Tests written in this study in two different settings are compared to determine which setting produces the optimum testing instrument. In this study the length of the test was ignored, since the number of items on each form was the same.

The following articles deal with test variance and reliability. Test variance was studied in this investigation to determine if it

increased in the open-book test setting. This would have implied that the open-book setting was more reliable than the closed-book setting. Other means of measuring the reliability of test items and tests are also discussed. Where it was possible these aspects of the data have been discussed in chapter four to determine the reliability of a test written in each test setting.

Gulliksen (1945, pp 79-91) dealt with the effects of items difficulty on item intercorrelations, test variance, and test reliability in a "well constructed" test. Under certain assumptions he showed that raw score variance increases as the (A) variance of item difficulties decreases for any given average item difficulty, (B) average item standard deviation increases, and (C) average item intercorrelation increases. Looking at item intercorrelations in more detail he also showed that the correlation approaches one only when items have the same difficulty value. Later work by Gulliksen (1950) substantiated these findings. His major finding again was that raw score variance increases as the average index of reliability, $r_{xg}\, s_g$, increases, where r_{xg} is the product-moment correlation of item g with the total test, and s_g is the standard deviation of item g.

Swineford (1959, pp 26-30) derived multiple regression equations to predict the standard deviation of scores, test reliability, and item test mean correlations. She worked with tests that were corrected for guessing and those where only the right responses were counted. Her general investigations confirmed Gulliksen's results

that raw score variance increases as the variance of item difficulties decreases and indicated that this conclusion can be extended to the case where scores are adjusted for chance success. A second finding showed that test variance increased as the average correlation of items with the test increases. Test reliability has been studied by both Gulliksen and Swineford. The measure of test reliability in both cases was the Kuder-Richardson formula 20 (KR-20). Gulliksen (1945) showed for a "well-constructed" test that test reliability increases as:

(A) the average item intercorrelation increases;

(B) the average correlation between the item and the total test increases;

(C) the variance of item standard deviations (or difficulties) decreases;

(D) mean item difficulty approaches 50%.

In his later study Gulliksen (1950) pointed out an obvious implication of the Kuder-Richardson Formula 20. The test reliability also increases as the average item variance decreases relative to the total test variance.

Work by Zimmerman (1968, p 41) and others showed that reliability can only be directly compared using Gulliksen's model when the mean and variance of the observed scores in the two samples are equal. Modifications of the formula were given for cases when the above condition did not hold. A second article by Zimmerman (1967) and others established that guessing introduces an error value that lowers

reliability.

Working with a predicted and real model of scores Payne and Anderson (1968) explored the characteristics of the KR-20 and came to a number of conclusions. Their findings generally supported previous findings about reliability as well as placing new emphasis on the large effect of score distribution on KR-20, the inverse relationship of range, number of items and KR-20, and marked relationship of population composition and stability of KR-20.

Further work by Swineford (1959) was concerned with developing multiple regression equations for predicting test reliability from a measure of the test standard deviation and the inverse of the squared average biserial correlation of items with the total test. Using the KR-20 as a measure confirmed parts (B) and (C) of Gulliksen's 1945 work and also indicated that test reliability increases as test variance increases.

More recent work by Ebel (1969) shows that relationship between reliability and the number of choices per item. His function is close to the Spearmen-Brown formula. The increase in reliability reaches a maximum when choices go from two to three. An example of his model states that a good test of 100 items with four alternatives each should have a KR-20 $\geq$ .86.

Woffard and Willoughly (1969) report the work of Cox who looked at reliability from yet another viewpoint. He studied the relationship of item difficulty, test length, size of upper and lower critical

groups, item selection methods and confidence levels. He found higher reliability for longer tests (10 versus 22 items), the lowest reliability for a difficulty range of 0 to 1.00 and no change in reliability over a difficulty range from .25 to .75. A secondary result of his study was that difficulty or test length did not affect the concurrent validity of the test. Thus his work yielded practical results to be used when relating reliability to other test measures.

Many individuals have studied the effects of validity in test construction. Validity is considered only briefly in this study. The main concern during this investigation was the validity of the examinations' content. Some background information dealing with validity is given in the following paragraphs. Where possible the results of these studies were duplicated using the data collected during the investigation to determine the validity of the examinations.

One of the first individuals to be concerned with validity was Thelma G. Thurstone (1932). She explored the influence of item difficulties on the diagnostic value of a test. Composing a number of subtests of homogeneous difficulty from a large spelling test and computing Pearson's r between the subtest scores and total test scores, she found the validity coefficients for each subtest. The highest validity coefficient was for a subtest composed of item difficulties ranging from .45 to .49.

Tucker (1946) using a different approach studied factors which increase reliability but at the same time decrease validity. He investigated the relation of item discrimination and item intercorrelation to the correlation between a test and a perfect measure of the ability the test was supposed to measure. His assumptions were (A) all items measure the same characteristic, (B) have equal reliabilities and (C) have equal difficulties. The item difficulty and number of test items were varied. The results of his investigation showed that, in order to maximize validity for tests with more than a single item, the item correlations and discriminating power had to be diminished. For example, if validity was to be maximized for a 10 item test, its item intercorrelations should approach .50 and item discriminating power should be 1.13 (0 was perfect). Thus tests composed of items of equal difficulty have maximum validity when items have less than perfect discriminating power and item intercorrelations.

Brogden (1946) also studied ways of maximizing validity. He attempted to determine the distribution of item difficulties which would give the largest product-moment correlation of a test with a perfect measure of the characteristic (normalized true score). He considered four different difficulty patterns. He noted that for the normal curve pattern, one of the four he considered, item intercorrelations of .6 and .8 had higher validities. In comparing validity

coefficients and Kuder-Richardson reliabilities, he also found that validity, in contrast to reliability, does not increase as the average item intercorrelation increases and as item variability decreases. It increases as average item difficulty approaches .50 only for tests with item intercorrelation of .6 and .8.

The insensitivity of reliability over a wide range of difficulty levels reported by Cox strengthens Brogden's conclusions, indicating that a test will have near maximum validity for many different difficulty levels and yet retain a high reliability.

Further work on the problems introduced by Brogden was carried out by Cronback and Warrington (1952). They examined special difficulties patterns to determine the relationship of validity to changes in variability of item difficulties and the precision of items (closely related to item-total correlations). Their major conclusions showed that as $s_d + s_y$ increases, overall test validity increases up to a maximum and then declines, where s_d is the variance of a particular measure of item precision (the higher the variance is, the less precise are the items) and s_y is the variance of the distribution of item difficulties. He found the maximum validity occurs when $s_d + s_y$ is about .50. It does not occur at high levels of item intercorrelations.

This section indicated some of the work that has been done in the area of test validity. Additional work by Kaiser and Carter (1971) and Horn (1971) has been completed which confirms the conclusions

reached in this section.

Many studies have been made concerning the relationship of anxiety to academic performance. McCandless and Castenda (1956) administered the children's form of the Manifest Anxiety Scale to a large school population and calculated correlations between it and various aspects of school achievement. They concluded that anxiety was significantly correlated with the complexity of the task of the task or subject. For example, students suffered from interference by anxiety while doing arithmetic rather than routine spelling and it also became more important in the higher grades.

Atkinson (1964) investigated McClelland's theory of achievement motivation. He concluded that anxiety level had a significant affect on achievement. Studies showed that a highly anxious individual would give a less accurate performance on a complicated task. An optimum level of anxiety produced the most accurate performance in a learning situation. Need for achievement was also found to correlate negatively with the psychological symptoms of anxiety. Atkinson's work showed that anxiety level, achievement and performance were closely related. His work implied a moderate level of anxiety would result in the most accurate performance or highest level of achievement being attained.

A study conducted at the elementary level by Reese (1961)

showed that an inverse relationship existed between manifest anxiety and the number of correct responses on achievement tests. He pointed out that IQ has little effect on the correlation between manifest anxiety and achievement, but prediction of achievement was not significantly improved by combining manifest anxiety and intelligence.

These studies sample some of the work conducted on the relationship of performance and anxiety. Little work appears to have been done on the relationship of different types of testing, anxiety levels, and performance on these tests. Most of the work done with examination types are logical arguments explaining why writing examinations in open-book setting should be less stressful than writing examinations in closed-book settings. More work must be done in this area to determine the empirical relationships that exist between anxiety and the type of examination the student writes.

The questionnaire used in this study to obtain a general measure of student anxiety to an examination setting was the "Anxiety Differential". This instrument is a variation of the anxiety differential developed by Husek and Alexander (1963). Their work was based on Osgood et al (1957) idea of a semantic differential which utilizes the relation between selected adjectives and concepts to express the difference in meaning among concepts. The fundamental assumptions underlying the use of such an instrument as an anxiety measure are that a person who is in an anxious state perceives things differently than a person who is not, that these different perceptions produce

changes within the individual, and that among the changes are changes in the meaning of things. An instrument which measures the differences in meaning allegedly measures the differences between anxiety states too. Marco (1966) describes the actual development of this instrument in detail. The complete questionnaire consists of 22 items, of which seven are fillers.

The student responding to the questionnaire is asked to indicate with an "x" the point on the line that represents his feeling about a particular word. High reliability has been indicated for this test in several different administrations. Reliability coefficients (Cronback's alpha) range from .58 to .75. The correlation scores on the questionnaire for a neutral setting as compared to examination setting is low (.58).

All the results accumulated so far indicated the questionnaire is a measure of anxiety. Marco's study showed the anxiety level was sufficiently lower in the neutral setting than in four separate testing settings. Since motivation and nervous tensions should be low in a neutral setting, the questionnaire scores confirmed expectations. Marco also carried out other measures of construct validity such as correlating the questionnaire results with the Guilford-Zimmerman Temperament Survey (GZTS) scales of Emotional Stability (E) and Objectivity (O), which were related to the general anxiety factor measured by Cattell's 16 P.F. Both in Cattell's work and Marco's work the correlation of these two factors with an anxiety measure produced

large negative loadings. In summary it appears to be an accurate measure of test anxiety and also a readily administered test for group situations.

Negative and positive attitudes directed toward mathematics by students influence their mathematical performance. Various studies have been conducted analyzing the relationship of students achievement in a course and his attitude toward it. The following studies are concerned with the effect a positive or negative attitude has on achievement in mathematics or arithmetic.

Churchill, reported in Biggs (1959), suggested that the cause of strong dislike or even fear, which many adults show towards arithmetical operations, may be faulty development of number concepts. This dislike or fear has its foundations in elementary school where children are taught to calculate without sufficient understanding.

Many opinions similar to the one above have been given as a reason for the origination of a dislike of mathematics. Once students have these attitudes though, how do they explain them? A study conducted by Dutton (1964) at the junior high school level asked the student to list why they liked or disliked mathematics. Reasons they gave for disliking mathematics were lack of understanding, too difficult and complicated, poor achievement and boring and repetitive. Students listed practicality, interest and challenge of mathematics as the reasons for liking it.

Two studies conducted in Britain showed a significant

relationship between strain and dislike of mathematics. Pritchard (1935) showed that boys and especially girls dislike arithmetic because of a feeling of incapacity and strain when dealing with difficult items in the curriculum. A second study by Freeman (1948) points to inability to master technical difficulties as the most common reason for not liking arithmetic.

Other studies have been conducted to determine reasons for a student disliking mathematics. Factors that have been investigated are effects of parent's attitudes (Proffenberger and Norton, 1961), effects of teacher-student rapport (Pritchard, 1935 and Biggs, 1959), and effect of curriculum (Remai, 1965). The results of these studies indicate that a dislike of mathematics may stem from type of curriculum or the actual material being taught. The extent that parents or teachers influence attitudes in mathematics is as yet largely undetermined.

A great deal of work remains to be done in determining the relationship of attitudes towards mathematics and the achievement of students in mathematics. Various methods of appraising attitudes in the learning of mathematics are discussed by Corcoran and Gibb (1961, p 106). They feel that appraisal must cover attitudes toward specific mathematics courses and such specific aspects of mathematics as computations, problem solving, figure construction, and the reasons why he studies them. In their article they discuss various means of doing this. However, at this time little has been done to test these

specific areas. In fact it is difficult to find a reliable test of attitude for the subject of mathematics, without worrying about the topic areas within the subject. A measure of attitude to the test situation is also extremely difficult to locate. Little empirical research has been done in this area.

Mortlock (1969) modified the original attitude opinionnaire, developed by Aiksen and Dreger, for use in a senior high school mathematics program. The modified form consisted of twenty attitude statements about mathematics. The results of Mortlock's administration showed the opinionnaire to be a reliable instrument. The students responded to the opinionnaire by marking each item on a 5 point scale ranging from strongly disagree to strongly agree.

Similarly, Mortlock (1969) modified the questionnaire, developed by Mandler and Sarasen, on attitude toward mathematics testing. The questionnaire consists of 15 items in its final form. Each item has a line segment with the end points marked with written descriptions of extreme anxiety reactions and an indicated mid point. The purpose of the questionnaire is to have the student rate himself on items descriptive of anxiety reactions in test situations. The student is asked to indicate with an "x" the point on the line that represents his attitude to the testing situation. The test-retest reliability of the questionnaire is .91. Various highly correlated scoring

techniques for the questionnaire have been reported.

SUMMARY

In summary, it appears that the evidence gathered about group examinations administered in an open-book setting is neither complete nor very consistent. Much of the work done has been of a descriptive nature with little effort to establish general theory applicable to many situations. Also the work done on those variables associated with the different test settings being examined in this study is scarce and incomplete. A good beginning has been made but more research is needed on a larger scale and at the senior high school level if the results are to be applicable to the Alberta educational scene.

It is hoped that this present study will replicate and generalize some of the results cited in the literature above. As many of the relationships concerning achievement, validity, reliability and affective measures have been considered as possible.

CHAPTER 3: METHODOLOGY

DESIGN FOR THE STUDY

Six hundred and seventy grade twelve Mathematics 30 students were selected according to a provincial pre-testing grid which divided the province into 12 sections, based on population density and geographical area. The open-book and closed-book examinations, anxiety and attitude scales were administered to the students as part of the normal pre-testing program during the first week of June, 1971.

Table 1 shows the distribution of classes used in the study for each section of the provincial grid.

TABLE 1

Distribution of Classrooms Used in Study

	Urban A Pop. = over 25,000	Urban B Pop. = 5,000 to 25,000	Urban C Pop. = 2,500 to 5,000	Pop. = less than 2,500
North Alberta	10	1	4	1
Central Alberta	2		2	
Southern Alberta	3	1	2	1

The students were randomly assigned to four groups by classroom lot. The randomness was within each of the provincial sections so no regional differences would influence the study. Each group had a different order of setting and test forms to prevent any systematic differences occurring because of order of administration.

Four instruments were prepared to examine the open-book setting versus the closed-book setting for examinations. They were:

1. Anxiety Scale (Appendix 1) *semantic differential*
2. Attitude to Testing (Appendix 1) *questionnaire on attitudes to testing*
3. Attitude to Mathematics (Appendix 1) *attitude to mathematics opinionnaire*
4. Two parallel Mathematics 30 Achievement Tests (Appendix 1)

Explanations and justification for the use of each of the above questionnaires is given in Chapter 2.

The following section lists the schedule of activities carried out, plus a brief description of each activity.

1. The two parallel test forms, Forms A and B, were constructed by dividing the January 1971 Mathematics Departmental Examination into two parts. The taxonomy and subject area classification (Appendix 1) was considered and the division of items resulted in two forms that retained the same taxonomy and area proportions as the original examination. The taxonomy and subject area classification is indicated (Appendix 1) for each form. The correlation between Form A and B and the original examination is given (Appendix 1).

2. A brief description of the differences between writing tests in an open-book setting and closed-book setting was sent to the participating schools as well as preparation hints for students preparing for tests in the two different settings, (Appendix 1). Also included on this information sheet for teachers and students were the dates for each of the two testing periods. This information sheet reached the school approximately two weeks before the first testing period.
3. The Anxiety Scale was administered during a neutral setting chosen by the classroom teacher during the week prior to the administration of the actual testing program. Directions given to the teacher asked her to administer the questionnaire during the last or first ten minutes of an average instructional period according to the included instructions.
4. On the first testing day the test administrator again administered the Anxiety Scale to the students following the same procedure as used before. Next he administered the appropriate test form (either A or B) in the setting selected for the day. Both the form of test and setting for that particular testing period were randomly determined several weeks prior to the testing period. The students were informed what to expect on that particular day approximately two weeks in advance (see point 1). Following the administration of the test form, the students were asked to record their impressions of the testing situation on the Attitudes to

Testing questionnaire. This concluded the first testing session.

5. The Attitude to Mathematics questionnaire was administered by the classroom teacher between the first and second testing session. The teacher was asked to select 10 - 15 minutes during an average instructional period and administer the questionnaire according to the included instructions.

6. On the second day, approximately four school days after the first day, the test administrator again administered the Anxiety Scale to the students following the same procedure as used before. Next he administered the appropriate test form in the setting selected for that day. Both the form and setting were the opposite to that used the previous testing day. For example, if the class had written Form A in an open-book setting on the first day, they now would write Form B in a closed-book setting. Again the students were informed in advance what type of test and setting to expect (see point 2). Following the administration of the test form, the students were asked to record their impressions of the testing situation on the Attitude to Testing questionnaire. This concluded the second testing session. This program is summarized in the following table.

SCHEDULE OF TESTING

GROUP	INFORMATION SHEET	ANXIETY SCALE 0	ANXIETY SCALE 1	TEST FORM 1	SETTING 1	ATTITUDE TO TESTING 1	ATTITUDE TO MATH	ANXIETY SCALE 2	TEST FORM 2	SETTING 2	ATTITUDE TO TESTING 2
GROUP I $N_1 = 230$ $N_2 = 210$	*	*	*	Form A	Open	*	*	*	Form B	Closed	*
GROUP II $N_1 = 210$ $N_2 = 211$	*	*	*	Form A	Closed	*	*	*	Form B	Open	*
GROUP III $N_1 = 105$ $N_2 = 97$	*	*	*	Form B	Open	*	*	*	Form A	Closed	*
GROUP IV $N_1 = 125$ $N_2 = 107$	*	*	*	Form B	Closed	*	*	*	Form A	Open	*

Thus, the following set of data was to be collected for each student participating in the testing program.

1. Three Anxiety Scale scores.
2. Two Attitude to Testing scores.
3. One Attitude to Mathematics score.
4. Two achievement test scores, each one recorded after a different testing setting.

Due to administration difficulties and student attendance, it was not possible to collect all of this data for every student.

Depending on the hypotheses being considered, the above information was analyzed according to the total sample, by classroom lot, according to individual scores, or by groups of students that possess certain characteristics.

NULL HYPOTHESES

Hypotheses concerning examination performance on examinations written in a closed- or open-book setting:

1. There will be no significant difference between the January, 1971 departmental means and the June, 1971 experimental means expressed as percentage of items correct.
2. There will be no significant difference in mean between open-book setting and closed-book setting.
3. There will be no significant difference in mean between Form A and Form B.
4. There will be no significant difference in mean between Time 1 and Time 2.
5. There will be no significant difference in the difficulty rating of questions classified as knowledge, comprehension and

application questions when written in either an open or closed-book setting.

6. There will be no significant difference in the reliability of questions classified as knowledge, comprehension and application questions when the two settings are compared.

Hypotheses related to test variance, test reliability and validity:

7. There will be no significant difference in the variance of examination scores written in an open-book setting as compared with the variance of examination scores written in a closed-book setting.

8. There is no significant difference in the examination reliabilities of the examination written in an open and closed-book setting.

9. There is no significant difference in the validities of the examinations written in an open and closed-book setting.

Hypotheses related to individual items:

10. There is no significant difference in student responses to items written under a closed-book setting as compared with an open-book setting. (Descriptive survey of sample items).

Hypotheses related to student anxiety level:

11. There is no significant difference in student anxiety levels before the examination tested in an open-book setting as compared with a closed-book setting.

12. There is no significant difference in student anxiety levels in a neutral situation and before writing an examination in a closed-book setting.

13. There is no significant difference in student anxiety levels in a neutral situation and before writing an examination in an open-book setting.

14. There is no significant relationship between the level of anxiety and student achievement on either the open or closed-book settings for examinations.

Hypotheses related to student attitudes:

15. There is no significant difference in the student's attitude toward writing examinations in an open-book or a closed-book setting.

16. There is no significant relationship between the student's attitude toward mathematics and achievement in mathematics with respect to the two different examination settings.

17. There is no significant relationship between the student's attitude toward the testing situation and his achievement in mathematics.

The hypotheses could be considered in summary as:

1. those dealing with achievement and test statistics (Hypotheses 1, 2, 3, 4, 5, 6, 7, 8, 9, 10).

2. those dealing with anxiety levels (Hypotheses 11, 12, 13, 14).

3. those dealing with attitudes (Hypotheses 15, 16, 17).

CHAPTER 4: ANALYSIS OF DATA AND RESULTS

INTRODUCTION

The present chapter provides data about the questions that have been asked, some supporting previous findings and some indicating new information. It begins with pertinent statistical information regarding the two achievement tests used in the study, Form A and Form B, and then considers student achievement, anxiety and attitudes for the two settings.

The research hypotheses are considered in their statistical form, and tests of significance are reported. Most of the computations were carried out on the IBM 360/67 computer, Computing Services Section, University of Alberta. Computer programs were furnished by the Division of Educational Research Service, Faculty of Education, University of Alberta.

PRELIMINARY COMPARISONS OF FORM A AND B

Final examination papers were prepared approximately one year in advance for each end-of-term writing session. The paper used in this study had been prepared for the January, 1971 session. The actual preparation of the paper had been carried out as indicated in Chapter 2, History of Alberta's Departmental Examinations. The Mathematics 30 Departmental Examination was administered in a closed-

book setting to 6,000 grade XII Alberta students. The mean, standard deviation and reliability of the 70 item examination were:

Mean	Standard Deviation	Reliability
43.713	11.399	.898 (KR-20)

For the purposes of this study, the paper was divided into two parallel papers - each 36 items long. Since the items on the original paper were classified according to content area and thought level (Appendix 1), items within each category were randomly divided into two sets. The procedure insured adequate coverage of the course in each set. The two sets of items were labelled Form A and B (Appendix 1). Each thought level formed a sub-set on the test. The test items on the test were divided into three sub-sets. Form A had 6 knowledge items, 14 comprehension items and 16 application items. Form B had 5 knowledge items, 14 comprehension items and 17 application items. When comparisons were made between the knowledge or application items on Form A and Form B adjustments were made for the unequal lengths. The means, standard deviations and reliabilities based on the original population of 6,000 students resulted in the following values for each 36 item test:

TABLE 2

Means, Standard Deviation and Reliability of Original Forms

Form	Mean	Standard Deviation	Reliability (KR-20)
Form A	21.555	6.039	.7781
Form B	22.152	5.863	.7802

By comparing the means and standard deviations given above, it can be seen that Form A and B have equal difficulty levels and grouping of students. The difference between the two reliability values was very small.

Further comparisons were computed to determine that the two forms were both statistically and content parallel. The original test, Form A and Form B were compared by a series of correlations based on the original population. The resulting correlations were:

TABLE 3

Correlation Matrix of Form A, Form B and Original

Test	Form A	Form B	Original
Form A	1.000	0.834	0.959
Form B		1.000	0.956
Original			1.000

The table of correlations indicate that, although Form A and B are both shorter tests than the original, they measure the same abilities and yield a very similar ranking of students. The correlation between the two forms is very high - being .834. It seems fair to say that A and B are parallel to each other and parllel to the original.

Form A and B were tested in an open- and closed-book setting. The testing was carried out in a number of different sequences to insure no systematic mean differences resulted between groups of students as a result of sequencing. The actual testing procedures used

are described in detail in Chapter 3. The next section contains a systematic comparison of the four groups of students involved in this study.

RELATIONSHIP OF STUDENT ACHIEVEMENT TO THE TIME OF ADMINISTRATION, SETTING AND FORM

The design used to study the relationship of student achievement of the time of administration, setting and form was a three-way analysis of variance. A description of the selection process of the students used in the study is given in Chapter Three. The only difference that occurred between the four groups during the interval of the study was the time and order of administration of the different forms to each group. The testing sequence of the four groups of students is summarized in Table 4.

TABLE 4

Sequences of Form Administration

	Time 1	Time 2
Group 1	Form A - Open	Form B - Closed
Group 2	Form A - Closed	Form B - Open
Group 3	Form B - Open	Form A - Closed
Group 4	Form B - Closed	Form A - Open

Initially the classes were placed in the four groups randomly so there were an equal number of classes in each group. However, due to some classes dropping out of the study during the administration of the testing program, fewer classes participated in Groups 3 and 4

than in Groups 1 and 2. Since Groups 3 and 4 do not represent a random selection of Alberta Mathematics 30 classrooms, the results from these two groups must be interpreted with care. The actual number of students in each group is given in Chapter 3.

The null hypotheses that were tested by this design are:

H_1: There is no significant difference between the January, 1971 departmental means and the June, 1971 experimental means expressed as % items correct.

H_2: There is no significant difference in mean between open-book setting and closed-book setting.

H_3: There is no significant difference in mean between Form A and Form B.

H_4: There is no significant difference in mean between Time 1 and Time 2.

The linear model assumed for this analysis was:

Y = Error + Mean + A + B + AB + C + BC + AC + ABC, where the desired alpha level was 0.05. The interaction effects were not used in this study, therefore; they were not reported here. A complete set of tables containing both main and interaction effects is reported in (Appendix 2). Comparisons were made for knowledge items, comprehension items, application items, and total test items.

In Table 5, the assumed population general means were determined from the January, 1971 Mathematics 30 examination (closed setting). The actual sample general means were calculated as a result of the administration of Form A and B (closed) during the testing period of this study. The means for setting, form and time were also calculated

from student scores formed during the two test administration periods. All these values are given in percents in Table 5. Pairs of significantly different means are marked with an asterisk.

TABLE 5

Comparison of Setting, Form, Time and General Means

$\alpha = .05$

TABLE 5

MAIN EFFECTS

EFFECTS	Knowledge Items	Comprehension Items	Application Items	Total Test Items
Setting				
1) Open	73.79	64.80	56.62	62.43*
2) Closed	65.27*	61.64*	56.78	60.03*
Form				
1) Form A	71.89	61.36	56.46	60.85
2) Form B	67.17*	65.08*	57.04	61.61
Time				
1) Time 1	69.12	62.37	54.86	59.98
2) Time 2	69.94	64.17	58.54*	62.48*
General Means				
1) January, 1971	70.00	65.00	60.00	63.00
2) June, 1971	69.00	63.22*	56.70*	61.23*

The values in Table 5 show that the means for open set examinations were higher in three of the four cases - Knowledge, Comprehension and Total test. No difference existed for the Application sub-test. The form means show that Form B was slightly easier. A significant difference was found in two of the sub-tests - Knowledge and Comprehension. However, the means for the Application sub-test, which composed half of the total test, were not significantly different

between the two forms. Similarly, the means for the total test were not significantly different. A more detailed look at the Knowledge sub-test shows that Form A is the easier while a similar look at the Comprehension sub-test shows that Form B is the easier. Thus this significant difference between the two forms cancels out when the total test is considered and the difficulty level of the total tests is equal. There was a consistent relationship between the two times with the means for Time 2 always higher than the means for Time 1. This difference was significant for the Application sub-test and total test means. Finally, there was a significant difference between the means for the General Means in three cases - Comprehension sub-test, Application sub-test and total test. The January means were consistently higher with the largest difference being 4.30%.

All the experimental situations have been analyzed in this section. The results have led to the rejection of null hypotheses 1 and 2 for three of the four cases. Null hypotheses 3 and 4 have been rejected in two of the four cases.

ANALYSIS OF INDIVIDUAL TEST ITEMS

The individual test items on each form were analyzed in several ways. First, an item analysis was computed for the four groups of students' responses on each sub-test and the total test for Form A and B, open and closed. This provided for each item its

(a) Difficulty level

(b) Biserial correlation (measure of reliability)

(c) reliability (Spearman-Brown)

Since Form A was written in an open setting in groups 1 and 4 and in a closed setting in groups 2 and 3, this information was used to investigate differences between the closed and open settings. Form B was compared in the same way since it had been administered opposite Form A throughout the study and thus provided a replication of results.

Conducting an item analysis at each of the thought levels - Knowledge, Comprehension and Application, as well as for the total test, provided additional information. If just the complete test analysis had been considered, significant differences that occurred only at a particular thought level would have been overlooked.

MEAN, VARIANCE AND RELIABILITY OF FORM A AND B

The following tables compare variance and reliability for each of the sub-tests and total test for Form A and Form B under the two settings open and closed.

TABLE 6

Means, Variances and Reliabilities of Total Test

EFFECTS	FORM A				FORM B			
	Open		Closed		Open		Closed	
	group				group			
	1	4	2	3	2	3	1	4
Mean	.565	.664	.574	.631	.635	.636	.570	.625
Variance	.0210	.0257	.0267	.0252	.0264	.0250	.0285	.0269
Reliability	.7238	.7958	.7799	.7794	.7901	.7845	.7969	.7968

TABLE 7

Means, Variances and Reliabilities of Knowledge Sub-Test

Mean	.742	.778	.663	.673	.718	.714	.608	.648
Variance	.0389	.0392	.0450	.0400	.0508	.0536	.0564	.0576
Reliability	.2031	.3144	.2364	.1943	.2299	.3347	.2830	.3538

TABLE 8

Means, Variances and Reliabilities of Comprehension Sub-Test

Mean	.567	.677	.585	.622	.670	.677	.597	.661
Variance	.0296	.0345	.0357	.0352	.0318	.0284	.0338	.0298
Reliability	.5020	.6151	.5686	.5688	.5786	.5378	.5838	.5528

TABLE 9

Means, Variances and Reliabilities of Application Sub-Test

Mean	.489	.608	.530	.616	.580	.580	.536	.588
Variance	.0338	.0383	.0376	.0358	.0384	.0368	.0376	.0369
Reliability	.5901	.6531	.6451	.6463	.6702	.6565	.6453	.6520

Looking at the student means on the total test and each of the sub-tests very small differences are noted. Considering the total test values first, the four open means were .565, .635, .636 and .664. No significant differences were found between these means ($\alpha = .05$) when they were compared by means of an F-ratio. It should be noted that means, variances and reliability for Group 1 (open) were lower than would be expected throughout the set of data. A considerable number of students from two classes in Group 1 were present for only the first

test administration and missed the second closed-book set examination. These classes were in a large urban high school which had a lower mathematics attitude score than the general population. In the closed-book set examination, Group 1 has similar values to the other groups. The students involved in the first setting only contributed to the lower statistic values for that group on the open setting.

The four closed means for the total test were .570, .574, .631 and .625. Again no significant differences were found between these means ($\alpha = .05$) when they were compared by means of a F-ratio. Equal means for all open-set forms and equal means for all closed-set forms show again that the groups of students were equivalent. This information later permits the comparison of different groups of students with the assumption they are equal under identical situations.

Comparisons similar to those above can be made for each of the sub-test means. It should be noted that these means supported relationship established earlier between the open and closed-book setting. Means in all cases were slightly higher for open-book set test forms in comparison to their opposing closed-book set test forms. No significance tests were computed here since this relationship had already been investigated.

Homogeneity of variance for each of the total tests and sub-tests was considered next. Analyzing each set of variance, using an approximation of Bartlett's homogeneity of variance test, it was found that:

a) forms written in different settings did not have a significantly different variance.

b) forms written in the same setting did not have a significantly different variance.

c) sub-tests written in different settings did not have a significantly different variance.

d) sub-tests written in the same setting did not have a significantly different variance.

This indicates that variance does not increase significantly in an open-book set examination versus a closed-book set examination.

Thus, the null hypothesis 7:

> There will be no significant difference in the variance of examination scores written in an open-book setting as compared with the variance of examination scores written in a closed-book setting.

was not rejected for the total test of the three sub-tests.

The final information that is considered from these tables are the reliability values. Internal consistency reliability estimates were computed by Spearman-Brown's formula. This permitted comparison of tests of different length. The reliability values for the Knowledge sub-test were very low. This was caused by the small number of items on each sub-test. Very little significance should be attributed to the Knowledge sub-test reliabilities as a result. A comparison of the reliability values for the other sub-tests and the total test indicate very small differences between groups. The reliabilities for total test do not differ by more than .05 between the two settings.

Null hypothesis 8:

> There is no significant difference in the examination reliabilities of the examinations written in an open and closed-book setting.

was not rejected for the total test or three sub-tests. A visual comparison of values showed no consistent difference between the two settings.

The composition of the original groups was randomly determined, therefore chance differences in reliability were all that could be expected to arise if setting did not influence reliability. In these cases the open-book test setting had higher reliabilities estimated in 6 of the 12 cases. This additional information seems to confirm that the reliability differences were randomly distributed between the two settings.

Validity of Form A and B

The content validity of the original test was determined by a committee of highly qualified teachers who constructed the original items, reviewed pre-test results and developed the original paper. The end result of their work was a test that was representative of the objectives and content of the Mathematics 30 course. When the original test was divided into Forms A and B, the emphases of the original paper was carefully retained. The means, variances and correlations given earlier verify this fact. Tables 10, 11 and 12 give measures of mean item difficulties, variance of mean difficulties and

mean biserial correlations of each sub-test (Form A) for the four groups of students involved in the study. Groups 1 and 4 are open-book examination values and groups 2 and 3 are closed-book examination values. The results for Form B are similar.

TABLE 10

Mean Difficulties, Variance of Difficulties and Mean Biserial Correlation Coefficients of Knowledge Sub-Test

Group	Mean Difficulty	Variance (Dif)	Mean Biserial Correlation
1	.7778	.0121	.6647
2	.6925	.0231	.6095
3	.6620	.0072	.6075
4	.7407	.0151	.6125
Total	.7181	.0151	.6235

TABLE 11

Mean Difficulties, Variance of Difficulties and Mean Biserial Correlation Coefficients of Comprehension Sub-Test

Group	Mean Difficulty	Variance (Dif)	Mean Biserial Correlation
1	.6764	.0314	.5890
2	.6211	.0219	.5179
3	.5851	.0407	.5268
4	.5679	.0407	.5033
Total	.6126	.0299	.5342

TABLE 12

Mean Difficulties, Variance of Difficulties and Mean Biserial Correlation Coefficients of Application Sub-Test

Group	Mean Difficulty	Variance (Dif)	Mean Biserial Correlation
1	.6084	.0229	.5693
2	.6148	.0332	.5523
3	.5313	.0357	.5464
4	.5097	.0271	.4992
Total	.5660	.0300	.5418

The values in Table 10, 11 and 12 show differences between settings to be very small. There does not appear to have been any distinct difference between the validity measure for the two settings - open and closed. Mean difficulties and biserial correlations were very similar. Only for the Knowledge sub-test were the means of Group 1 and 4 (open) consistently higher than the means of Group 2 and 3 (closed).

The null hypothesis 9:

> There is no significant difference in the validities of the examinations written in an open and closed-book setting.

was not rejected.

DESCRIPTIVE COMPARISON OF INDIVIDUAL ITEMS

An overview of items on Form A has been made in the two settings- open and closed- to determine if any differences exist in the way

students respond to the same item in two different settings. The survey provided information for null hypothesis 10:

> There is no significant difference in student responses to items written under a closed-book setting as compared with an open-book setting.

Biserial correlations and difficulties, discussed earlier, were not considered.

The distribution of students over each of the alternatives did not greatly change between the two settings. Poor students, writing in an open-book setting, were unable to take advantage of their textbook or other references and still selected incorrect responses. However, in many items the Z-score on the distractor was lower in the open-book setting indicating that average students were able to secure the necessary information and select the correct response. In some cases the greater number of average students selecting the correct response caused the biserial correlation to be lower for the open-book item. The good student generally selected the correct response in both settings. This may indicate that he received little help from the open-book arrangement since he already knew the correct response.

Perhaps more marked differences in student responses could be seen if items were designed especially for an open-book examination. At present there seems to be little difference between the two settings on an examination originally designed for a closed-book setting.

COMPARISON OF ITEM DIFFICULTIES ON FORM A AND B

The item difficulties for each of the items were calculated as part of the item analysis. These values were compared for the open-book and closed-book setting for each of the sub-tests. The results of this comparison are given in Table 13. Group 1 (open) was compared with Group 2 (closed) and Group 3 (open) was compared with Group 4 (closed). A chi-square test was used to make the comparison. The chi-square value of 1 versus 2 is the first number in the square and 3 versus 4 is the second value.

TABLE 13

Comparison of Item Difficulties, Form A and B

$\alpha = .20$

	FORM A		FORM B	
	chi-square	probability	chi-square	probability
Knowledge Sub-Test	7.539 4.145	.1835 .5287	16.000 10.464	.0030 .0333
Comprehension Sub-Test	21.632 9.486	.0101 .3937	14.092 10.030	.1191 .3480
Application Sub-Test	5.859 3.447	.7539 .9439	2.149 3.623	.9889 .9344

The probability of a chi-square occurring greater than that observed is listed in the probability column in Table 13. For the Knowledge sub-test three of the four group comparisons had low probabilities. This indicates that the difficulty of Knowledge sub-test items differ

significantly between the two settings- open and closed. For the Comprehension sub-test two of the four group comparisons had low probabilities. This indicates that there was a significant difference between the difficulty levels of Comprehension items in two cases when the settings were compared, but not as consistent a trend as with the Knowledge sub-test items. The Application sub-test had no low probabilities for the item difficulty comparisons between the two settings. The results of this analysis confirmed the findings on the student achievement scores. Null hypothesis 5:

> "There will be no significant difference in the difficulty rating of questions classified as knowledge, comprehension and application questions when written in either an open or closed-book setting."

was rejected for Knowledge items only.

COMPARISON OF ITEM BISERIAL CORRELATION COEFFICIENTS ON FORM A AND B

A biserial correlation coefficient was calculated for each item as part of the item analysis. The biserial correlation coefficient is continuous on one variable, total test score and dichotomous on the other, individual item. Because of the structure of biserial correlation coefficients, it is difficult to compare sets of these values. The objective of this section was to determine if the individual biserial correlation coefficients confirmed the previous findings with sub-test reliabilities for the two test settings- open and closed.

The biserial correlation coefficients for both Form A and B in

in the open and closed-setting were compared. The data for Form A are given in Table 14, 15 and 16. Similar results were found to hold for Form B. In each case the closed value was subtracted from the open value. Comparisons were made for each sub-test.

For knowledge items the biserial correlation coefficient increased in the open-book setting 7/12 of the time. This indicates that having materials available for knowledge items decreased slightly the number of random errors students made while answering them. The good student made fewer errors on easy items. The same small trend held for comprehension items with the biserial correlation coefficient for the open-book setting increasing 4/7 of the time. However, in the application items an equal number of biserial correlation coefficients increased and decreased. This indicates that other factors, besides setting, have the major influence on reliability at this thought level. Student accuracy is not increased on application questions if outside materials are available.

This set of data generally supports the findings on total and sub-test reliability of no significant difference. The slight advantages for open-book settings given for the Knowledge and Comprehension sub-tests indicate at best a slight trend to favor one setting. On Form B there was no increase in biserial correlation coefficients for the Comprehension sub-test, only the Knowledge sub-test. Thus, with an equal increase and decrease of biserial correlation coefficients between the two settings, they must be called equally reliable testing procedures.

COMPARISON OF BISERIAL CORRELATION COEFFICIENTS (OPEN - CLOSED)
FOR ITEMS ON EACH SUB-TEST

TABLE 14 - KNOWLEDGE Sub-Test

Item	Group 1	Group 2	Group 1-2	Item	Group 3	Group 4	Group 3-4
1	0.609	0.732	-0.123	1	0.654	0.709	0.055
2	0.568	0.535	0.033	2	0.514	0.597	0.083
3	0.382	0.580	-0.198	3	0.421	0.424	0.003
4	0.896	0.750	0.146	4	0.682	0.644	-0.038
5	0.796	0.484	0.312	5	0.664	0.593	-0.071
6	0.743	0.581	0.162	6	0.716	0.714	-0.002

TABLE 15 - COMPREHENSION Sub-Test

7	0.583	0.571	0.012	7	0.583	0.535	-0.048
8	0.707	0.473	0.234	8	0.603	0.273	-0.330
9	0.602	0.297	0.305	9	0.433	0.542	0.109
10	0.421	0.593	-0.172	10	0.414	0.400	-0.014
11	0.897	0.558	0.339	11	0.497	0.588	0.091
12	0.513	0.167	0.346	12	0.415	0.423	0.008
13	0.531	0.632	-0.101	13	0.570	0.580	0.010
14	0.544	0.714	-0.170	14	0.607	0.543	-0.064
15	0.519	0.619	-0.100	15	0.650	0.518	-0.132
16	0.405	0.586	-0.181	16	0.536	0.606	0.070
17	0.512	0.497	0.015	17	0.383	0.425	0.042
18	0.789	0.645	0.144	18	0.532	0.624	0.092
19	0.641	0.539	0.102	19	0.547	0.498	-0.049
20	0.596	0.374	0.222	20	0.619	0.505	-0.114

TABLE 16 - APPLICATION Sub-Test

21	0.326	0.605	-0.279	21	0.576	0.450	-0.126
22	0.796	0.752	0.044	22	0.694	0.421	-0.273
23	0.615	0.354	0.261	23	0.529	0.506	-0.023
24	0.627	0.505	0.122	24	0.568	0.530	-0.038
25	0.599	0.531	0.068	25	0.635	0.631	-0.004
26	0.701	0.621	0.080	26	0.591	0.651	0.060
27	0.686	0.441	0.245	27	0.459	0.573	0.114
28	0.836	0.583	0.253	28	0.622	0.663	0.041
29	0.769	0.733	0.036	29	0.475	0.594	0.119
30	0.460	0.679	-0.219	30	0.500	0.342	-0.158
31	0.305	0.458	-0.153	31	0.528	0.588	0.060
32	0.657	0.706	-0.049	32	0.640	0.468	-0.172
33	0.479	0.384	0.095	33	0.544	0.267	-0.277
34	0.273	0.272	0.001	34	0.307	0.318	0.011
35	0.491	0.669	-0.178	35	0.646	0.662	0.016
36	0.505	0.560	-0.055	36	0.444	0.340	-0.104

Null hypothesis 6:

> "There will be no significant difference in the reliability of questions classified as knowledge, comprehension and application questions when the two settings are campared."

was not rejected.

ANALYSIS OF STUDENT ANXIETY AND ATTITUDE SCORES

This portion of the chapter deals with the student responses to the two achievement tests, anxiety scale and two attitude questionnaires. Student responses to each instrument are first discussed separately and then the possible interrelationships between them are considered.

Student achievement has already been considered for the total test and sub-tests and the results reported. Thus, in this section, achievement is considered only as it relates to the other factors mentioned above.

ANALYSIS OF ANXIETY SCORES

A measure of anxiety level was obtained in three situations. The Anxiety Differential was administered by the classroom teacher in a neutral setting during the week prior to the first testing session. It was again administered by the test administrator prior to the writing of Form A and the writing of Form B. The students had been informed what setting to expect for each form they wrote and came to the testing situation prepared to write under the conditions determined for that day.

Each testing day the students filled out the Anxiety Differential with responses indicating how they felt about the approaching test they would write. A more detailed description of administration is given in Chapter 3.

The means of the anxiety scale for the three times were

Time Neutral	45.3678
Time Open	48.0191
Time Closed	49.3755

A survey of the means show that the general level of anxiety was highest for students before writing a closed-book set examination and lowest in the neutral situation. Analysis of the data as a single factor experiment with repeated measures yielded the results shown in Table 17.

TABLE 17

ANALYSIS OF ANXIETY SCALES AS A REPEATED MEASURE

$\alpha = .05$ $n = 367$

Source of Variation	SS	DF	MS	F-ratio
Between People	76,209.0	366	208.221	
Within People	51,642.0	734	70.357	
Treatment	3,289.0	2	1,644.500	24.8955
Residual	48,353.0	732	66.056	
Total	127,851.0	1100		

Letting $\alpha = .05$, this analysis showed that there were significant differences between the three anxiety levels. The anxiety level of students appeared to rise significantly when measured from a neutral setting to a testing setting. Further comparisons between the three anxiety scale set of responses yielded the following tables.

TABLE 18

COMPARISON OF NEUTRAL AND OPEN ANXIETY LEVELS

$\alpha = .05$ n = 410

Source of Variation	SS	DF	MS	F-ratio
Between People	59,952.0	409	146.582	
Within People	31,105.0	410	75.866	
Treatments	2,060.0	1	2,060.000	29.0081
Residual	29,045.0	409	71.015	
Total	91,057.0	819		

TABLE 19

COMPARISON OF NEUTRAL AND CLOSED ANXIETY LEVELS

$\alpha = .05$ n = 367

Source of Variation	SS	DF	MS	F-ratio
Between People	52,075.0	366	142.281	
Within People	27,467.0	367	74.842	
Treatments	3,045.0	1	3,045	45.6339
Residual	24,422.0	366	66.727	
Total	79,542.0	733		

TABLE 20

COMPARISON OF OPEN AND CLOSED ANXIETY LEVELS

$\alpha = .05$ n = 522

Source of Variation	SS	DF	MS	F-ratio
Between People	96,270.0	521	184.779	
Within People	31,739.0	522*	60.803	
Treatments	480.0	1	480.000	8.003
Residual	31,259.0	521	59.998	
Total	128,009.0	1043		

*522 students were involved in the last analysis as classes and students who did not have a score on the neutral anxiety scale could be used.

The results of the first two tables confirmed that both the open and closed-book test setting produced a significant increase in anxiety level. This conclusion allowed the rejection of null hypotheses 12 and 13, with an α - level = .01.

Null hypothesis 12 states:

> There is no significant difference in student anxiety levels in a neutral situation and before writing an examination in a closed-book setting.

Null hypothesis 13 states:

> There is no significant difference in student anxiety levels in a neutral situation and before writing an examination in an open-book setting.

Table 20 showed there is a significant difference in anxiety level between the two settings. It has already been shown that a testing situation produced a significant increase in anxiety over the neutral situation. The means given earlier indicated that the open setting had a lower anxiety level than the closed setting. The F-ratio calculated was significant.

Thus, Null hypothesis 11:

> There is no significant difference in student anxiety levels before the examination in an open-book setting as compared with a closed-book setting.

was rejected, with an α - level = .05.

ANALYSIS OF ATTITUDE SCALES

Two different attitude measures were obtained during the study. Immediately after the students had written their open-book examination, and again after they had written their closed-book examination,

they were asked to record their feelings toward the testing situation on the Questionnaire on Attitudes to Testing. The score on this questionnaire was a measure of the student's attitude toward the kind of mathematics test he had just written. The second type of attitude measure obtained was the Attitude to Mathematics Opinionnaire. This opinionnaire was administered in a neutral setting between the two testing days. A more detailed account of the administration of both measures is given in Chapter 3.

Each of the attitude measures are considered separately first. Later in this section the analysis includes relationships that exist between the different attitude measures.

The two Questionnaires on Attitudes to Testing were compared for the open and closed-book examination. The comparison of the two settings is indicated in Table 21.

TABLE 21

Comparison of Open and Closed-Book Attitudes to Testing

n = 452

Source of Variation	SS	DF	MS	F-ratio
Between People	280,911.0	451	622.862	
Within People	33,313.0	452	73.701	
Treatment	28.000	1	28.000	0.3794
Residual	33,285.0	451	73.803	
Total	314,224.0	903		

The F-ratio is very small and not significant. Thus, null hypothesis 15

> There is no significant difference in the student's attitude toward writing examinations in an open-book or a closed-book setting.

was not rejected. It appears that the student who has a favourable attitude to one test setting also has a favourable attitude to the other test setting.

RELATIONSHIP OF ACHIEVEMENT TO ATTITUDE AND ANXIETY VARIABLES

A series of regression equations was structured for each of the sub-tests and total test for Form A and B in the two settings - open and closed. The purpose of these equations was to determine if some or all of the attitude and anxiety variables could be used to predict level of achievement. Another direct result of this investigation would be to show which variables have a significant influence on achievement and thus provide information to use when considering null hypotheses 14, 16 and 17.

Null hypothesis 14 states:

> There is no significant relationship between the level of anxiety and student achievement on either the open or closed-book settings for examinations.

Null hypothesis 16 states:

> There is no significant relationship between the student's attitude toward mathematics and achievement in mathematics with respect to the two different examination settings.

Null hypothesis 17 states:

> There is no significant relationship between the student's attitude toward the testing situation and his achievement in mathematics.

Regression equations for Form A are considered first. The open-setting is inspected first, and then the closed-setting. All the relevant information is reported in the form of tables. A discussion of the findings follows.

MULTIPLE REGRESSION ANALYSIS ON FORM A, OPEN-BOOK

Neutral Anxiety (Var. 1), Open Anxiety (Var. 2), Open Attitude (Var. 3) and Mathematics Attitude (Var. 4) are used to predict achievement on (a) knowledge sub-test (Var. 5)

(b) comprehension sub-test (Var. 6)

(c) application sub-test (Var. 7)

(d) total test (Var. 8)

A correlation matrix between the eight variables is given below.

TABLE 22

Correlation Matrix of Anxiety, Attitude and Achievement Values for Open-Book Examinations, Form A

n = 175

	1	2	3	4	5	6	7	8
1	1.000	0.341	-0.118	-0.144	-0.066	-0.069	-0.055	-0.076
2		1.000	-0.352	-0.127	-0.071	-0.139	-0.098	-0.132
3			1.000	0.464	0.211	0.371	0.439	0.459
4				1.000	0.314	0.373	0.426	0.474
5					1.000	0.414	0.267	0.548
6						1.000	0.581	0.861
7							1.000	0.883
8								1.000

A brief survey of the correlation matrix shows that neither of the two anxiety values were significantly related to the four achievement scores. All correlation values were approximately zero, indicating a random relationship. The attitude values appear to be significantly related to the four achievement scores. The relationships between the anxiety values and attitude values appear to be merely chance, as the correlation values were approximately zero. The near zero correlations confirm findings given earlier, when null hypothesis 15 was not rejected.

Regression equations for these variables were developed. First, a series of equations were developed for each sub-test (Knowledge, Comprehension, Application)* and then for the total test. The regression equations for Form A (total test) under an open setting are given below. They show which variables can be used to predict achievement on an examination written in an open-book setting.

Form A, Total Test, Open Book

Step No. 1

Variable entering	4 (Math Attitude)
F value for variable entering	50.236
Probability level for variable entering	0.000
Percent variance accounted for	22.504
Standard error of predict y	4.918

Achievement = .47 (Math Attitude) + 12.561

*The regression equations for each of the sub-tests under the two settings - open and closed may be found in Appendix two. These equations show in detail the prediction relationship of each variable to each of the sub-tests.

Step No. 2

Variable entering	3 (Open Attitude)
F value for variable entering	17.768
Probability level for variable entering	0.000
Percent variance accounted for	29.759
Standard error of predicted Y	4.697

Regression equation:

Achievement = .30 (Open Attitude) + .33 (Math Attitude) + 7.895

Variables three and four account for over 90% of the accounted variance. The other two variables have no relationship to achievement on the total test.

In the open-book setting, it appears attitude to the setting used and to the subject content has a significant effect on student achievement. Anxiety levels do not appear to have any effect. Form A is now considered in a closed setting to determine if the same relationships exist.

MULTIPLE REGRESSION ANALYSIS ON FORM A, CLOSED-BOOK

Neutral Anxiety (Var. 1), Closed Anxiety (Var. 2), Closed Attitude (Var. 3) and Mathematics Attitude (Var. 4) are used to predict achievement on

(a) knowledge sub-test (Var. 5)

(b) comprehension sub-test (Var. 6)

(c) application sub-test (Var. 7)

(d) total test (Var. 8)

A correlation matrix between the eight variables is given below.

TABLE 23

Correlation Matrix of Anxiety, Attitude and Achievement Values for Closed-Book Examinations, Form A

n = 112

	1	2	3	4	5	6	7	8
1	1.000	0.295	-0.104	0.020	-0.072	0.179	0.129	0.137
2		1.000	-0.362	-0.097	-0.252	-0.033	-0.108	-0.131
3			1.000	0.401	0.109	0.238	0.271	0.283
4				1.000	-0.023	0.224	0.240	0.229
5					1.000	0.216	0.297	0.489
6						1.000	0.624	0.857
7							1.000	0.901
8								1.000

A brief survey of the correlation matrix shows that neither of the two anxiety values are significantly related to the four achievement scores. All correlation values were approximately zero, indicating a random relationship. The attitude values appear to be significantly related to the dour achievement scores. The correlation coefficients, however, were not as high as in the open-book setting. The relationships between the anxiety values and attitude values appear to be merely chance, as the correlation values are approximately zero. The near zero correlations confirm findings given earlier, when null hypothesis 15 was rejected. The regression equations for these variables for the total test (Form A - closed) are now considered.

Form A, Total Test, Closed Book

Step No. 1

Variable entering	3(Closed Attitude)
F value for variable entering	9.566
Probability level for variable entering	0.003
Percent variance accounted for	8.001
Standard error of predicted Y	5.227

Regression equation:

Achievement = .28 (Closed Attitude) + 13.784

According to this sequence of regression equations only one variable, Closed Attitude, has a significant relationship with achievement on the total test. This variable accounts for approximately 2/3 of the accounted variance. The other three variables have little relationship to achievement on the total test.

In the closed-book setting, it appears attitude to the setting is the variable that consistently has a significant effect on student achievement. The relationship of anxiety levels and attitude of mathematics to achievement varies from sub-test to sub-test. In general, the relationships given in these closed-book regression equations are weak, accounting for only a small amount of variance.

The regression equations referring to the total test for Form B in both the open and closed setting are given in Appendix 3. The same prediction relationships seem to exist for Form B as existed for Form A.

COMPARISON OF RESULTS ON RELATIONSHIP OF ACHIEVEMENT TO ATTITUDE AND ANXIETY VARIABLES

Four cases were considered for prediction of achievement from attitude and anxiety variables. The four cases are summarized in Table 24. The percentage of variance accounted for by the prediction equations and the variables that are consistently significant are given.

TABLE 24

Comparison of Prediction Equations

	Form A (Open)	Form A (Closed)	Form B (Open)	Form B (Closed)
Percent Variance of Total Test	29.792%	12.866%	16.182%	23.890%
Significant Variables on Total Test $\alpha = .05$	Math Attitude Open Attitude	Closed Attitude	Math Attitude	Math Attitude

From the percentage of accounted variance for each test form, it can be seen that these prediction equations would not be a very accurate tool to use. Other variables must be considered that have a stronger relationship with achievement in the classroom if an accurate prediction equation is to be formed. However, these regression equations are useful in determining which variables have a significant relationship with achievement both on the sub-tests and total tests.

The anxiety variables, neutral anxiety, open anxiety and closed anxiety, do not have a significant effect on any achievement scores. The few times one of the anxiety variables appeared to be significant the general status of the sequence of regression equations was so poor that the results must be placed in serious doubt. Null hypothesis 14 was not rejected.

The attitude variables, open attitude, closed attitude and math attitude, appear to have a significant effect on achievement in three of the four cases ($\alpha = .05$). The one case, Form A (closed) where mathematics attitude is not significant, the total set of regression equations are very weak. None of the relationships in that case may be very significant. Likely, other factors have affected this sample as it appears different from the other three cases. In general, the results seem to indicate that null hypothesis 16 should be rejected. The correlation matrices confirm this decision. Setting attitude, open and closed, has a significant effect on achievement scores half of the time ($\alpha = .05$). This occurs for Form A (open) and Form A (closed). Approximately 2/3 of the correlation coefficients confirm the significant effect on achievement ($\alpha = .05$). These facts indicate that null hypothesis 17 should be tentatively rejected. Further work is needed to determine more precisely the relationship of attitude to setting and achievement.

CHAPTER 5: DISCUSSION OF RESULTS AND SUMMARY

In previous chapters the hypotheses have been stated, the methodology has been discussed and the results have been presented. In this chapter the results are interpreted in the light of the theory and research reviewed in Chapter 2 where it is possible. Where no relevant literature can be brought to bear on the issue, the writer attempts to formulate her own explanation of the results. The purpose of the present research was to explore as many relevant relationships as possible. It is hoped that the new ideas and theories presented in this study will lead to further study and refinement of the issues involved in open-book examinations. In addition to the interpretation of the results, some suggestions for use of the results and further research are given.

SUMMARY

The purposes of this study were

(a) to determine the effects of open- and closed-book test settings on achievement test performance;

(b) to identify any anxiety or attitude levels that differed between the two settings, and thus affected test performance;

(c) to determine the effect of open- and closed-book settings on test variance, reliability, and validity.

Twenty null hypotheses were formed and tested in the study.

The students participating in this study were grade XII Mathematics 30 students located in over 25 classrooms throughout the province of Alberta. (N = 670) Classroom lots of students were randomly assigned to one of four groups.

The following data were collected for each student:

(a) open-book test score

(b) closed-book test score

(c) neutral, open and closed anxiety scores

(d) open and closed attitude scores to testing

(e) attitude to mathematics score

These data were analyzed by means of three-way analysis of variance, item analysis, chi-square, repeated measures, correlations and regression equations depending on the factors involved.

Both of the tests that the student wrote were initially parallel and contained 36 items. The thought levels represented in the items were knowledge, comprehension and application. The classification was made according to Bloom's taxonomy. The content of each test covered the total Mathematics 30 course.

The anxiety scales were administered in a neutral setting and then immediately prior to each of the test settings. The Anxiety Differential was described as a measure of specific test anxiety.

The attitude to testing scales were administered immediately after each of the test settings. The attitude to mathematics scale was administered in a neutral setting between the two testing sessions.

Both scales, Attitude to Mathematics Opinionnaire and Questionnaire on Attitudes to Testing, attempt to measure attitudes to a specific situation or subject.

The primary results of the study were the following:

(a) Achievement scores were significantly higher on open-book examinations for knowledge item, comprehension item and total test item scores. No significant differences were found for application item scores.

(b) Little difference existed between the values for test variances, reliabilities and validities between the two settings.

(c) The students were significantly more anxious in either the open- or closed-setting than in the neutral setting. They were significantly less anxious in the open-book setting than in the closed-book setting.

(d) There was no significant difference between attitude to open-book testing and closed-book testing.

(e) There was a significant relationship between achievement and attitude values but not between anxiety values and achievement.

INTERPRETATION OF THE RESULTS

The results of the study, summarized in the last section, are now interpreted. In this section they are discussed in the order they were presented in the last chapter. Each topic is reviewed briefly before the related results are interpreted.

ORIGINS OF FORM A AND B

Form A and B were developed from a reliable (KR - 20 = .898) grade XII Mathematics 30 Departmental Examination. The forms were developed as parallel papers. The criteria for parallelism and validity for each paper has been given in Chapter two, three, and four. In addition each form had a correlation of nearly one with the original examination. This well established parallelism allowed various comparisons between the two settings-open and closed-book to be made.

RELATIONSHIP OF ACHIEVEMENT TO GENERAL MEAN, TIME OF ADMINISTRATION, SETTING AND FORM

Comparison of General Means

The general means of the original examination and the combined Form A and B were compared. The general means of combined Form A and B were lower than the original examination's general means for each of the sub-tests.

These findings provide information on the differences that exist between writing a series of test items in a final examination situation and a pre-test situation. It is surprising that the difference was not greater than it was, considering the supposed effects of review and different testing environments. The null hypothesis of no significant difference was rejected in three of the four cases - Comprehension sub-test, Application sub-test and Total test.

Comparison of Settings

The null hypothesis of no significant difference between settings was rejected in three of the four cases - Knowledge sub-test, Comprehension sub-test and Total test. This led to the rejection of null hypothesis 1 and the conclusion that there was a significant difference between the open- and closed-book set examination for the three cases above.

There was no significant difference between the Application sub-tests. A number of possible explanations could be given. First, the students may have found the questions too complex or unrelated to specific details in their notes or textbooks to receive significant help when they were writing the open-book examination within the time given. The application section may have been effectively closed-book under both settings. Second, the very nature of application questions indicates the student should not be able to find the answer - ready made - in his notes. The student is required to apply what he knows to a new situation. This finding indicates that tests which contain a large proportion of application, analysis and synthesis questions will likely not vary in difficulty level as a result of the setting in which they are written. However, the use of the open-book setting will permit more items to be constructed in this area than would otherwise be the case. The student will be able to have more facts available to use in solving such problems. Further investigation of the relationship of higher order items to the setting in which they

are written must be carried out. These types of questions are to receive the most emphasis and use in the future. Thus they must not be ignored when different item types are considered.

The rejection of null hypothesis 1 with respect to the other three cases was not in exact agreement with previous research. Stalnaker and Stalnaker (1935) found no difference in mean achievement on open- and closed-book tests. Kalish (1958) found there was no significant difference in the number of errors per examination in the open- or closed-book groups. Marco (1966) found that tests written in an open-book setting appeared to have slightly higher overall means on both the knowledge and application sub-tests. According to his classification scheme, the application sub-test contained comprehension and application items - the latter classification being based on Bloom's taxonomy. However, his results were not generally significant. Differences were significant for two of the four knowledge sub-tests and for one of the four application sub-tests. Marco blamed the lack of significant results on the type of test item he used - objective multiple choice and the lack of time students had to make the most use of the open-book setting. He felt the characteristics of both settings were too similar. The same conclusions might be applied to Kalish's test as his items were also multiple choice and his test was more speeded than Marco's test. Kalish gave an hour examination with 40 items. In both cases the tests effectively may have been closed-book tests even though given under the open-book test

setting in that students had no time to use available references.

In the present study multiple-choice items were again used. However, they seemed to cover a wider range of possible problems than in either Kalish's or Marco's study. One of the main reasons for their use was the repetition of existing open-book and closed-book testing procedures in the Department of Education. Only by investigating existing practises, can new improved practises be instituted. The finding of no significant difference between the application sub-tests shows that this type of item is not easier for the student who has materials available. If Marco had separated his Comprehension and Application items he may have found the same result. The fact setting was important for the Knowledge and Comprehension sub-tests in this study may be a function of the time the student had available while writing the test. The student had 60 minutes to complete a 36 item test in each setting. This additional time allowance likely permitted him to gain from the materials he had available in the open-setting. More work should be done on the relationship of time available to the effect of setting.

Comparison of Forms

The null hypothesis of no significant difference between forms was rejected in two of the four cases - Knowledge and Comprehension. However, this effect was not consistently in favour of one form, thus the comparison of total tests is equal. The differences on sub-test means likely occur as a result of the construction procedure of Form A

and B. The total difficulty level on the two forms was matched but difficulty levels for each sub-test were not matched perfectly. These differences might have been eliminated if the original examination had contained a larger number of items, thus allowing more ideal matching of item difficulties without losing any content validity. However, since the effect is not consistently on one form, one can conclude the two forms in total are equal and parallel, and use the results accordingly. No such direct comparisons of test forms have been made in this context previously. Also the same form in two different settings allows some interesting original comparisons to be made. These comparisons will follow later in this section.

Comparison of Time of Administration and Interaction Effects

The null hypothesis of no significant difference between times of administration was rejected in two of the four cases - Application and Total test. Generally the second administration took place four to five days after the first administration. Since these tests were conducted three and two weeks prior to the end of term when students would be writing their final mathematics examination and during that period the majority of teachers conducted review and both forms covered the same course content, it is not surprising to find the second time of writing slightly easier for the Application sub-test. Application items are by definition more complicated and thus the student would likely improve most in this area during a review

period. As indicated earlier, this sub-test was the largest part of each form - 16 items on Form A and 17 on Form B out of 36 items. Thus the score students received on this sub-test influenced their final score on the total test, making the total test scores significantly different for different administration times. This effect should not influence the total design, however, since the sequence of open- and closed-set examinations was randomly ordered over the two times of administration.

The interaction effects were not rejected in 11 of the 16 cases. The only significant rejections occurred with the interaction of form and time of administration. Possible reasons for differences resulting in means as a result of time of administration have already been given. The random assignment of classes of students in this study should counterbalance any over-all effect this might cause. Prior to this study no effective investigation of these interaction effects had been made.

Conclusions

The results of this section lead to the rejection of no significant difference between settings for the knowledge sub-test, comprehension sub-test and total test. Other possible variables that might have had a significant effect on the study were investigated, discussed and discounted. The results of this analysis permit the interchangeable use of forms in the rest of this study.

MEAN, VARIANCE, RELIABILITY AND VALIDITY OF FORM A AND B

The comparison of means for each of the total test and sub-tests supported the findings discussed earlier. Means for either form administered under the same setting were very close. This permits the assumption that the four groups of students participating in the study were equally representative of the population. The means showed a consistent difference in setting between open- and closed-book for each form, the open-book means generally being higher.

Considering the variance values the following conclusions can be made.

(a) Forms written in different settings did not have a significant difference in variance.

(b) Forms written in the same setting did not have a significant difference in variance.

(c) Sub-tests written in different settings did not have a significant difference in variance.

(d) Sub-tests written in the same setting did not have a significant difference in variance.

These findings are not in complete agreement with Marco's study. By the use of correlations he was able to show that variance increased for the open-book setting when certain test construction assumptions were met. A discussion of these assumptions is given in the literature of this study. The work of Gulliksen (1945) and Swineford (1959) showed that raw score variance increases as the average item standard deviation increases and as the variance of item difficulties decreases, Marco shows this to be true for his data, but the correlation values he

cites are low. A possible reason why the variance relationships cited by Marco were not duplicated in this study was the difference that existed in test means and the spread of item difficulties. In this study the means were significantly higher than in Marco's study and the spread of item difficulties was more divergent. If maximum variance is to be achieved, specifications for item construction must be rigorously defined. The actual findings in this section did not allow the rejection of no significant difference in variance.

The null hypothesis of no significant difference between reliabilities on the total tests and sub-tests was not rejected. The findings in this study were consistent with information given about reliability in the literature. (Swineford, 1959; Gulliksen, 1945; Zimmerman, 1967, 1968; Payne, 1968; Ebel, 1969). Reliabilities were higher on those sub-tests that met or nearly met the conditions specified in the literature for maximum reliability values. The open-book examination reliabilities, however, were not consistently higher than the closed-book examination reliabilities. This does not agree with Marco's work which suggested that reliabilities are consistently higher for open-book set tests. He based this conclusion on trends in his data that were not significant enough to reject his null hypothesis. A possible reason for the difference in findings may have been his small sample size and the low values of his 16 reliabilities, which ranged from - .26 to .61. The reliabilities in this study ranged from .20 to .80.

The validity of the two forms was considered under both settings. The content of the two forms has already been discussed and established. General validity has been discussed in the literature. The literature indicated the highest values of validity are found when item difficulties are close to .5, item intercorrelations are not high and the item variances (sum of variance of difficulty and variance of item intercorrelation) are not high, approximately .5. As item biserial correlations approached one, the item intercorrelations would be very high. It seems item intercorrelations would yield the highest validity as mean biserial correlations approached .5. The tables in Chapter 4 gave measures of mean item difficulties, variance of mean difficulties and mean biserial correlations of each sub-test for the four groups of students involved in the study. The results agree with the criteria established in the literature for high validity. (Thurstone, 1932; Tucker, 1946; Brogden, 1946; Cronback and Warrington, 1952; J. Horn, 1971). The results reported in those articles concerning tests composed of items with low precision, (that is, low item intercorrelations) are particularly appropriate to results of this study. Since validity scores are interrelated to reliability and variance measures there does not appear to be any significant difference in validity measures between the two settings. These findings again differ with Marco's results as he found small indications that the validities of open-book set tests were higher. However, his correlation values in some cases were so low that he doubted the

significance of his results. More work must be done in this area before general conclusions can be reached.

ANALYSIS OF INDIVIDUAL ITEMS

Individual items were compared descriptively and statistically. The difficulty levels and biserial correlations of items were compared under the two settings - open and closed. The descriptive comparison of student responses showed that the distribution of students over each of the alternatives did not greatly change between the two settings. There were some indications that the average student gained the most from the open-book setting. The difficulty values were compared using a chi-square test. The findings in the comparison supported the differences found between open- and closed-book achievement scores earlier in this study. The significant differences in favour of open-book set examinations occurred for the Knowledge sub-test and Comprehension sub-test. The comparison of item biserial correlation coefficients supported the results indicated when the reliability coefficients were compared. No definite pattern was established in item reliabilities between the two settings. A slight trend towards increased reliability in an open-book setting was noted for Knowledge sub-test items and Comprehension sub-test items.

ANALYSIS OF STUDENT ANXIETY SCORES

Two main results were noted when the student anxiety scores were analyzed.

1) Significant differences were found to exist between neutral, open and closed anxiety scores.

2) The students were most anxious in the closed-book setting, and least anxious in the neutral setting.

The rejection of null hypotheses of no significant differences between anxiety levels confirmed the trend noted by Marco in his study.

The findings that test anxiety is greater under the closed-book setting and is consistent with Feldhusen's findings that students reported feeling less anxious under the open-book test setting (1961). Marco confirmed the trend noted by Feldhusen in his study. Marco was able to show that students seemed less anxious in an open-book setting with the same measurement instrument, $.10 < p < .15$ for F - ratios computed. It is also consistent with what many people claim to be one of the advantages of the open-book examinations, a less tense writing situation. One problem with this type of anxiety measure is that it is obtained prior to the test administration. The difference in anxiety levels might be higher if the anxiety score represented the internal anxiety level during the examination. A more sensitive instrument might increase the degree of significance of the change of open-closed anxiety in relationship to neutral anxiety. More detailed investigation of the relationship of anxiety level and other related variables is needed.

The affect of anxiety levels on achievement was studied by a series of correlations and sequence of regression equations. The

results of the regression equations and correlations in both settings showed that anxiety level was not significantly related to achievement. These findings confirmed work done with anxiety levels by Marco. He also found no relationship. These findings are in contrast to those suggested in the literature. It appears that the anxiety levels measured in a laboratory setting and those found in the actual classroom may not be identical. More factors than anxiety must be involved in successful completion of classroom tasks.

ANALYSIS OF STUDENT ATTITUDE SCORES

Students did not indicate a significant difference in attitude between the two settings in which they participated. This was an important result of the analysis of student attitude scores. In some previous studies students had reported they liked the open-book examination best. Yet in this situation the majority of students did not differentiate between the two settings. Since the attitude questionnaires were administered immediately after each test was written, the students should have been aware of the difference between the open- and closed-book settings. A significant difference might have occurred if the open setting had been more different from the closed setting, for example, if the open setting had involved a take home essay examination.

The affect of attitude levels on achievement was studied by a series of correlations and sequence of regression equations. Level of

mathematic attitude and attitude to testing, both open and closed, had a significant predictive relationship with achievement. If the student did not like the test setting of subject matter being tested, his achievement score was lower than the reverse case's score.

As indicated in the literature section of this study very little research has been compiled on the relationship of attitudes and achievement. The brief findings reported in the literature are confirmed by this study. Students achieve better in situations they like. More sensitive testing must be conducted to determine if different test settings influence their achievement. According to this study the setting had no influence on the attitude to testing held by the majority of students. This does not support some of the earlier discussions on the topic that felt students liked the open-book setting best (Feldhusen, 1961; Tussing, 1951).

These results confirm trends and strengthen conclusions previously reported. It is hoped these results will increase the sensitivity and accuracy of the evaluation process.

SUGGESTIONS FOR USE OF THE RESULTS

The advantages and disadvantages of open- and closed-book testing have been examined in this study. In view of the findings of this study the following recommendations can be made.

Tests that are composed of mainly higher thought level items can be administered in either setting. The level of student

achievement will not change. Since open-book examinations are not easier than closed-book examinations they can be used in their place when other objectives call for them. For example, if a group of students have a positive attitude toward open-book examinations, these examinations can be used without concern over the measurements of student progress attained as a result.

A close relationship between student attitude to the subject, to the testing mode, and achievement in the subject was noted. This indicates the importance of attitude in the successful completion of a task. Measurement of student attitudes should be made to determine if maximum results are being produced in a subject area during the school year.

The varied responses in attitudes to testing indicates that different individuals achieve best under many kinds of evaluation modes. With the increase in individualized programs, different ways of evaluating a unit of work should be open to the student. This study has compared two forms of evaluation that can be used in many situations interchangeably.

These suggestions for use of the results have considered the major findings of the study. More detailed suggestions have been included in the previous sections in this chapter.

SUGGESTIONS FOR FURTHER STUDY

The previous section contained a summary and interpretation of results of this study. Some additional questions that now need to be

answered include the following.

1. What is effect of different item formats (essay, take-home, multiple-choice, etc.) on achievement under the two test settings?

2. What precise relationship exists between test anxiety and achievement on open- and closed-book examinations?

3. What relationship exists between various types of physical settings and achievement on open- and closed-book examinations?

4. What relationship exists between attitudes to different test settings and achievement on tests in these settings?

5. What further relationships exist to account for changes in test variance, reliability and validity in the two settings?

In addition to answering these questions concerned with open- and closed-book examinations more work must be done with the other types of evaluation. Similar studies could be done with each type. Only when all the information is available can the correct choices of measurement tools be made.

REFERENCES

Atkinson, J.W., An Introduction to Motivation, New York, D. Van Nostrand Company, Inc., 1964.

Avital, S.M., Shettleworth S.J., Objectives for Mathematics Learning, The Ontario Institute for Studies in Education, 1968.

Biggs, J.B., "Attitudes to Arithmetic - Number Anxiety", Educational Research, 1959, 2, 9.

Bloom, B.S., Taxonomy of Educational Objectives, New York, David McKay Company, Inc., 1956.

Brodgen, H.E., "Variation in test validity with variation in the distribution of items difficulties, number of items, and degree of their intercorrelation", Psychometrika, 1946, 11, 197-214.

Chalmers, J.W., Teachers of the Foothill Province, The Alberta Teachers' Association, University of Toronto Press, 1968.

Corcoran, M., and Gibb, E.G., "Appraising Attitudes in the Learning of Mathematics", The Nation Council of Teachers of Mathematics: Evaluation in Mathematics, 1961, 26, 106-121.

Cowley, W.H. "The Open-book Examinations", Journal of Higher Education, 1934, 4, 399.

Department of Education, High School and University Matriculation Board Minutes, Edmonton, Alberta, Department of Education, 1966.

Dutton, W.H., "Attitudes of Junior High School Pupils, School Review, 1964, 72, 18-22.

Ebel, R.L., Measuring Education Achievement, New Jersey, Prentice Hall, Englewood Cliffs, 1965.

Ebel, R.L., "Prospects for Evaluation of Learning", The Bulletin of the National Association of Secondary Principals, 1968, No. 332, 32-43.

Ebel, R.L., "Expected reliability as a function of choices per item", Educational and Psychological Measurement, 1969, 29, 565-570.

Feldhusen, J.F., "An Evaluation of College Students Reactions to Open-Book Examinations", Educational and Psychological Measurement, 1961, 21, 637-645.

Fouch, R.S., "Overview and Practical Interpretations", The National Council of Teachers of Mathematics: Evaluation in Mathematics, 1961, 26, 167-180.

Freeman, W.A., An Enquiry into the Attitudes of Secondary School Boys and Girls Towards Arithmetic, Unpublished Master's Thesis, University of Birmingham, Birmingham, 1948.

Furst, E.J., Constructing Evaluation Instruments, New York, Longmans, Green, 1958.

Gulliksen H., "The Relation of item difficulty and inter-item correlation to test variance and reliability", Psychometika, 1945, 10, 79-91.

Gulliksen, H., Theory of Mental Tests, New York: Wiley, 1950.

Hartung, M.L., "Basic Principles of Evaluation", The National Council of Teachers of Mathematics: Evaluation in Mathematics, 1961, 26, 21-42.

Horn, J., "Integration of Concepts of Reliability and Standard Error of Measurement, KR-20 and KR-21", Educational and Psychological Measurement, 1971, 31, 57-47.

Husek, T.R., Alexander, S., "The Effectiveness of and Anxiety Differential in Examination Stress Situations", Educational and Psychological Measurement, 1963, 23, 309-318.

Jones, P.S., (Editor), A History of Mathematics Education, The National Council of Teachers of Mathematics, 1970.

Kaiser, H.F., Carter, H.D., "Geometric Representation of the Notions of the Reliability Relevance and Validity", California Journal of Educational Research, 1971, 22, 122-125.

Kalish, R.A., "An Experimental Evaluation of the Open-book Examination" Journal of Educational Psychology, 1958, 49, 220-240.

Lord, F.M., "The Relation of the Reliability of Multiple-Choice Test to the Distribution of Item Difficulties", Psychometrika, 1952, 17, 181-194.

Marbler, C., Smallenburg, H., "Effects of Testing Programs on the Attitudes of Students, Teachers, Parents, and Community", Yearbook of the National Society for the Study of Education, 1963, 62, 103-125.

MacArthur, R.S., Hunka, S., School Examination Practises and Standards in Alberta, University of Alberta, 1960.

Marco, G.L., Psychological and Psychometric Correlates of Open and Closed Book Achievement Test Modes, Unpublished Doctoral Dissertation, University of Illinois, 1966.

McCandless, B., Castenda, A., "Anxiety in children School Achievement, and intelligence," Child Development, 1956, 27, 297-382.

Mortlock, R.S., Provision for Individual Differences in Eleventh Grade Mathematics Using Flexible Grouping Based on Achievement of Behavorial Objectives, Unpublished Doctoral Dissertation, The University of Michigan, 1969.

Nunnally, J.C., Tests and Measurement, Toronto, McGraw-Hill, 1959.

Osgood, C.E., Suci, G.J., Tannenbausm, P.H., The Measurement of Meaning, Urlana, Illinois: University of Illinois Press, 1957.

Payne, W.H., Anderson, D.E., "Significance levels for the Kuder-Richardson Twenty", Educational and Psychological Measurement, 1968, 28, 23-39.

Phillips, B.N., "School Stress as a Factor in Children's Responses to Tests and Testing", Journal of Educational Measurement, 1971, 8, 21-26.

Pritchard, R., "The Relative Popularity of Secondary School Sugjects At Various Ages:, British Journal of Education Psychology, 1935, 5-6, 157-179.

Reese, H.W., "Manifest Anxiety, Achievement, and Test Performance", Journal of Education Psychology, 1961, 52, 132-135.

Remai, H.A., An Experimental Investigation Comparing Attitudes Toward Mathematics of Modern and Traditional Mathematics Students at the Junior High School Level, Unpublished Master's Thesis, University of Alberta, Edmonton, 1965.

Remmers, H.H., Gage, N.L., Rummel, J.F., A Practical Introduction to Measurement and Evaluation, New York, Harper & Row, 1960.

Sharon, A.T., "Measurement of College Achievement by the College-Level Examination Program", Educational and Psychological Measurement, 1971, 31, 477-484.

Snyder, A., "The Open-Book Vocabulary Test", College English, 1952, 13, 458-460.

Stalnaker, J.M., Stalnaker, R.C., "Open-Book Examinations: Results", With the Technicians, 1935, 5, 214-216.

Sueltz, B.A., "The Role of Evaluation in the Classroom", The National Council of Teachers of Mathematics: Evaluation in Mathematics, 1961.

Swineford, F., "Some Relations Between Test Scores and Item Statics", Journal of Education Psychology, 1959, 50.

Taylor, P.A., Cowley, D.M., "New Dimensions of Evaluation", Readings in Curriculum, Dubuque, Iowa, W.C. Brown, 1972, 1-10.

Thurstone, T.G., "The difficulty of a test and its diagnostic value," Journal of Educational Psychology, 1932, 23, 335-343.

Tucker, L.R., "Maximum validity of a test with equivalent items", Psychometrika, 1946, 11, 1-13.

Tussing, Lyle, "A Consideration of the Open-Book Examination", Educational and Psychological Measurement, 1951, 11, 597-602.

Woffard, J.C., "Evaluation, An Unfinished Stretch of the Education Highway", Bulletin, Canadian Council for Research in Education, 1969, 1-40.

Zimmerman, D.W., Williams, R.H., Burkheimer, G.J., "Dependence of test reliabilities upon heterogeneity of individual and group score distributions", Educational and Psychological Measurement, 1968, 28, 45.

Zimmerman, D.W., William, R.H., Burkheimer, G.J., "Maximum reliability of a multiple choice test as a function of number of items, number of choices and group heterogeneity", Journal of Experimental Education, 1967, 36, 89-94.

APPENDIX 1

Name _____ Date _____

Semantic Differential

The purpose of this study is to measure the meaning of certain words to various people by having them judge them against a series of descriptive scales. In taking this test, please make your judgments on the basis of what these words mean to you. In the left-hand column of the next page you will find different concepts to be judged and to the right of them a set of scales. You are to rate a concept on the scale to the right of that concept.

Here is how you are to use these scales: If you feel that the concept at the left is very closely related to one end of the scale, you should place your checkmark as follows:

FATHER: fair x : ____: ____: ____: ____: ____: ____ unfair
FATHER: fair ____ : ____: ____:or____: ____: ____: x unfair

If you feel that the concept is quite closely related to one or the other end of the scale (but not extremely), you should place your checkmark as follows:

FATHER: strong ____: x : ____: ____: ____: ____: ____ weak
FATHER: strong ____: ____: ____:or____: ____: x : ____ weak

If the concept seems only slightly related to one side as opposed to the other (but not really neutral), then you should check as follows:

FATHER: active ____: ____: x : ____: ____: ____: ____ passive
FATHER: active ____: ____: or: ____: x : ____: ____ passive

The direction toward which you check, of course, depends upon which of the two ends of the scale seem most characteristic of the thing you are judging.

If you consider the concept to be neutral on the scale, both sides of the scale equally associated with the concept; or if the scale is completely irrelevant, unrelated to the concept, then you should place your checkmark in the middle space:

FATHER: safe ____: ____: ____: x : ____: ____: ____ dangerous

IMPORTANT:
(1) Place your checkmark in the middle of the space, not on the boundaries.

This Not this
____: ____: x : ____: ____x____: ____

(2) Be sure you check every scale; do not omit any.
(3) Never put more than one checkmark on a single scale.

FIG 1

Sometimes you may feel as though you've had the same item before on the test. This will not be the case, so do not look back and forth through the items. Do not try to remember how you checked similar items earlier in the test. Make each item a separate and independent judgment. Work at fairly high speed through the test. Do not worry or puzzle over individual items. It is your first impressions, the immediate "feeling" about the items, the way you feel about the concepts at this moment, that we want. On the other hand, please do not be careless, because we want your true impressions.

SCALES

ME:	helpless	_____:_____:_____:_____:_____:_____:_____	secure
CLASSROOM:	hot	_____:_____:_____:_____:_____:_____:_____	cold
BREATHING:	tight	_____:_____:_____:_____:_____:_____:_____	loose
SEAT:	hard	_____:_____:_____:_____:_____:_____:_____	soft
SCREW:	strong	_____:_____:_____:_____:_____:_____:_____	weak
HANDS:	wet	_____:_____:_____:_____:_____:_____:_____	dry
TODAY:	loose	_____:_____:_____:_____:_____:_____:_____	tight
BLACKBOARD:	constricted	_____:_____:_____:_____:_____:_____:_____	spacious
ME:	frightened	_____:_____:_____:_____:_____:_____:_____	fearless
GERMS:	deep	_____:_____:_____:_____:_____:_____:_____	shallow
INSTRUCTOR:	serious	_____:_____:_____:_____:_____:_____:_____	humorous
HANDS:	good	_____:_____:_____:_____:_____:_____:_____	bad
BREATHING:	careful	_____:_____:_____:_____:_____:_____:_____	carefree
CLASSROOM:	large	_____:_____:_____:_____:_____:_____:_____	small
FINGERS:	stiff	_____:_____:_____:_____:_____:_____:_____	relaxed
ME:	calm	_____:_____:_____:_____:_____:_____:_____	jittery
TEXTBOOK:	good	_____:_____:_____:_____:_____:_____:_____	bad
SCREW	loose	_____:_____:_____:_____:_____:_____:_____	tight
WINDOW:	opaque	_____:_____:_____:_____:_____:_____:_____	transparent
ME:	carefree	_____:_____:_____:_____:_____:_____:_____	worried
ASSIGNMENTS:	long	_____:_____:_____:_____:_____:_____:_____	short
ANXIETY:	clear	_____:_____:_____:_____:_____:_____:_____	hazy

QUESTIONNAIRE ON ATTITUDES TO TESTING

NAME: _____ DATE: _____

This questionnaire is designed to give you an opportunity to indicate how and what you feel in regard to mathematics test.

One of the main reasons for construction of this questionnaire is that very little is known about people's feelings toward taking various kinds of tests. We can assume that people differ in the degree to which they are affected by taking a test. What we are particularly interested in here is how widely people differ in their opinions of and reactions to testing situations.

The value of this questionnaire will in large part depend on how frank you are in stating your opinions, feelings and attitudes. Your answers will be kept confidential.

For each question there is a line or scale on the ends of which are statements of opposing feelings or attitudes. In the middle of the line you will find either the word "midpoint" or a phrase, both of which are intended to reflect a feeling or attitude which is in-between the statements of opposing feelings described above. You are required to put a mark (X) on that point of the line you think best indicates the strength of your feeling or attitude about the particular question. The midpoint is only for your guidance. Do not hesitate to put the mark (X) on any point of the line as long as that mark reflects the strength of your feeling or attitude.

If you have any questions at this time please ask them.

THERE ARE NO CATCH QUESTIONS IN THIS QUESTIONNAIRE. PLEASE READ EACH QUESTION AND EACH SCALE VERY CAREFULLY. THERE IS NO TIME LIMIT.

THE MIDPOINT IS ONLY FOR YOUR GUIDANCE. DO NOT HESITATE TO PUT THE MARK (X) ON ANY POINT OF THE LINE AS LONG AS THAT MARK REFLECTS THE STRENGTH OF YOUR FEELING OR ATTITUDE.

The following questions relate to your attitude toward and experience with mathematics tests. More specifically we are concerned with the attitude you have toward the kind of mathematics test you have just written. Please try to remember how you usually reacted toward this type of test and how you felt while taking them.

FIG 2

1. How valuable do you think mathematics tests are in determining a person's ability?

very valuable	Valuable in some respects and valueless in others	valueless

2. Do you think that mathematics test should be used more widely than at present to grade students?

should be used less widely	should be used	should be used more widely

3. Would you be willing to stake your grade in a math course on the outcome of one mathematics test which has previously predicted success in a highly reliable fashion?

very willing	uncertain	not willing

4. If you know that you are going to take a math test, how do you feel beforehand?

feel very unconfident	midpoint	feel very confident

5. After you have taken a math test, how confident do you feel that you have done your best?

feel very unconfident	midpoint	feel very confident

THE MIDPOINT IS ONLY FOR YOUR GUIDANCE. DO NOT HESITATE TO PUT A MARK (X) ON ANY POINT ON THE LINE AS LONG AS THAT MARK REFLECTS THE STRENGTH OF YOUR FEELING OR ATTITUDE

6. When you are taking a mathematics test, to what extent do your emotional feelings interfere with or lower your performance?

do not interfere at all	midpoint	interfere a great deal

7. Before taking a mathematics test, to what extent are you aware of an "uneasy feeling"?

am very much aware of it	midpoint	am not aware of it at all

8. While taking a mathematics test, to what extent do you experience an accelerated heart-beat?

heart-beat does not accelerate at all	midpoint	heart-beat noticeably accelerated

9. Before taking a mathematics test, to what extent do you experience an accelerated heart-beat?

heart-beat does not accelerate at all	midpoint	heart-beat noticeably accelerated

10. While taking a mathematics test, to what extent do you worry?

worry a lot	midpoint	worry not at all

11. Before taking a mathematics test, to what extent do you worry?

worry a lot	midpoint	worry not at all

12. While taking a mathematics test, to what extent do you perspire?

perspire not at all	midpoint	perspire a lot

13. Before taking a mathematics test, to what extent do you perspire?

perspire not at all	midpoint	perspire a lot

14. In comparison with other students, how often do you think of ways of avoiding a mathematics test?

less often than other students	midpoint	more often than other students

15. Do your emotional feelings interfere with your performance on a mathematics test more than on tests of similar importance in most other subjects?

interfere more on a mathematics test	midpoint	interfere less on a mathematics test

ATTITUDE TO MATHEMATICS OPINIONNAIRE

Name:_____ Date:_____

Directions:

Write your name and the date. Each of the statements on this opinionnaire expresses a feeling which a particular person has toward mathematics. You are to express on a five-point scale, the extent of agreement between the feeling expressed in each statement and your own personal feeling. The five points are:

Strongly Disagree	(SD)
Disagree	(D)
Undecided	(U)
Agree	(A)
Strongly Agree	(SA)

You are to circle the letter which best indicates how closely you agree or disagree with the feeling expressed to each statement as it concerns you.

1.	I do not like mathematics. I am always under a terrible strain in a math class.	SD	D	U	A	SA
2.	I do not like mathematics, and it scares me to have to take it.	SD	D	U	A	SA
3.	Mathematics is very interesting to me. I enjoy math courses.	SD	D	U	A	SA
4.	Mathematics is fascinating and fun.	SD	D	U	A	SA
5.	Mathematics makes me feel secure, and at the same time it is stimulating.	SD	D	U	A	SA
6.	I do not like mathematics. My mind goes blank and I am unable to think clearly when working math.	SD	D	U	A	SA
7.	I feel a sense of insecurity when attempting mathematics.	SD	D	U	A	SA

(Fig 3)

8.	Mathematics makes me feel uncomfortable, restless, irritable and impatient.	SD	D	U	A	SA
9.	The feeling I have toward mathematics is a good feeling.	SD	D	U	A	SA
10.	Mathematics makes me feel as though I'm lost in a jungle of numbers and can't find my way out.	SD	D	U	A	SA
11.	Mathematics is something I enjoy a great deal.	SD	D	U	A	SA
12.	When I hear the word math, I have a feeling of dislike.	SD	D	U	A	SA
13.	I approach math with a feeling of hesitation -- hesitation resulting from a fear of not being able to do math.	SD	D	U	A	SA
14.	I really like mathematics.	SD	D	U	A	SA
15.	Mathematics is a course in school which I have always liked and enjoyed studying.	SD	D	U	A	SA
16.	I don't like mathematics. It makes me nervous to even think about having to do a math problem.	SD	D	U	A	SA
17.	I have never liked math, and it is my most dreaded subject.	SD	D	U	A	SA
18.	I love mathematics. I am happier in a math class than in any other class.	SD	D	U	A	SA
19.	I feel at ease in mathematics; and I like it very much.	SD	D	U	A	SA
20.	I feel a definite positive reaction to mathematics; it's enjoyable.	SD	D	U	A	SA

DEPARTMENT OF EDUCATION

MATHEMATICS 30 EXAMINATION (FORM A)

All answers in this examination are to be machine scored.

Use the separate ANSWER SHEET and HB PENCIL.

Candidates are permitted to use slide rules and mathematical tables. Knott's Mathematical Tables will be supplied by the Presiding Examiner.

You have 55 minutes to complete 36 multiple-choice questions worth one mark each. Time yourself accordingly.

There will be no deduction for errors. Therefore, if you find a question difficult, make as intellignet a choice as possible and go on to the next one. Do not spend too much time on any one question. If there is time left over you may go back and check your answers.

Do not put any marks on this test booklet.
Do not bend or fold the separate answer sheet in any way.
BOOKLET, ANSWER SHEET and PENCIL must be returned at the end of the period.

(FIG 4)

DEPARTMENT OF EDUCATION

MATHEMATICS 30 EXAMINATION (FORM B)

All answers in this examination are to be machine scored.

Use the separate ANSWER SHEET and HB PENCIL.

Candidates are permitted to use slide rules and mathematical tables. Knott's Mathematical Tables will be supplied by the Presiding Examiner.

You have 55 minutes to complete 36 multiple-choice questions worth one mark each. Time yourself accordingly.

There will be no deduction for errors. Therefore, if you find a question difficult, make as intelligent a choice as possible and go on to the next one. Do not spend too much time on any one question. If there is time left over you may go back and check your answers.

Do not put any marks on this test booklet.
Do not bend or fold the separate answer sheet in any way.
BOOKLET, ANSWER SHEET and PENCIL must be returned at the end of the period.

(FIG 5)

SUBJECT Mathematics 30 Form A

EXAMINER _____

EXAM NO _____

BLUEPRINT #2

MAJOR TOPICS OF COURSE CONTENT

(use as many topics as is appropriate - probably 4 to 9)

MENTAL ACTIVITY (Cognitive Process or Level)

	Topic or Content Areas									Total Items	Emphasis
	I	II	III	IV	V	VI	VII	VIII	IX		
	Lang.	Function	Series	Complex Numbers	Quad. Relation	Probability					
	Item Numbers										
1.00 KNOWLEDGE To answer items at this level the student needs only to recognize or remember materials learned directly from textbooks or through classroom instruction		1			2	3,5,6	4			6	
2.00 COMPREHENSION At this level the student must make a simple transfer or generalization using well comprehended knowledge	12	7,10,11	8,9	13,14	15,16	18	17,19, 20			14	
3.00 APPLICATION At this level the student must solve a problem of transfer dealing with an unfamiliar situation, and the solution is generally a multi-step procedure		25,26	21,22,23 24	27,28	29,30 32	31,34 35,36	33			14	
4.00 ANALYSIS At this level the student does not have available a set of procedure or method of solution. He must be able to examine the material and derive his own relationships to solve the problem										2	
5.00 SYNTHESIS At this level the student must be able to put together given elements in an entirely new way to find the solution											
6.00 EVALUATION At this level the student must be able to form a judgment about the value of material and methods for given purposes											
AFFECTIVE DOMAIN WHERE APPLICABLE											
TOTAL NUMBER OF ITEMS	1	6	6	4	6	8	5			36	
EMPHASIS	2	17	17	11	17	22	14				100%

(FIG 6)

SUBJECT Mathematics 30 Form B

EXAMINER _____

EXAM NO _____

BLUEPRINT #2

MAJOR TOPICS OF COURSE CONTENT

(use as many topics as is appropriate - probably 4 to 9)

MENTAL ACTIVITY (Cognitive Process or Level)

	Topic or Content Areas									Total Items	Emphasis
	I	II	III	IV	V	VI	VII	VIII	IX		
	Lang.	Functions	Series	Complex Numbers	Quad. Relations	Probability					
	Item Numbers										
1.00 KNOWLEDGE To answer items at this level the student needs only to recognize or remember materials learned directly from textbooks or through classroom instruction			1		2	3,5	4			5	
2.00 COMPREHENSION At this level the student must make a simple transfer or generalization using well comprehended knowledge	9	6,10 11,12	7,8	13,14	15	16	17,18, 19			14	
3.00 APPLICATION At this level the student must solve a problem of transfer dealing with an unfamiliar situation, and the solution is generally a multi-step procedure	25	24	20,21, 22,23	26,34	27,28, 29,35	30,31, 32,33	36			14	
4.00 ANALYSIS At this level the student does not have available a set of procedure or method of solution. He must be able to examine the material and derive his own relationships to solve the problem										3	
5.00 SYNTHESIS At this level the student must be able to put together given elements in an entirely new way to find the solution											
6.00 EVALUATION At this level the student must be able to form a judgment about the value of material and methods for given purposes											
AFFECTIVE DOMAIN WHERE APPLICABLE											
TOTAL NUMBER OF ITEMS	2	5	7	4	6	7	5			36	
EMPHASIS	6	14	19	11	17	19	14				100%

(FIG 7)

SUBJECT Mathematics 30

EXAMINER

EXAM NO Test 1

BLUEPRINT #2

MAJOR TOPICS OF COURSE CONTENT

(use as many topics as is appropriate - probably 4 to 9)

MENTAL ACTIVITY (Cognitive Process or Level)

	Topic or Content Areas									Total Items	Emphasis
	I	II	III	IV	V	VI	VII	VIII	IX		
	Lang.	Function	Series	Complex Numbers	Quadr Relation	Arrange-ment	Prob.				
	Item Numbers										
1.00 KNOWLEDGE To answer items at this level the student needs only to recognize or remember materials learned directly from textbooks or through classroom instruction		2	1		3, 4	5,6,9 10,11	7, 8			11	15
2.00 COMPREHENSION At this level the student must make a simple transfer or generalization using well comprehended knowledge	18,24	12,13,19 20,21,22 23	14,15 16,17	25,26, 29	30, 31, 32	34, 35	33,36,37 38,39,40			28	40
3.00 APPLICATION At this level the student must solve a problem of transfer dealing with an unfamiliar situation, and the solution is generally a multi-step procedure	51	49, 50 52	41,42, 43,44, 45,46, 47,48	53,54, 55,66	56,57 58,59 60,67 68	61,62, 63,64, 65	69,70			25	35
4.00 ANALYSIS At this level the student does not have available a set of procedure or method of solution. He must be able to examine the material and derive his own relationships to solve the problem										5	10
5.00 SYNTHESIS At this level the student must be able to put together given elements in an entirely new way to find the solution											
6.00 EVALUATION At this level the student must be able to form a judgment about the value of material and methods for given purposes											
AFFECTIVE DOMAIN WHERE APPLICABLE											
TOTAL NUMBER OF ITEMS	3		13	9	12	12	10			69	
EMPHASIS	4	16	19	13	17	17	14				100%

(FIG 8)

DIVISION OF ORIGINAL EXAMINATION BY

TOPIC AREAS TO CONSTRUCT FORM A AND FORM B

ORIGINAL EXAMINATION CONTAINS THE FOLLOWING ITEMS:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70

FORM A CONTAINS THE FOLLOWING ITEMS:

2 3 5 7 9 11 13 14 16 20 22 24 25 27 30 32 33 35 38 39
41 43 45 47 49 52 53 55 57 59 62 64 67 69

FORM B CONTAINS THE FOLLOWING ITEMS:

1 4 6 8 10 12 15 17 18 19 21 23 26 29 31 34 36 37 40 42
44 46 48 50 51 54 56 58 60 61 63 65 66 68 70

MATHEMATICS 30 JANUARY 1971

VARIABLE	MEAN	STD. DEV.
1	21.665	5.841
2	22.042	6.054
3	43.713	11.399

	1	2	3
1	1.000	0.8461	0.9566
2	0.8461	1.000	0.9596
3	0.9566	0.9566	1.000

(FIG 9)

INFORMATION SHEET FOR MATHEMATICS 30 EXPERIMENTAL

TYPE OF EXAMINATION, TIME OF ADMINISTRATION AND ADMINISTRATOR

1. Semantic Differential Questionaire to be administered at any time during a regular Mathematics 30 period by your teacher.

2. An open-book examination to be administered on _____ by a Departmental representative.

3. A closed-book examination to be administered on _____ by a Departmental representative.

REASON FOR EXAMINATIONS

To obtain information about the effects of writing open-book examinations in mathematics.

CONTENT OF EXAMINATIONS

Both mathematics examinations cover the complete course. Student scores for these examinations will be sent to participating schools shortely after the testing date.

PREPARATION HINTS

When you are preparing for the closed-book examination, conduct a general review of the material covered in this course. Work sample questions as you review. Finally, study previous tests and determine why errors occured. On the day of the examination bring a copy of Knott's Mathematical Tables and a slide rule if you plan to use one.

When you are reviewing for the open-book examination prepare as you did for the closed-book examination. The following additional hints may be helpful. A student writing an open-book examination is permitted to bring slide rule, and any other helpful reference materials. Thus, you should make sure your notes are in order so you are able to locate concepts and facts you may wish to check. However, you should not depend on using your notes and test constantly as you write the examination. Texts, notes and reference materials should be used only as background material to clarify a particular fact,definition or method that may be forgotten or confused for the moment.

(Fig 10)

Important sections of your text can be marked for easy reference. Important terms constants, and formulas you may want to refer to can be listed on a sheet of paper. This will save time when you are writing the examination and need a particular reference.

We hope the above information will be helpful. Thank you in advance for participating in this project.

APPENDIX 2

The following section tests the effects of setting, form and time. Each sub-test is considered in turn for the following effects:

1. Mean - deviation of sub-test's mean from general mean
2. A - main effect of setting
3. B - main effect of form
4. AB - effect of form and setting interaction
5. C - main effect of time of administration
6. BC - effect of form and time interaction
7. AC - effect of setting and time interaction
8. ABC - effect of setting, form and time interaction

KNOWLEDGE SUB-TEST $\alpha = .05$

Source	SS	DF	MS	F-ratio	Prob	Decision
Mean	233.997	1	233.997	0.489	0.484	NS
A	19,299.500	1	19,299.500	40.385	0.000	SIG
B	5,953.140	1	5,953.140	12.457	0.000	SIG
AB	26.604	1	26.604	0.056	0.814	NS
C	183.084	1	183.084	0.383	0.536	NS
BC	1,751.980	1	1,751.980	3.666	0.056	NS
AC	446.972	1	446.972	0.935	0.334	NS
ABC	240.062	1	240.062	0.502	0.479	NS
Errors	567,735.000	1188	477.891			

COMPREHENSION SUB-TEST $\alpha = .05$

Source	SS	DF	MS	F-ratio	Prob	Decision
Mean	3,364.420	1	3,364.420	10.310	0.001	SIG
A	2,650.490	1	2,650.490	8.122	0.004	SIG
B	3,701.030	1	3,701.030	11.342	0.001	SIG
AB	426.455	1	426.455	1.307	0.253	NS
C	970.719	1	970.719	2.975	0.840	NS
BC	7,580.430	1	7,580.430	23.230	0.000	SIG
AC	2,899.240	1	2,899.240	8.885	0.003	SIG
ABC	37.679	1	37.679	.115	0.734	NS
Errors	387,659.000	1188	326.312			

APPLICATION SUB-TEST $\alpha = .05$

Source	SS	DF	MS	F-ratio	Prob	Decision
Mean	11,585.500	1	11,585.500	31.207	0.000	SIG
A	7.441	1	7.441	.020	0.887	NS
B	125.769	1	125.769	.339	0.560	NS
AB	973.951	1	973.951	2.623	0.106	NS
C	3,588.650	1	3,588.650	9.666	0.002	SIG
BC	10,471.400	1	10,471.400	28.206	0.000	SIG
AC	984.517	1	984.517	2.651	0.104	NS
ABC	108.021	1	108.021	.290	0.590	NS
Errors	441,046.000	1188	371.251			

TOTAL TEST

$\alpha = .05$

Source	SS	DF	MS	F-ratio	Prob	Decision
Mean	3,328.810	1	3,328.810	12.912	0.000	SIG
A	1,646.620	1	1,646.620	6.387	0.012	SIG
B	151.526	1	151.526	.588	0.443	NS
AB	467.365	1	467.365	1.813	0.178	NS
C	1,654.230	1	1,654,230	6.417	0.011	SIG
BC	7,474.820	1	7,474.820	28.993	0.000	SIG
AC	1,463.260	1	1,463.260	5.676	0.017	SIG
ABC	22.149	1	22.149	0.086	0.769	NS
Errors	306,275.000	1188	257.807			

APPENDIX 3

Form A, Knowledge Sub-Test, Open Book

Step No. 1

Variable entering	4(Mathematics Attitude)
F-value for variable entering	18.956
Probability level for variable entering	0.000
Percent variance accounted for	9.875
Standard error of predicted Y	1.130

Regression equation:

Achievement = .31 (Math Attitude) + 3.311

Step No. 2

Variable entering	3(Open Attitude)
F-value for variable entering	1.045
Probability level for variable entering	0.308
Percentage variance accounted for	10.419
Standard error of predicted Y	1.130

Regression equation:

Achievement = .08 (Open Attitude) + .28 (Math Attitude) + 3.039

Step No. 3

Variable entering	1(Neutral Anxiety)
F-value for variable entering	0.0534
Probability level for variable entering	0.817
Percentage variance accounted for	10.447
Standard error of predicted Y	1.133

Regression equation:

Achievement = -0.02 (Neutral Anxiety) + 0.08 (Open Attitude)
+ .27 (Math Attitude) + 3.149

Step No. 4

Variable entering	2(Open Anxiety)
F-value for variable entering	0.001
Probability level for variable entering	0.975
Percentage variance accounted for	10.448
Standard error of predicted Y	1.136

Regression equation:

Achievement = -0.02 (Neutral Anxiety) -0.00 (Open Anxiety) +
0.08 (Open Attitude) + .27 (Math Attitude) + 3.161

Form A, Comprehension Sub-Test, Open-Book

Step No. 1

Variable entering	4(Math Attitude)
F-value for variable entering	28.009
Probability level for variable entering	0.000
Percent variance accounted for	13.935
Standard error of predicted Y	2.288

Regression equation:

Achievement = .37 (Math Attitude) + 5.411

Step No. 2

Variable entering	3(Open Attitude)
F-value for variable entering	10.513
Probability level for variable entering	0.000
Percent variance accounted for	18.892
Standard error of predicted Y	2.228

Step No. 3

Variable entering	2(Open Anxiety)
F-Value for variable entering	0.078
Probability level for variable entering	0.780
Percent variance accounted for	18.929
Standard error of predicted Y	2.234

Regression equation:

Achievement = -0.02 (Open Anxiety) + 0.24 (Open Attitude) + (Math Attitude) + 3.996

Step No. 4

Variable entering	1(Neutral Anxiety)
F-value for variable entering	0.002
Probability level for variable entering	0.963
Percent variance accounted for	18.929
Standard error of predicted Y	2.240

Regression equation:

Achievement = 0.00 (Neutral Anxiety) -0.02 (Open Anxiety) + .24 (Open Attitude) + .26 (Math Attitude) + 3.967

Form A, Application Sub-Test, Open Book

Step No. 1

Variable entering	3(Open Attitude)
F-value for variable entering	41.342
Probability level for variable entering	0.000
Percent variance accounted for	19.288
Standard error of predicted Y	2.861

Regression equation:

Achievement = .44 (Open Attitude) + 2.477

Step No. 2

Variable entering	4(Math Attitude)
F-value for variable entering	14.511
Probability level for variable entering	0.000
Percent variance accounted for	25.567
Standard error of predicted Y	2.755

Regression equation:

Achievement = .31 (Open Attitude) + .28 (Math Attitude) + 1.147

Step No. 3

Variable entering	2(Open Anxiety)
F-value for variable entering	0.558
Probability level for variable entering	0.456
Percent variance accounted for	25.809
Standard error of predicted Y	2.759

Regression equation:

Achievement = .05 (Open Anxiety) + .33 (Open Attitude) + .28 (Math Attitude) + 0.200

Step No. 4

Variable entering	1(Neutral Anxiety)
F-value for variable entering	0.009
Probability level for variable entering	0.923
Percent variance accounted for	25.813
Standard error of predicted Y	2.767

Regression equation:

Achievement = .01 (Neutral Anxiety) + .05 (Open Anxiety) + .33 (Open Attitude) + .28 (Math Attitude) + 0.124

Form A, Knowledge Sub-Test, Closed Book

Step No. 1

Variable entering	2(Closed Anxiety)
F-value for variable entering	7.489
Probability level for variable entering	0.007
Percent variance accounted for	6.374
Standard error of predicted Y	1.197

Regression equation:

Achievement = -.25 (Closed Anxiety) + 5.244

Step No. 2

Variable entering	4(Math Attitude)
F-value for variable entering	0.199
Probability level for variable entering	0.657
Percent variance accounted for	6.780
Standard error of predicted Y	1.206

Regression equation:

Achievement = -.24 (Closed Anxiety) + .04 (Closed Attitude)
-.06 (Math Attitude) + 5.228

Step No. 3

Variable entering	1 (Neutral Anxiety)
F-value for variable entering	0.005
Probability level for variable entering	0.946
Percent variance accounted for	6.785
Standard error of predicted Y	1.211

Regression equation:

Achievement = .01 (Neutral Anxiety) -.24 (Closed Anxiety) +
.05 (Closed Attitude) -.07 (Math Attitude) +5.202

Form A, Comprehension Sub-Test, Closed Book

Step No. 1

Variable entering	3(Closed Attitude)
F-value for variable entering	6.577
Probability level for variable entering	0.012
Percent variance accounted for	5.642
Standard error of predicted Y	2.495

Regression equation:

Achievement = .24 (Closed Attitude) + 5.608

Step No. 2

Variable entering	1(Neutral Anxiety)
F value for variable entering	5.053
Probability level for variable entering	0.027
Percent variance accounted for	9.823
Standard error of predicted Y	2.451

Regression equation:

Achievement = .20 (Neutral Anxiety) + .25 (Closed Attitude)
+ 2.946

Step No. 3

Variable entering	4(Math Attitude)
F-value for variable entering	1.966
Probability level for variable entering	0.164
Percent variance accounted for	11.435
Standard error of predicted Y	2.440

Regression Equation:

Achievement = .20 (Neutral Anxiety) + .20 (Closed Attitude) + .14 (Math Attitude) + 2.304

Step No. 4

	2(Closed Anxiety)
Variable entering	
F-value for variable entering	0.003
Probability level for variable entering	0.955
Percent variance accounted for	11.438
Standard error of predicted Y	2.451

Regression equation:

Achievement = .20 (Neutral Anxiety) -.01 (Closed Anxiety) + .14 (Math Attitude) + 2.376

Form A, Application Sub-Test, Closed Book

Step No. 1

	3(Closed Attitude)
Variable entering	
F-value for variable entering	8.736
Probability level for variable entering	0.004
Percent variance accounted for	7.358
Standard error of predicted Y	2.826

Regression equation:

Achievement = .27 (Closed Attitude) + 4.997

Step No. 2

	1(Neutral Anxiety)
Variable entering	
F-value for variable entering	3.022
Probability level for variable entering	0.085
Percent variance accounted for	9.857
Standard error of predicted Y	2.800

Regression equation:

Achievement = .16 (Neutral Anxiety) + .29 (Closed Attitude) + 2.645

Step No. 3

	4(Math Attitude)
Variable entering	
F-value for variable entering	2.133
Probability level for variable entering	0.147
Percent variance accounted for	11.602
Standard error of predicted Y	2.786

Regression equation:

Achievement = .17 (Neutral Anxiety) -.07 (Closed Anxiety)
+ .20 (Closed Attitude) + .15 (Math Attitude)
+ 2.869

APPENDIX 4

Multiple Regression Analysis on Form B, Open Book

Neutral Anxiety (Var.1), Open Anxiety (Var. 2), Open Attitude (Var. 3) and Mathematics Attitude (Var. 4) are used to predict achievement on the total test (Var. 8).

A correlation matrix between the five variables given above and three additional variables is given in the table below.

The three additional variables are

(a) Knowledge sub-test (Var. 5)

(b) Comprehension sub-test (Var. 6)

(c) Application sub-test (Var. 7)

TABLE

n = 118

	1	2	3	4	5	6	7	8
1	1.000	0.285	-0.182	-0.046	0.073	-0.107	0.045	-0.005
2		1.000	-0.331	-0.115	-0.149	-0.158	-0.079	-0.141
3			1.000	0.414	0.193	0.213	0.204	0.230
4				1.000	0.194	0.331	0.346	0.384
5					1.000	0.347	0.346	0.540
6						1.000	0.570	0.830
7							1.000	0.912
8								1.000

Form B, Total Test, Open Book

Step No. 1

Variable entering	4(Math Attitude)
F-value for variable entering	20.072
Probability level for variable entering	0.000
Percent variance accounted for	14.751
Standard error of predicted Y	4.958

Regression equation:

Achievement = .38 (Math Attitude) + 14.890

Step No. 2

Variable entering	2(Open Anxiety)
F-value for variable entering	1.306
Probability level for variable entering	0.255
Percent variance accounted for	15.708
Standard error of predicted Y	4.952

Regression equation:

Achievement = .09 (Open Anxiety) + .37 (Math Attitude) + 17.571

Step No. 3

Variable entering	3(Open Attitude)
F-value for variable entering	0.342
Probability level for variable entering	0.560
Percent variance accounted for	15.961
Standard error of predicted Y	4.966

Regression equation:

Achievement = -.08 (Open Anxiety) + .06 (Open Attitude) + .35 (Math Attitude) + 16.343

Step No. 4

Variable entering	(Neutral Anxiety
F-value for variable entering	0.298
Probability level for variable entering	0.586
Percent variance accounted for	16.182
Standard error of predicted Y	4.981

Regression equation:

Achievement = .05 (Neutral Anxiety) - .09 (Open Anxiety) +
.06 (Open Attitude) + .35 (Math Attitude) + 15.370

Multiple Regression Analysis on Form B, Closed Book

Neutral Anxiety (Var. 1), Closed Anxiety (Var. 2), Closed Attitude (Var. 4) are used to predict achievement on the total test (Var. 8)

A correlation matrix between the five variables given above and three additional variables is given in the table below. The three additional variables are

(a) Knowledge sub-test (Var. 5)

(b) Comprehension sub-test (Var. 6)

(c) Application sub-test (Var. 7)

TABLE

n = 153

	1	2	3	4	5	6	7	8
1	1.000	0.435	-0.259	-0.209	-0.151	-0.113	-0.137	-0.150
2		1.000	-0.408	-0.174	-0.111	-0.160	-0.056	-0.120
3			1.000	0.451	0.262	0.276	0.264	0.309
4				1.000	0.389	0.440	0.401	0.476
5					1.000	0.464	0.543	0.885
6						1.000	0.693	0.928
7							1.000	0.928
8								1.000

Form B, Total Test, Closed Book

Step No. 1

Variable entering	4(Math Attitude)
F-value for variable entering	44.235
Probability level for variable entering	0.000
Percent variance accounted for	22.657
Standard error of predicted Y	5.532

Regression equation:

Achievement = .48 (Math Attitude) + 11.562

Step No. 2

Variable entering	3(Closed Attitude)
F-value for variable entering	2.169
Probability level for variable entering	0.143
Percent variance accounted for	23.760
Standard error of predicted Y	5.511

Regression equation:

Achievement = .12 (Closed Attitude) + .42 (Math Attitude) + 9.188

Step No. 3

Variable entering	1(Neutral Anxiety)
F-value for variable entering	0.208
Probability level for variable entering	0.649
Percent variance accounted for	23.866
Standard error of predicted Y	5.525

Regression equation:

Achievement = -.03 (Neutral Anxiety + .11 (Closed Attitude) + .42 (Math Attitude) + 10.504

Step No. 4

Variable entering	2(Closed Anxiety)
F-value for variable entering	0.046
Probability level for variable entering	0.831
Percent variance accounted for	23.890
Standard error of predicted Y	5.543

Regression equation:

Achievement = -.04 (Neutral Anxiety) + .02 (Closed Anxiety) + .12 (Closed Attitude) + .42 (Math Attitude) + 10.009