This is to certify that the

dissertation entitled

MATHEMATICS PLACEMENT PROCEDURES

AND PSYCHOMETRIC DECISION THEORY

presented by

James Andrew McComb

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Department of

Counseling, Educational Psychology and

Special Education

William A. Mehrens
Major professor

Date July 25 1983

MATHEMATICS PLACEMENT PROCEDURES AND PSYCHOMETRIC DECISION THEORY

By

James Andrew McComb

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology
and Special Education

1983

ABSTRACT

MATHEMATICS PLACEMENT PROCEDURES AND PSYCHOMETRIC DECISION THEORY

By

James Andrew McComb

The purpose of this investigation was to examine Math placement procedures within the framework of psychometric decision theory. The MSU Math placement test and the American College Test Math score were considered as placement variables. Alternate treatments ending in a common class were compared to determine if they compensated for initial differences in ability.

Within the psychometric decision framework, aptitude treatment interactions (ATI) should exist if there is justifiable reason for assigning students with differing characteristics to alternate treatments intended to produce similar learning outcomes.

A long algebra sequence was compared with a short algebra sequence and a long calculus sequence was compared with a short calculus sequence. Freshmen entering the university Fall terms 1978 to 1981 who completed Math classes were considered.

Analysis of covariance was used to determine if there were aptitude treatment interactions.

The Johnson-Neyman technique was used to determine which cutting score ranges were significantly different.

No consistent ATI were discovered for either the algebra or calculus sequences when either the MSU Math placement test or the ACT Math score was used as the placement variable. Students completing the short algebra sequence outperformed students completing the long algebra sequence during both 1978 and 1979 when the MSU Math placement test was used as a covariate. There was a reversal in the calculus treatment which produced the best result between 1978 and 1979: the reversal may be attributed to grade deflation and stricter enforcement of placement policy. Students enrolling in classes at the level recommended tended to outperform students who enrolled in higher than recommended classes. Comparative treatments did not tend to compensate for differences in initial ability.

When only students who enrolled consistent with placement policy were examined, no discontinuities in regression lines appeared which would indicate the existance of ATI. The predictive validities for both the MSU placement test and the ACT Math score were low. Correlations between the first grade in a sequence and the last grade were moderate to low. The inconsistency in treatment outcomes indicates that placement policy should not be developed and maintained on data from a single year. When testing costs are considered the ACT may be as useful as the local test for placement.

Dedicated to

Helen

and

Susan

ACKNOWLEDGMENTS

To Dr. William Mehrens, my committee chairman and major advisor, whose help and support I greatly appreciate.

To Dr. John Vinsonhaler, my research director, who helped to formulate this study in its early stages and provided financial support.

To the other members of my committee for their ongoing willingness to help me: Dr. Walter Hapkiewicz, Dr. Fred Ignatovich, and Dr. Sarah Sprafka.

To Dr. Charles Eberly who shared his experiences and helped me to collect the data.

To Dr. William Rosenthal who helped me greatly with data analyses.

To Dr. James Hamilton who was always very supportive of my efforts.

To Richard Lee who helped me with proofreading and editing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: The Problem

Introduction

Michigan State University like many other similar institutions uses a locally developed Math test in order to place new students into an appropriate course in the Math sequence. The test is 30 items long and has KR20 reliability estimates ranging from .83 to .88. The standard error is approximately 2.19. Students taking Math are usually placed into one of six courses on the basis of their Math placement test score. The title of each course and the current cutting range associated with it are displayed in Table 1.

Freshman students take the placement test and enroll for classes during academic orientation. An individual enrolling in Math at orientation may not take a class above the level indicated by their placement test score without receiving special permission from a Math guidance counselor. Permission to enroll in a higher level course may be granted if the counselor feels it is justified upon the basis of past Math performance, an informal verbal quiz, standardized test scores, or the belief that the student is highly motivated to succeed. Students may take a placement retest if they wish to improve their

Table 1

The Math classes offered and the cutting score ranges associated with each class.

Class	Title	Score Range	Comment
Math 081/103	Elements of Algebra	0-7	081 & 103 are taken concurrently - credit is given for 103 only
Math 082/104	Intermediate Algebra	8-13	082 & 104 are taken concurrently - credit is given for 104 only
Math 108	College Algebra & Trig. I	14-17	Math 108 & 109 are a two course sequence
Math 109	College Algebra & Trig. II	-----	see 108
Math 111	College Algebra with Trig.	18-23	Math 111 or Math 108/109 may be taken to prepare for Math 112
Math 112	Calculus & Analytic Geometry I	24-30	

scores and receive permission to enroll in a higher level course.

Course enrollment is not monitored after orientation. During registration freshman students may change their course schedules from those that were reserved during orientation. Returning students may enroll at any point in the Math sequence.

Ideally, students would begin the sequence at the point for which they are prepared and continue in the sequence until they have completed the Math requirements of their major (see Table 2). Not all students follow the ideal sequences. Some of the patterns displayed by students may include: taking courses in-sequence beginning consistent with placement test scores; taking courses out-of-sequence beginning above or below the level indicated by placement test scores; or taking alternative courses which satisfy the requirements of their major.

Over a period of three academic terms (Fall 1979, Winter 1980, and Spring 1980) students taking Math 081, 082, 108, 109, or 111 fall term displayed approximately two hundred patterns of Math coursework. Patterns included retaking the course, taking sequential courses, taking non-sequential courses, taking no additional Math coursework, taking courses lower in the sequence, and dropping out of the institution. Similar patterns were found during the 1980-81 academic year.

Table 2

Ideal course sequences for each beginning level of placement for students who will complete Math 112 or higher.

Beginning Placement	Ideal Sequence
081/103	081/103, 082/104, 108, 109, 112
082/104	082/104, 108, 109, 112
108	108, 109, 112
111	111, 112

Need for Study

Many students perform poorly in the Math courses they take. In reviewing fall term grade distributions for Math 081, 082, 108, 109, 111, and 112 for the years 1976-1979 it is found that between 28.5% and 55.0% of all students enrolled in any one class received a grade below a 2.00. The mean Math grade of students taking Math classes (all classes) during the 1979-1980 academic year was 2.29 with a standard deviation of 1.255 (n=7270). During the 1980-1981 academic year 8825 students took Math courses and earned an average grade of 2.16 (standard deviation = 1.259).

There are many potential factors influencing a student's failure in Math: inappropriate placement, instruction inappropriate to the learner, lack of motivation, uncontrolled environmental factors, and lack of student ability. It seems likely that for many students it is a combination of unfavorable conditions which ultimately leads to failure. High failure rates have important implications for the institution and for individuals. The institution must support more sections of classes to meet the demands of students who must retake a class, while students must pay for the courses a second time. Students who become extremely discouraged or fall too far behind their peers may change their majors or may even leave the institution. Not all students should be expected to succeed in all courses; however, decreasing the failure rate would be beneficial to both the institution and to the involved students.

Math education at the university level is a complex process involving several components: setting and review of educational objectives, coordination of educational goals, placement and guidance, instruction, and student motivation. These factors act both independently and in conjunction with each other to affect outcomes. If the educational objectives are judged sound (which seems likely since they are usually developed by relevant faculty), a logical point to begin an investigation of the education process is with placement. Willingham (1974) reported that students do best when placed appropriately in the educational sequence.

> Placement is intended to get students started at the right level in a subject according to thier preparation and moving at their own speed. (p. 55)

Without adequate placement an individual's chance to gain the greatest benefit in the most efficient manner decreases.

If the placement process is not adequate too many students are likely to be misplaced. Students who are placed above their level of readiness are more likely to fail than are students who are appropriately placed. Those individuals placed below their readiness level lose valuable time studying material they already grasp and may not be sufficiently motivated to perform optimally. Experiencing extreme difficulty or failing a Math course may leave the student with a negative view of Math which may affect his/her future Math performance (Bloom, 1976).

Often students are given alternate treatments which are intended to result in similar learning outcomes. If one treatment is more difficult, expensive, or time consuming than the other it is necessary to determine the relative merits of each treatment for students with different characteristics or ability levels. Individuals in a treatment designed to match their entering characteristics should perform better in the treatment than do individuals whose entry level of ability is below that for which the treatment is designed.

Theoretically students will perform best when placed at the proper entry point in the learning sequence or in the treatment most suited to their characteristics. Improper placement reduces the efficiency and effectiveness of instruction. It is important to investigate placement procedures to determine their appropriateness and relative worth. Poor placement procedures are likely to increase the utilization of resources unnecessarily and to have a negative impact on student achievement.

Purpose of the Study

This study had several objectives. The major purpose of the study was to fit MSU placement procedure data to the Psychometric Decision Model developed by Cronbach and Gleser (1965), in order to determine if the current cutting scores result in the maximization of outcomes as conceptualized in the Decision Theoretic framework. Within the context of the model, aptitude treatment interactions are expected to occur in the vicinity of the current cutting scores.

In conjunction with the major purpose of the study a number of other issues were investigated. Descriptive data concerning the Math placement procedure, the outcomes of the procedure, and student performance data were collected and interpreted. The comparative usefulness of the American College Test Math subscore (ACTms) and the Michigan State University Math Placement Test (MSUpt) was considered. The performance of students who enrolled in courses suggested for their placement test scores were compared with the performance of students who enrolled in courses at higher levels than those suggested for their placement test scores.

Educational Decisions

Education involves making many decisions at many points. This study was undertaken within the framework of Psychometric Decision Theory (Cronbach and Gleser 1957, 1965). Placement of students into different courses constitutes one form of educational decision. Decisions are usually made on the belief that they will result in the greatest good when all things are considered. Institutional resources as well as pedagogical beliefs play a role in most educational decisions.

Cronbach and Gleser (1957, 1965) discuss the nature of educational placement decisions. They conclude that when students are placed into different courses on the basis of a single score or the combination of several scores there should be evidence of differential payoff. If students with different characteristics do not perform better under alternate treatments then there is little reason to provide more than one treatment. Discovering aptitude treatment interactions which are useful for placement may be difficult (Cronbach and Snow, 1981). The complexity of aptitude treatment interactions may preclude their discovery in many situations. The usefulness of interactions for sectioning independent groups of individuals into different treatments is open for investigation. Psychometric Decision Theory as discussed by Cronbach and Gleser 1957, 1965), Hills (1971), Mehrens and

Lehmann (1973), Willingham (1974), and Cronbach and Snow (1981) will be discussed in detail in Chapter 2.

Hypotheses

The hypotheses which were investigated in this study are stated in general form below.

1. The cutting scores currently used for sectioning students into either Math 108 or Math 111 are set appropriately in relation to the Psychometric Decision Model.

2. The cutting scores currently used for sectioning students into either Math 081/103 or 082/104 are set appropriately in relation to the Psychometric Decision Model.

3. Cutting scores may be developed using the ACTms which are appropriate in terms of the Psychometric decision model.

4. Students who enroll in selected courses suggested for their MSUpt scores will not perform differently than students who enroll in courses at a higher level than are suggested for their placement scores.

5. Students completing the Math 108, 109, 112 sequence will not perform differently in Math 112 than will students completing the Math 111, 112 sequence.

Overview

In Chapter 2 related literature will be reviewed. The methodology of the study is presented in Chapter 3. The sample, measures, design, hypothesis, and method of analysis are considered. In Chapter 4 the results of the study are presented and conclusions are discussed in Chapter 5.

Chapter 2: Review of the Literature

Decision Theory

Decisions involve the choice between two or more alternate courses of action. In educational situations test information is often used for decision making. Decisions are an integral component of education and the processes by which they are made deserve careful consideration. Mehrens and Lehmann (1973) point out the centrality of decision making in educational situations.

> The direct involvement of everyone in education means that every person must at some time make educational decisions....This brief introductory section is intended to focus the readers attention on the basic notions that educational decisions must be made, that these decisions should be based on information, that this information should be accurate, and that the responsibility of gathering and imparting that information belongs to educators. (p. 5)

Statistical Decision Theory was developed more than thirty years ago within the field of Mathematical statistics and deals largely with estimating the relative benefits which can be expected from different courses of action (Willingham, 1974). Early work in decision theory centered around economics and the maintenance of quality control in industrial operations. The application of decision theory to educational problems was pioneered by Cronbach (Cronbach and Gleser 1965, Hills 1971, Willingham 1974). The underlying principle of decision theory as it is applied to educational problems is that individuals with differing degrees of relevant characteristics will perform differently when conditions vary. The choice concerning which treatment an individual should undertake depends upon which course of action is expected to produce the greatest benefit when cost and other limiting factors are considered. Whenever possible, students should be given the treatment which will result in the greatest good.

Cronbach and Gleser (1957, 1965), Hills (1971), Mehrens and Lehmann (1973), and Willingham (1974) among others have discussed the nature of educational decision making. Mehrens and Lehmann (1973) have summarized the components of decision theory discussed by other authors.

> Psychologists, economists, political scientists, educational administrative theorists, and others have been studying the whole process of decision making and have built various models describing how people should make decisions. Although these models vary somewhat in detail, they have several things in common.
>
> 1. Decision making is defined as the act of chosing among various courses of action or their alternatives.
> 2. Each alternative (or course of action) has several possible outcomes.
> 3. Each outcome has, at least theoretically, some given probability (chance) of occurrence.
> 4. Each outcome has a certain utility value or desirability).
> 5. The expected utility value (or desirability) of each alternative in the decision making process can be obtained by considering the probabilities and utilities of all possible outcomes for each alternative.

6. The alternative with the highest expected utlility value should then be chosen.

> According to the models, then, a person making a decision should be aware of (1) all the alternatives, (2) the possible outcomes of every alternative, and (3) the probabilities and utilities of the outcomes. The more information one has about these variables, the better the decision is likely to be. The decision-making process portrayed by the model is time consuming, but is less expensive in the long run than the result of making poor decisions. (p. 4)

In its simple form the decision model suggests selecting the alternative course of action which is likely to result in the greatest gain when the value placed on a particular outcome and the probability of that outcome occurring is considered. Educational placement involves assigning an individual to the most suitable treatment. Since decision theory deals with the selection among alternate courses of action in an attempt to maximize a specified outcome, it is particularly suitable for application to educational placement problems. In the following sections the role of decision theory in test theory will be discussed, the components of the decision model will be defined, the application of the model to placement problems will be expanded upon, and complexities of the application of the model will be discussed.

Classical Test Theory and Decision Theory

In classical test theory, tests are predominantly conceptualized as measuring instruments and major importance is attached to the accuracy of measurement on a continuous scale. The purpose of tests tends to be either to describe or to make a decisions about people or groups (Cronbach, 1971). Cronbach and Gleser (1965) maintain that the final purpose of any personnel testing is to reach qualitative decisions and that the value of a test depends upon factors in addition to the accuracy of the test or the size of the validity coefficient.

> Our society continually confronts people with decisions for which they have inadequate information. It is for this reason that psychological and educational tests exist....Some of the problems on which tests are brought to bear are purely individual....Equally numerous are the occasions on which an administrator, teacher, or clinician turns to tests for assistance in making decisions about many people....
>
> It is therefore desirable that a theory of test construction and use consider how tests can best serve in making decisions. Little of present test theory, however, takes this view. Instead the test is conceived as a measuring instrument, and test theory is directed primarily toward the study of accuracy of measurement on a continuous scale. Hull (1928, p. 298) voiced a principle that has been the root of nearly all work on test theory: "The ultimate purpose of using aptitude tests is to estimate or forecast aptitudes from test scores." It is this view that we propose to abandon. We acknowledge the usefulness of accurate estimation - but we maintain that the <u>ultimate</u> purpose of any personnel testing is to arrive at qualitative decisions....
>
> The value of a test depends on many qualities in addition to its accuracy. Especially to be considered are the relevance of the measurement to the particular decision being made, and the loss resulting from an erroneous decision. (p. 1)

In the perspective expressed by Cronbach and Gleser, the value of a test does not depend solely on a high degree of accuracy but depends also upon how helpful it is in improving the decisions that are to be made in a particular situation. A test with a moderate or low predictive validity coefficient may be very valuable in a decision situation where it predicts something that is not otherwise easily discernible. A test with a high validity coefficient may be of little use in a situation where the information it provides is readily available from other sources. Since test information is so often used for decision making, Cronbach and Gleser (1965) argue that test theory should consider the decision situations within which test information will be used. Current test theory deals only with a limited range of situations where test results are used to choose among separate courses of action. Cronbach and Gleser (1965) conclude that there is no integrated theory of testing.

> In distinguishing between so many different decision problems, this chapter makes clear the need for broadening existing test theory. Any test theory well developed at present deals only with a very limited area within our category system....We are forced to conclude that there is no "theory of mental tests" at the present time, although there are many fragmentary theories. (p. 17)

Components of the Decision Model

The components of the decision model are discussed in detail in the following sections. Willingham (1974) identifies three basic elements of the decision model: (1) an assessment variable, (2) an outcome criteria, and (3) treatments. For there to be a rationale for assigning individuals to different courses of action there should be evidence that persons with different characteristics will be more or less successful in different situations. If an aptitude is not differentially related to success in two or more treatments it is relatively useless for placement of students into the treatment. Therefore, within the decision framework there must be an aptitude treatment interaction for there to be any rationale to assign individuals to different treatments on the basis of measured characteristics.

> It is apparent that the key to successful use of alternate treatments is the trait treatment interaction. (Willingham, 1974, p. 23)

If a characteristic is relevant for differential assignment, students with low scores on the assessment variable should do better in one treatment while students high on the assessment variable should do better in the other treatment. If all participants do better, on the average, in a single treatment regardless of their position on the assessment continuum, then it is most productive to assign all students to the treatment in which they are expected to do best.

Decision making involves chosing the action which is expected to result in the greatest gain in the long run. There are a number of concepts inherent to the decision model which need to be defined and discussed in order to present a complete picture of the process. The general concepts of the decision model are presented below.

<u>Strategies</u> A strategy is the rule or set of rules by which decisions are made. A strategy specifies the course of action that will be selected when a decision maker encounters any possible contingency. A matrix consisting of conditional probabilities concerning the likelihood of the choice of a particular action from among several can be developed for any strategy. The strategy matrix indicates the probability that any one choice will be made given a particular set of information. Figure 1 displays a general strategy matrix.

Figure 1

Strategy matrix for the placement of individuals into three sequential courses depending upon their predictor test score.

Test Score	Course A	Course B	Course C
y1-y10	pa/y1-y10	pb/y1-y10	pc/y1-y10
y11-y20	pa/y11-y20	pb/y11-y20	pc/y11-y20
y21-y30	pa/y21-y30	pb/y21-y30	pc/y21-y30

<u>Note.</u> pa/y1-y10 is the probability that an individual with a placement score between 1 and 10 will be assigned to Course A.

It is possible to empirically compare the proposed strategy and the actual strategy that is applied. Comparison of the stated strategy with the decision maker's actual practice is often useful for identifying weaknesses in decision procedures. Inconsistencies between intended strategies and actual practices may indicate that the decision maker either uses information not included in the strategy or makes inconsistent decisions. The best strategy is the one which results in the greatest cumulative good.

Treatments A treatment is any alternative course of action that may be specified in the decision process. Placing students at different levels in the same subject constitutes assignment to different treatments. Sorting students into alternate curriculum on the basis of aptitude scores is another example of providing different treatments. In admissions decisions both acceptance and rejection are treatments.

Fixed and Adaptive Treatments Fixed treatments are determined prior to the collection of information about the individuals who are to be assigned to the treatments. Individuals are then given the treatment which is expected to best suit their characteristics. No attempt is made to modify treatments in light of group characteristics.

Treatments may vary continuously. Minor changes along a number of treatment dimensions may be made until the treatment is optimal for the characteristics of individuals involved. The use of adaptive treatments usually results in greater benefits. Unfortunately, in many situations adaptive treatments are hard to develop, hard to administer, and costly. In its extreme form the use of adaptive treatments results in individualized instruction.

<u>Assessment Variable</u> The assessment variable is sometimes referred to as the information category or the predictor variable. It is the characteristic (or combination of characteristics) upon which the decision concerning alternative courses of action is made. The assessment variable is a measure of the individual's characteristics which are expected to interact with the treatment. Such characteristics vary continuously across individuals. In placement, the assessment variable is often a placement test score.

<u>Maximization</u> In institutional decision making the strategy chosen to guide decision making is generally the one which maximizes the average gain in outcome over many similar decisions (Cronbach and Gleser, 1965). The preferred course of action is the one which will result in the greatest average gain over many cases.

<u>Institutional and Individual Decisions</u> Institutional and individual decision strategies vary. For institutional decisions a large number of similar decisions are made using a constant strategy. The intended outcome of the process is to maximize the resulting outcomes. Benefits are cumulative over decisions. Individual decisions differ from institutional decisions in that they are often unique. If the decision is likely to occur only once or infrequently the average gain over decisions cannot be determined. The values people place on outcomes are

subjective. Any two individuals may place a different value on the same outcome. Since individuals have different value systems the worth of any outcome of similar decisions is not directly comparable over individuals.

Cronbach and Gleser (1965) report that current test theory is more applicable to institutional decisions than to individual decisions.

> Regression formulas, for example are designed to maximize the average squared error of estimate. Institutional decisions lend themselves to "strong" mathematical and statistical treatment, making generalizations possible. (p. 9)

Statistical decision theory concentrates on institutional decisions which can be averaged across individuals using a common utility scale. Decision strategies are developed in order to assign individuals to alternate courses of action which are expected to result in the greatest gain across individuals over time.

Utility Utility concerns the relative value or worth of possible outcomes. Each outcome must be evaluated and assigned a position on a common cardinal scale if outcomes are to be compared over treatments. The evaluated outcome (utility) is sometimes referred to as the payoff or benefit and the common scale is the utility scale. A gain in utility using strategy A instead of strategy B indicates that the use of A results in a more desirable outcome on the average. The utility of a test is determined by comparing the average outcome obtained using the test information with the average outcome realized if the test were not used. Without a common utility scale it is not possible to determine the relative gain from a strategy or a course of action. The utility scale will reflect that which the decision maker wishes to maximize (Hills, 1971).

> In order to define an expected gain and maximize it, it is necessary to make an important assumption about the utility scale on which the outcomes of possible decisions are evaluated. We must assume that the value of various outcomes can be expressed in "equal units of satisfaction" or the like, which are additive over many decisions. (Cronbach and Gleser, 1965. p. 10)

A common utility scale makes it possible to determine which placement strategy is likely to produce the most desirable outcome for individuals with a particular predictor score. It is possible to estimate the expected payoff for such individuals by summing the products of the evaluated outcomes (outcomes expressed on the utility scale) and their respective probabilities.

The probabilities of the outcomes are based on _a priori_ knowledge of the joint distribution of the assessment variable and the criterion variable which is expressed on a utility scale. If a test or other instrument is used for gathering information its cost should be expressed on the common utility scale and subtracted from the expected payoff.

Using one set of information for decision making may result in greater gain than using another set. The cost of collecting information must be considered in the gain and reflected in the utility. If the average learning outcome is increased only slightly by use of expensive test information over currently available records, it may not be justifiable to use the test for the decision making procedure. The gain in utility using test scores must be judged against the utility which is achievable through using *a priori* information.

A major difficulty in applying decision theory to educational placement is the development of a cardinal utility scale. It may not be possible to express dollar costs and learning outcomes on the same scale (Willingham, 1944).

Cronbach and Gleser (1965) provide a method for estimating the utility of a set of decisions averaged over a large number of individuals. The expected utility for a large number of decisions is determined by summing the products of the expected payoff for each predictor score and the probability of the predictor score. The appropriate formulas are provided by Cronbach and Gleser (1965, pp. 54-63).

The determination of the overall utility will ultimately be a subjective weighting of values. This may not be an untenable situation since many educational values are subjective. If the decision maker and other vested parties are satisfied with the utility scale it may be appropriate for the purpose for which it is being used.

Outcomes The outcome of a decision entails all consequences of the course of action taken which are of relevant interest to the decision maker. The outcome which occurs from a specific course of action will depend upon the characteristics of the individual, the characteristics of the treatment, and uncontrolled environmental variables. An actual outcome for any specific individual cannot be predicted with certainty. The probability of an outcome for an individual or the probability distribution for a group of individuals can be determined.

Validity Matrix Outcomes are predicted from assessment variable information (i.e. predictor test scores). A validity matrix providing the empirical relationship between the assessment or predictor variable and each outcome or criterion can be developed for each treatment. The validity matrix for a treatment is a set of conditional probabilities specifying the likelihood of an outcome given a certain assessment or predictor score. The typical validity matrix is provided in Figure 2 below.

Quotas Quotas may be fixed or adaptive. With fixed quotas a certain number or proportion of individuals must be assigned to a certain treatment. In such a case it is likely ,for example, that all the highest scoring individuals would be assigned to the more difficult treatment. With adaptive quotas the numbers or proportions assigned are modified in order to maximize the expected benefit. Educational placement problems often allow adaptive quotas to be used.

Figure 2

Typical validity matrix providing conditional probabilities relating predictor scores to outcomes for a single treatment.

Predictor Score	Outcome A	Outcome A
1	pa/1t	pb/1t
2	pa/2t	pb/2t
3	pa/3t	pb/3t

Note. pa/2t is the probability of an outcome given than an individual with a predictor score of 2 is placed in treatment t.

Criterion The criterion is the measure of the outcome of the treatments. In decision theory it is expressed on the utility scale. Criterion measures may be end-of-course achievement or any educational outcome including persistence or satisfaction.

Relation of Treatments, Quotas, and Utility In general, the gain in utility over the best *a priori* strategy will be limited by whether treatments and quotas are fixed or adaptive. The greatest gain in utility can be achieved when both treatments and quotas may be adjusted. In many cases of educational placement there are *a priori* a fixed number of treatments, and quotas are adapted. Individuals are then assigned to the treatment which is expected to result in the best outcome for them.

Classification, Selection, and Placement All decision problems can be considered as cases of classification (Cronbach and Gleser, 1965). Classification involves assigning an individual to a category or treatment when more than one category or treatment is available. Both placement and selection are forms of classification. Most measurement using tests can be interpreted in terms of the placement model (Cronbach and Gleser, 1965).

> It is helpful to consider separately two special cases of classification that are very much simpler to analyze than the general problem. The first case is that where information is univariate. Even when information is obtained in terms of more than one score or dimension, it is a common practice to combine such multivariate information into a single composite score before making decisions. If scores on the same composite scale are used in making all decisions between treatments, the classification problem may be termed a placement problem....The most common examples are dividing students among sections to be taught at different rates and using a trade test for course grouping of applicants...We shall see later that "measurement" problems can be considered as a particular variant of the use of tests for placement decisions.
>
> Problems may also be differentiated according to whether or not rejection is allowed as one possible treatment, so that the person is eliminated from the institution. We can refer to these as selection problems; (p. 13)

Willingham (1974) identifies four classes of treatments: (1) assignment, (2) selection, (3) placement, and (4) exemption. The definition of placement for Willingham is somewhat more restrictive than it is for Cronbach and Gleser.

> [Placement is] Positioning students at the optimal point in an instructional sequence on the basis of how much the student knows about the subject. In

> this context, placement corresponds fairly closely to conventional use of the term. Students are placed on the basis of subject-matter tests in alternate treatments that vary on the basis of subject-matter content. Treatments vary in length (e.g. a one-course sequence versus a two-course sequence), but they always have a common subject-matter criterion at the end of the sequence. the general purpose of placement is to match the content of instruction with what the student needs to learn next. (Willingham, 1974, p. 19)

Hills (1971) makes a distinction between classification and placement similar to that of Cronbach and Gleser (1965) noting that placement is a more restricted form of the classification model. The purpose of educational placement is to start the student in the learning sequence (or treatment) at a point where the student will neither be overwhelmed nor completely unchalleged.

> One approach to placement consists of trying to locate the student at the proper point in the sequence of courses according to how much he already knows. A student may be exempted from, and perhaps given credit for, courses below the level into which he is placed. Another approach consists of placing the student according to how fast he might be expected to learn....
>
> The above kinds of placement are attempts to situate the student in the course or treatment that will challenge him but will not overwhelm him - to prevent his wasting time or being bored on the one hand and to prevent his failure due to lack of preparation or lack of sufficient repetition or explication on the other. It appears that there are two ways of doing this. However, both are cases of attempting to place him in an instructional setting that will maximize payoff. This is, then what placement is all about. (Hills, 1971, p. 702)

<u>Placement Models</u> Willingham (1974) identifies three placement models: (1) vertical sectioning,

(2) remediation, and (3) group pacing. Vertical sectioning entails starting a student at the point in the sequence for which he has mastered the prerequisites. The student may begin with the first course in a sequence or may be exempted from the first course if he has already mastered the material offered there.

Remediation is intended to provide students with content or skills needed in an introductory course. It differs from vertical sectioning in that it provides learning prerequisites to the sequence. Institutions of post secondary education often provide remedial courses for students who appear qualified in other areas but who have not accomplished what is considered to be pre-college level work in the particular discipline. The value of remediation is under dispute among some educators. Hills (1971) reports that "remedial courses are generally not very effective in improving subsequent grades or reducing withdrawal" (p. 707).

Different individuals are capable of learning the same materials at varying rates. Some students learn material in certain disciplines much more quickly than do other students. Group pacing is applied in order to match the rate at which material is presented with the rate at which the individual learns. Some students may be given a two term sequence covering the same material that other students can master in one term. Individualized instruction often involves pacing and it is common for college Math courses to be offered in one or two term sequences.

Educational Placement and the Decision Model

Decision theory provides a useful framework for determining the optimal assignment to educational treatments. As previously noted most placement within a subject-matter discipline is undertaken in order to maximize the overall benefits. Students are placed into an available instructional program which is expected to result in the greatest amount of learning. Underlying placement into alternate treatments is the inherent assumption or belief that students with varying characteristics will perform best under different treatments. If students would not perform differently depending upon the treatment they were administered then there would be little reason to provide different sequences other than student preference or institutional necessity.

Several authors have noted the importance of acknowledging the expectation of differential payoff in placement when assessing the value of the tests and setting cutting scores. Cronbach and Gleser (1965) discuss how placement is based on the conscious or unconscious assumption that aptitude treatment interactions exist. Treatments must have different validity coefficients and the regression lines relating the predictor score to the criterion measure must intersect for differential placement to be defensible when all relevant factors are considered in the criterion measure (i.e. a common utility scale is used).

> Even more to the point is the fact that the use of tests for educational placement implicitly assumes the existence of such different payoff functions. Our model therefore is merely a formal statement of what is everyday accepted without question, and examining its implicatons for testing is undoubtedly important. The realization that this assumption underlies practice is in itself an advance, because it makes clear the need for research on payoff functions....
>
> The assumption that payoff functions do exist making placement profitable, is consistent with available theories about instruction. A person who lacks readiness to profit from one experience may be able to learn from another....Accepting the concept of intersecting payoff functions will allow us...to raise fundamental questions regarding the construction and validation of placement tests. (Cronbach and Gleser, 1965, pp. 25-26)

Willingham (1974) acknowledges the need for differential payoff for valid placement. He goes on to point out that if there is a basis for valid placement (i.e. an aptitude treatment interaction exists) then high and low competency students need different treatments in order to make their best gains.

> Any time different students are placed in different treatments in order to facilitate optimal achievement, there is an implicit assumption that the trait and treatment interact, and the usefulness of the whole placement procedure rests on that assumption...the regression lines have to cross if the placement procedure is to result in greater overall achievement. It also makes explicit the fact that under these circumstances differential placement is necessary if instruction is to be

> effective for students at low and high levels of competency... (Willingham, 1974, p. 78)

Hills (1971) discusses the difference between traditional and decision theory approaches to placement.

> For placement to be worthwhile the placement test must have different regression slopes for the various treatments...The traditional model does not recognize this necessity, and indeed, traditional practice in college fails to recognize it also. (p. 714)

In his assessment of educational measurement for the seventies Thorndike (1971) discusses the need for differential payoff from educational treatments in relation to placement procedures.

> Different treatments may be majoring in different subject-matter fields, assignment to different sections of a common course taught in different manners or at different levels, or moving ahead to different units in a course sequence after completing unit X. For such decisions, the logic of classification holds and a test or test battery must be evaluated in terms of its differential validity for the alternate treatments. Recognition of the crucial role of differential validity has gradually spread from research workers concerned with personnel classification in jobs to those concerned with guidance of choices of field of study and on to those concerned with placement of students in differentiated treatments within some one course or program. As awareness of the requirements for the placement and classification use of tests in education increases, a different criterion will be applied to tests for these purposes - the criterion of differential validity. (pp. 9-10)

In summary, placement problems fit the decision model well. To place individuals into alternate treatments on the basis of univariate composite information implies that individuals lying at different points on the predictor continuum should perform differently under each treatment. if such an assumption is not accurate then differential placement on the basis of the univariate composite information will be futile. The existence of aptitude treatment interactions implies that there will be different validity coefficients for each treatment.

> Clearly, since the test may have a different validity coefficient for each treatment, utility is not a simple function of any single validity coefficient. However, the greater the correlations of the tests with the differences in payoff between treatments, the greater the utility from using the test for placement. (Cronbach and Gleser, 1965, p. 56)

The value of placement tests cannot be judged by considering the size of a single correlation coefficient. Traditional methods of test validation are inappropriate for placement tests (Cronbach and Gleser, 1965). Tests used for placement only have value in those particular situations where they differentially predict success in alternate treatments and the differential indication takes the form of an aptitude treatment interaction.

Although a number of authors have indicated that the assumption of existing aptitude interactions underly the differential placement of students there has been very little published research in this area. In fact, there is very little published research concerning the evaluation of placement procedures.

> The lack of published data concerning sound evaluations of remedial procedures is paralleled by the lack of data evaluating placement procedures.

> Review of the literature in this area yields few published studies. The College Entrance Examination Board provides a whole series of Achievement tests on a world-wide administrative basis several times a year. They are often suggested for use in college placement. Even such agencies as this have in the past conducted but few studies of the effectiveness of their test for placement purposes. (Hills, 1971, p. 708)

The lack of published placement evaluations may be due to the fact that many placement procedures are of local concern or result in somewhat inclusive findings. There is extremely little evidence of placement studies framed within the psychometric decision model. A little over ten years ago Hills (1971) reported knowledge of no such studies.

> The author cannot recall ever having witnessed a claim that a commercial or locally developed placement test was efficacious because it had different regression coefficients for different available treatments. (Hills, 1971, p. 714)

More recently Cronbach reported that to his knowledge no serious placement studies within psychometric decision theory guidelines had been undertaken.

> I regret to say that I cannot call to mind a single study in which a serious validation of "placement" along the lines of our theory was attempted. Given that just this day I have been reading an account of a court decision in which the demand was made for validity evidence specific to the placement decision (at least eight years ago), your study takes on added importance. (Cronbach, 1982, personal communication)

There has been substantial research on aptitude interactions in educational settings, however, as noted little has been done in the area of placement. For an extensive review of aptitude treatment interaction research the reader should refer to _Aptitudes and Instructional Methods: A Handbook for Research on Interactions_ (Cronbach and Snow, 1981). It is likely that little aptitude treatment interaction research has been done in the area of placement for the same reasons that little general placement research is to be found and that application of the model is difficult. Despite these restrictions the inherent strength of the model justifies further research.

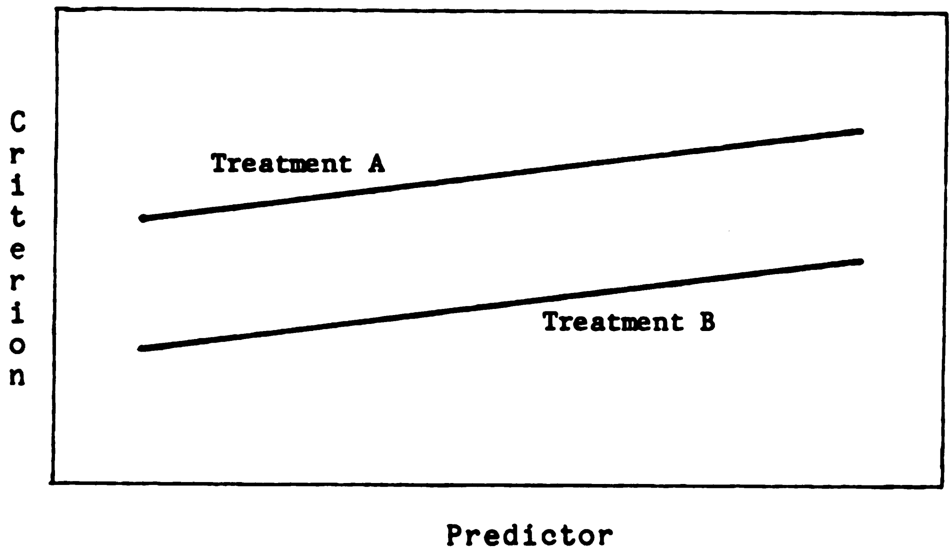
The Nature of Payoff Functions

Payoff functions take the form of regression lines for each treatment relating predictor scores and criterion scores which are measured on a continuous scale. When two treatments are used there are three basic relationship that may exist between the payoff functions: (1) they may have equal slopes and not intersect, (2) they may display disordinal interactions, or (3) they may display ordinal interactions. As noted in previous sections the value of a test for placement purposes depends upon the relationship between the payoff functions for the treatments and the appropriate place for the cutting scores are determined by the intersection of the payoff functions.

Figure 3 illustrates the case where there is no interaction between payoff functions. In this case everyone performs better in treatment A than in treatment B regardless of where the score falls on the predictor continuum. Under such conditions there is no basis for the placement of individuals into the different treatments. If possible everyone should be placed in treatment A. Individuals will be placed into treatment B only if it is impossible for the institution to provide them with treatment A.

Figure 3

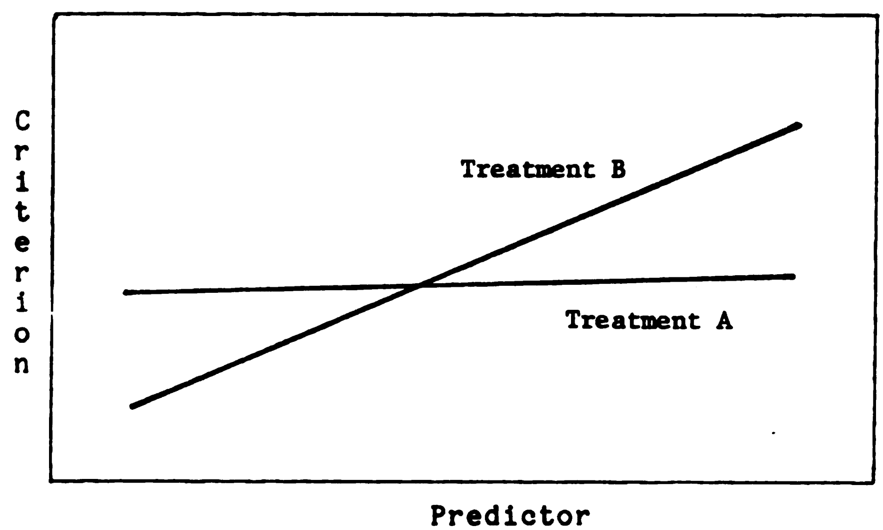
Illustration of payoff functions which do not display an interaction.



When sets of payoff functions display aptitude treatment interactions they provide a basis for the assignment of individuals. Figure 4 illustrates a cases where there is a disordinal aptitude treatment interaction. This outcome suggests that individuals with predictor scores below the point where the regression lines intersect do better in treatment A while individuals with predictor scores above the point of intersection do better in treatment B. The point of intersection of the payoff functions is the optimal point for setting cutting scores if the object is to maximize the outcome as measured on the criterion scale.

Figure 4

Illustration of payoff functions displaying a disordinal interaction.



In the above situation high ability students do not do much better in Treatment A than do low ability students. The nature of treatment A is such that the particular predictor score is not very good for predicting success within the treatment (i.e. it has a low predictive validity coefficient). However, low ability students may be expected to do better in treatment A than they would be expected to do in treatment B. As can be seen from the figure high ability students can be expected to do better in treatment B than in treatment A. The higher ability students should be given treatment B while the lower ability students should receive treatment A.

It is sometimes argued that it is not likely that poorer students will outperform better students in any treatment. In order to examine some of the arguments which have been presented concerning the differential achievement of less and more prepared students it is useful to consider the situation where treatment B is a one term calculus course while treatment A is the same course taught over a period of two terms. The following factors may singularly or in combination lead to situations where superior students do not markedly outperform less prepared individuals.

1. Superior students are not challenged because of the slow pace of the long sequence and fail to perform at their best due to boredom and poor study habits.

2. The long sequence gives the poor students more time to grasp the material and minimizes the advantage of the superior students.

3. The nature of the utility scale upon which the criterion is measured is such that all relevant aspects of the situation are considered in the outcome. The utility scale may reflect costs so that the criterion is more than an indication of achievement. Superior students placed in the two term sequence may outperform poor students, however, they lose in that they must spend two terms covering material they could accommodate in one term.

The above conditions may lead to aptitude treatment interactions. Certain treatment conditions may discourage or fail to encourage superior students while being positive for poor students. In some treatments superior students may have a clear advantage. When all factors are considered on the utility scale, slight increases in achievement for superior students may be outweighted by the waste of their time or talent. The point is further illustrated by the discussion of ordinal interactions presented in the following paragraphs.

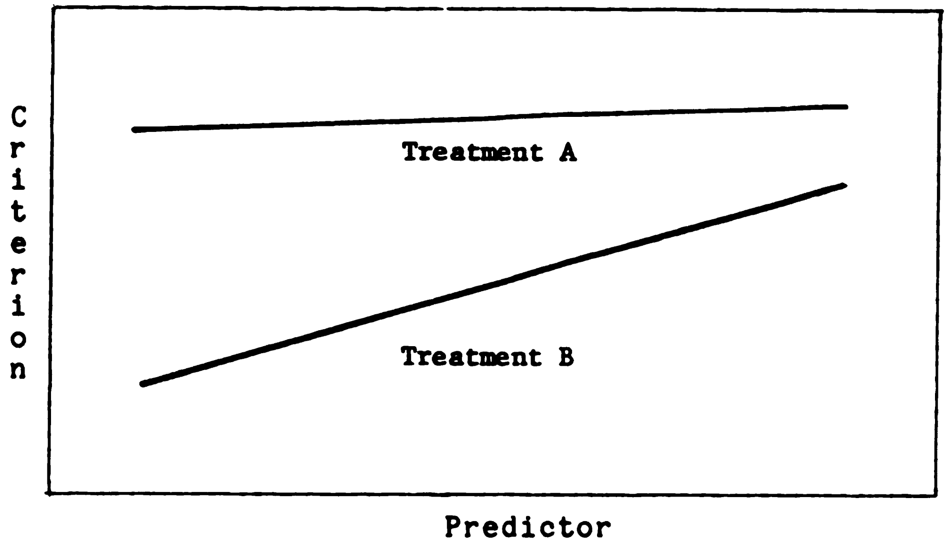
Ordinal interactions occur when the payoff functions have different slopes but do not intersect within the range of the predictor score. An ordinal interaction is illustrated in Figure 5. If the regression lines were extended (assuming linearity and no ceiling effects) they would eventually intersect.

In the situation illustrated in Figure 5 it would appear that treatment A is superior to treatment B for all individuals distributed along the predictor continuum. If the range of the predictor variable were increased and the regression lines remained linear, it is possible that for some individuals high on the continuum treatment B would be superior to treatment A.

Consider the situation discussed above where treatment A is a two term Math sequence and treatment B covers the material in one term. With the criterion scale used in Figure 5 all students do slightly better in the two term sequence than in the one term sequence. Students higher

Figure 5

Illustration of payoff functions which display an ordinal interaction.



on the predictor continuum tend to do almost as well in the one term sequence as they do in the two term sequence. If the additional cost of the extra term is included - that is a utility scale where all relative costs and outcomes are considered jointly - the true situation may be depicted in Figure 4. Once the costs associated with assigning well prepared students to a two term sequence are considered in the utility measure it becomes apparent that the overall benefit to the institution may be increased by assigning better prepared students to the shorter sequence.

<u>Aptitude Treatment Interactions in Research</u>

The existence of aptitude treatment interactions indicates that individuals with differing amounts of a relevant characteristic will do better under some treatments than others. Experimentalists are interested in determining which treatment is generally best. By randomly assigning students to the treatments under investigation and comparing mean outcomes, an experimentalist would conclude that treatment A as depicted in Figure 3 was best for everyone, that treatments A and B as depicted in Figure 4 produced the same results, and that treatment A was superior in Figure 5. The traditional experimental interpretation would be true for the case illustrated in Figure 3 but not for the cases illustrated in Figures 4 or 5 (assuming that an interaction would occur in Figure 5 if an appropriate utility scale were used). In the latter two cases an experimental treatment of the data would mask important information and lead to incomplete conclusions.

A researcher using a correlational approach would draw somewhat different conclusions. With respect to the situation illustrated in Figure 3 the correlationist would find that the test had good predictive validity and was useful for placement. The correlations for the combined groups represented in Figures 4 and 5 would be low and the correlationist would be likely to decide that the test was not very useful for placement.

Identifying aptitude treatment interactions may be more appropriate than traditional experimental or correlational approaches when the objective is to adapt instruction to individual differences. The experimental approach can lead to discovery of the single best treatment for everyone while the correlational approach can lead to the discovery of the treatment which is highly related to selected independent variables. Neither approach provides a complete nor extremely useful picture when the purpose is to assign individuals to treatments in which they will perform best on the basis of a predictor variable. Only when the predictor variable interacts with the treatment variable is there a sound basis for differential assignment to alternate treatments and neither traditional experimental nor correlational research procedures address such concerns.

Methods of Setting Cutting Scores

As noted above, the placement of students into alternate treatments, or at different points in the instructional sequence, assumes that individuals with differing characteristics will perform optimally under different conditions. This issue has not been widely articulated in studies aimed at validating placement tests or setting cutting scores. In many situations cutting scores have been set without ample consideration of the relationship between the placement score and the treatment in which individuals are likely to do best.

The value of a placement test depends upon the degree that it facilitates increasing the desired outcomes in educational situations. If the placement instrument measures a trait which is viable for selecting the best treatment then the value of the placement procedure depends in a large part upon the cutting score that is selected. If the wrong cutting score is chosen then some individuals will be misplaced and the cumulative benefits will be diminished. Various methods of setting cutting scores have been applied to placement procedures. Some methods are more appropriate than others. Several of the traditional methods for setting cutting scores are discussed below. The final method discussed is the one which flows from decision theory. It is the model on which this investigation centers and will be discussed in detail.

Fixed Quota Placement with fixed quotas involve the placement of higher scoring students into one treatment and lower scoring students into another. Less prepared students may be placed into a remedial section until all positions are filled while the best students may be put into the advanced section until it is filled. All other students would be placed into the regular sections.

The use of fixed quotas for setting cutting scores is only slightly better than random assignment. If there is not a nearly perfect positive relationship between the placement test and the outcome measure the justification for using the fixed quota method is weak.

> The first primitive advance over completely arbitrary choice often is choice of cutoff points so that certain quotas are met....The advance is primitive because no arbitrary quota is sensible in placement. The procedure would be tolerable from the individual's point of view if the correlation between the placement test and the criterion were positive and perfect. However, that situation is not likely to occur in practice. Correlations between placement test scores and criteria such as grades are more likely to be around .5 than any number greater than that, and often they will not be that high....
>
> If the placement test does not have a very high correlation with the grade criterion, but if students are placed in the remedial sections on the basis of such a placement test, one is guilty of suggesting to students that he has accurately determined what is wrong with them and is giving them the proper treatments for their difficulties, when in fact the whole procedure may be very doubtful from the placement instrument through the remediation (Hills, 1971, p. 710)

Even if the correlation between the placement test and the criterion is high setting cutting scores using quotas may not be appropriate. If a high proportion of the students are poorly prepared and need remedial work only the lowest scoring students will be assigned to the remedial sections. Other students equally in need of remediation may be placed into regular sections because the remedial quotas have been filled.

If the correlations between the placement test and the criterion are low there may be no cutting score that will result in assigning more than a few people to the extreme sections who would not perform adequately in the regular sections. In such a situation the placement instrument is not appropriate for selecting individuals to be assigned to remedial or advanced treatments.

Use of a Joint Distribution The joint distribution of placement test scores and criterion scores may be considered in determining cutting scores. Cutting scores are set so that students with a given score are more likely to pass the treatment that they are assigned to than to fail it. This is sometimes referred to as a system of "hits" (assigning students to a treatment in which they pass) and "misses" (assigning students to a treatment in which they fail). In some situations students who would be expected to get A's in a lower course may be placed into a higher course where their expected grade would be less than an A but greater than whatever grade is deemed unsatisfactory.

Predicted Probability of a specified Criterion Cutting scores may be determined by predicting the probable grades that individuals with given scores are likely to achieve in alternate treatments. Prediction equations are developed for each treatment and individuals are assigned to the course for which their predicted performance is best. When the object is to place the individual at the proper point in an instructional sequence cutting scores may be set which result in assigning individuals to the highest course in which they are likely to pass even if they would be likely to get a better grade in a less difficult course.

Task Analysis The learning sequence may be broken into tasks which are arranged in hierarchies. Students may then be placed at the point in the sequence where they had mastered the prerequisite tasks. This is typical of individualized instruction and mastery models. With course sequences or treatments involving different techniques it may be difficult to segment the content adequately enough for the task analysis procedure to be productive.

Comparable Performance Students may be exempted from the requirements of a lower course and allowed to take a higher level course if they perform at a level similar to students who have successfully completed the lower level course. This is sometimes classified as an exemption decision rather than a placement decision.

Student Feedback It is possible to adjust cutting scores on the basis of student feedback. When this technique is used students are placed on the basis of a preliminary set of cutting scores. After students complete their studies they are asked how appropriate they feel their placement was. On the basis of student response the cutting scores are raised or lowered.

Decision Theory Approach - Aptitude Treatment Interactions

An important purpose of placement is to increase some desired outcome which is generally an indication of student achievement. If there is a valid basis for placement, other than convenience, there should be an interaction between the placement variable and the treatments into which individuals may be placed. In the decision framework cutting scores are selected in the area where aptitude treatment interactions occur. Individuals are placed into the treatment which is expected on the average to result in the greatest gain in utility when all costs are considered.

The ideal way to determine cutting scores is to administer the placement instrument and then to randomly assign individuals to the treatments under consideration. Regression lines relating the outcome measures to the placement scores are developed for each treatment. Cutting scores are set at the points where the regression lines intersect. If no interaction occurs (see Figure 3) the placement variable is inappropriate for assigning individuals to the different treatments under consideration.

Figure 4 provides an example of a case where there is a basis for the assignment of individuals to the different treatments. Individuals with scores below the point of interaction should be assigned to treatment A, in which they are expected to do best on the average, while individuals with scores above the point of interaction should be assigned to treatment B.

Ordinal interactions may appear if the assessment scale has a low ceiling or the utility scale does not take all relevant factors into account. The interaction may take place outside of the range measured by the placement test so that individuals with higher aptitudes on the assessment variable would actually be better off in the treatment which appeared to be less effective for all individuals. If the restricted score is to be used then all individuals should be placed into the treatment that results in the best average performance.

The utility scale is a second factor that should be considered when ordinal interactions occur. One of the most difficult problems in the application of decision theory is the development of an appropriate utility scale. Within the decision framework the outcome measure is theoretically expressed on a scale which incorporates all relative costs. In actual practice it is not easy to express testing cost, instructional cost, and learning outcomes on the same scale. All utility scales actually used will be subjective to a degree. If the interaction is ordinal it may be that certain costs or benefits have not been considered in the construction of the utility scale.

In Figure 5 treatment A is superior to treatment B regardless of an individual's score on the predictor scale. If treatment A is very expensive and takes longer to administer than treatment B an adjustment in the utility scale may be appropriate. In such a case the relative utility of treatment B would increase when all relevant factors were incorporated into the outcome scale. The result would be a situation similar to the one displayed in Figure 4 where there was a valid basis for differential placement.

Ordinal interactions which are the result of the use of less than optimal utility scales may be common in placement. When grades or examination scores are used as outcome measures there are no objective procedures for considering testing, instructional, and human costs concurrently. By considering the related costs in a general way the investigator may adjust the position of the cutting scores to account for factors which are not directly considered when outcome scores are assigned. The cutting scores are then placed at the point of the interaction of the adjusted regression lines. This process may seem somewhat subjective. However, the value placed on any educational outcome and the relative costs of educational processes are ultimately subjective. If the adjustment is consistent with the values that the decision maker wishes to maximize and can be defended as logical then they are appropriate.

In many decision problems grades are used as the criterion of success despite their actual and potential weaknesses. Some potential problems associated with the use of grades include: (1) they may not be consistent across teachers or treatments, (2) they do not account for testing costs or instructional costs or, (3) they may not represent subject-matter competency. A number of researchers have recognized the difficulties of using grades as the outcome measure in decision problems.

> The models that have been developed for applying decision-theory approach to personnel decisions involve a number of assumptions that may not be met in practice. The first is that one have an acceptable measure of benefit or utility. That grades are a direct measure of utility and that the conventional numerical equivalents fairly represent the utility attached to each grade is quite a hazardous assumption, but one that is implicit in most treatments of grades. (Hills, 1971, p. 715)

> The criterion provided by course marks is notoriously unsatisfactory, but the ease of obtaining such data makes them the most common of all outcome measures. The difficulties are least serious when all the grades were assigned by a single teacher in a single class, since then, the students are likely to be located on the same scale, but there is no guarentee that this scale truely represents mastery of the course. (Cronbach, 1971, p. 491)

Even with their difficulties grades may prove useful as outcome measures in decision problems. In most institutions grades represent success and are considered to a large degree to be indicators of subject matter competency. When grades are used as a utility measure the decision maker should use them cautiously. The decision maker should be satisfied that either his aim is to maximize grades or whatever it is that grades represent.

It is best whenever possible to use grades from a common course as the utility scale rather than using grades from various courses. For example, if students either begin with a regular course or take a remedial course prior to the regular course the most appropriate criterion would be the regular course grades after both groups have completed the regular course. If the remedial course grades were used for the one group and the regular course grades were used for the other group, the decision maker could not make the argument that a common criterion or learning objective was being considered.

Ideally, individuals should be randomly assigned to treatments when an attempt to determine aptitude treatment interactions is undertaken. Random assignment is seldom possible in educational systems. Generally, placement tests are developed because faculty wish either to place students into alternate treatments or to locate them at the appropriate level in a course sequence. The test is often used to place students before it is validated.

> One of the difficulties that is experienced in practical operation is that the decision to place students is often made before the research on how to place students soundly is finished, or even commenced. There is usually great reluctance to place all students randomly into the routine and alternate treatments for a year or two while the placement instrument is being evaluated. If that is not done, it is very hard to evaluate the instrument because the basic question to be asked is how well this instrument would predict in advance which students would have trouble with the routine course or would be able to handle a more

> (or less) advanced or accelerated or a differently presented course. (Hills, 1971, p. 703)

When random assignment can not be undertaken the investigator has several options. A second procedure involves partial randomization. Students with high scores are assigned to the higher level treatment while students with lower scores are assigned to the lower level treatment. Individuals with middle range scores are then randomly assigned to either the more or the less difficult course. Faculty will often accept this solution since there is often no clear justification for which treatment individuals should actually be in.

When partial randomization is not possible it may be necessary to extrapolate the regression lines to project where the interaction would take place. Extrapolation of the regression lines should be undertaken with caution. It is possible that the extended regression might be curvilinear even if the truncated line is linear. The variability around the extrapolated regression lines, of course, is open to question.

> If there is simply no alternative, one can carry partial randomization to its logical extreme - i.e., no randomization, or assignment of all students to one of two treatments depending on whether they score above or below a particular score on a placement test. In this case one plots the regression lines for the two treatment groups and extends those lines to determine the point of intersection. This method is much less stable and less persuasive, though it can provide useful information when the same objective criterion is used for both treatment

> groups and when students are assigned to the two groups strictly on the basis of the placement test. (Willingham, 1974, p. 30)

If students are not forced to enroll or remain enrolléd consistent with their placement test results there are likely to be individuals distributed along the score continuum in every treatment. It is possible to develop regression lines and investigate interactions in such situations. Caution should be used since individuals who self-place out of a course recommended for their placement test score may be different from those individuals who follow placement advice. However, the use of such a procedure may be useful in situations where random assignment can not be implemented. The appropriateness of such cutting scores can be checked by considering "hits" and "misses" when the scores are applied to independent groups.

<u>Summary</u>

There are several methods which may be used for setting cutting scores. Some methods are more appropriate than others. The practical situation that one must work within will often influence the process chosen to set cutting scores.

Decision Theory is conceptually powerful. If there is a sound educational basis for assigning students to different treatments then there should be evidence of differential outcomes. Students should be given the available treatment which is expected to maximize their chance of success. A variable which does not differentially predict success in different treatments has questionable use for placement.

While a number of methods for setting cutting scores are available there has been little research dealing with the evaluation of placement procedures. Even less research concerning placement and decision theory has been undertaken.

Difficulties in the application of decision theory to placement problems are partially responsible for the lack of sound research. The development of appropriate scales and problems with random assignment make the decision theory approach difficult to apply in pure form. Fortunately, well planned deviations from the technique may be undertaken and still result in useful conclusions.

Hills (1971) reports that problems with the Cronbach and Gleser decision model include: (1) the degree to which the model fits practical educational problems has not been determined, (2) the parameter values for the model are not sufficiently well known to make estimates in varied situations, and (3) there is no clear way to develop a utility scale on which costs and educational outcomes can be expressed. Even with its difficulties Hills (1971) recognizes the value of the theory.

> By this time so many problems have been posed and such an apparition of

> complexity raised that a natural response is to reject the decision theory approach and to retreat to the good old simple days. However, it appears to the author that those good old days never existed but were a figment of inadequate analysis. This is the reason that one can find almost no sound studies demonstrating the effectivness of placement as is is often naively done, with no consideration of the interaction between traits and treatments as the affect criteria. Once one accepts the correctness of the fundamental idea that different students can be taught most effectively through the use of different methods, the rest of decision theory rationale becomes inescapable. But this fundamental idea is just as basic to placement as it is currently done as it is to placement as it should be done taking advantage of decision theory as an intellectual tool. Thus there appears to be no retreat available. (Hills, 1971, p. 729)

Due to difficulties in applying the decision model to placement problems the approach may not generate expected interactions. If interactions are found it will be necessary to conduct additional research to determine if the interactions remain consistent when the placement rules that are generated from them are applied to other groups. It may be that in some situations the grouping of students by a new placement rule will affect treatment outcomes. In light of the inherent logic of the decision model for placement and the lack of substantive research it is necessary to further investigate placement procedures. Additional research in this domain will help to determine the practical usefulness of the model for placement.

Chapter 3: Design of the Study

Sample

The population from which a sample was drawn consisted of all students who entered Michigan State University as freshmen fall terms 1978 through 1980. Samples of students used for analysis were selected on the basis of enrollment in relevant Math course sequences. The analysis involved determining the appropriateness of the current cutting scores used with the placement test in terms of the psychometric decision model proposed by Cronbach and Gleser (1965), determining the usefulness of the American College Test Math subscore (ACTms) for placement decisions, and determining the equality of final performance of students completing the same levels of Math by completing different sequences of courses. For the analysis concerning the setting of cutting scores only freshman students who reported both American College Test Scores and Michigan State University Math placement test scores are considered.

Between 88% and 92% of the freshmen entering MSU during the selected terms reported ACT scores. All entering freshmen were required to take the placement test. The size of the groups entering with both ACT scores and placement test scores were: 1978 (n=4588), 1979 (n=5765), and 1980 (n=5298). Specific groups which were selected from the larger groups are defined below.

1. Students completing Math 108, 109, and 112 (long calculus) during their first three terms.

2. Student completing Math 111 and 112 (short calculus) during their first three terms.

3. Students completing Math 081, 082, and 108 (long algebra) during their first three terms.

4. Students completing Math 082, 108 (short algebra) during their first three terms but not taking Math 081.

The groups of students entering Michigan State University Fall 1978, 1979, and 1980 were highly similar in terms of ability and background. Information concerning new freshman students enrolling during the three year period of the study who, reported ACT information to MSU, are provided in Tables 3 through 7 below.

Table 3

Academic abilities of students entering MSU Fall 1978, 1979, and 1980 who reported ACT information.

Category	Fall 1978	Fall 1979	Fall 1980
ACT composite mean	22.0	22.0	22.0
ACT composite standard deviation	4.8	4.8	4.6
ACT Math mean	22.5	22.5	22.4
Mean self-reported high school GPA	3.29	3.28	3.29
High school GPA standard deviation	.48	.48	.48
Female high school GPA	3.32	3.31	3.32
Female GPA standard deviation	.48	.47	.48
Male high school GPA	3.25	3.24	3.24
Male GPA standard deviation	.48	.48	.47
Mean Math high school GPA	3.03	3.02	3.03
High school Math standard deviation	.84	.83	.82

Table 4

Degree aspirations and self-predicted college GPA of students entering MSU fall 1978, 1979, or 1980 who reported ACT information to MSU.

Degree Aspiration	Fall 1978	Fall 1979	Fall 1980
Vocational or Technical	1%	0%	0%
Two Year	3%	3%	2%
Bachelor's degree	39%	37%	39%
Master's Degree	21%	24%	23%
Doctorate/Professional Degree	34%	34%	34%
Predicted Grade Point Average			
Self-predicted first year GPA	3.1	3.1	3.1
Standard deviation - predicted GPA	0.4	0.4	0.4

Table 5

Home community size of students entering MSU Fall 1978, 1979, and 1980 who reported ACT information.

Category	Fall 1978	Fall 1979	Fall 1979
Farm or open county	10%	9%	8%
Less than 500 population	1%	1%	1%
500-1999 population	5%	5%	4%
2000-9999 population	16%	16%	17%
10000-49000 population	30%	28%	29%
50000-249999 population	24%	24%	26%
250000-499999 population	4%	4%	4%
500000-999999	2%	2%	2%
Over 1000000	5%	5%	5%
Not given	4%	4%	3%

Table 6

Racial or Ethnic background of students entering MSU fall 1978, 1979, and 1980 who reported ACT information.

Category	Fall 1978	Fall 1979	Fall 1980
Afro-American	5%	6%	7%
American Indian/Eskimo	1%	1%	0%
Caucasian American	75%	77%	86%
Mexican American	0%	0%	1%
Oriental American	1%	1%	1%
Puerto Rican or Spanish American	3%	3%	2%
Prefer not to respond	10%	9%	3%
Not given	6%	5%	1%

Table 7

Percentage of students entering MSU Fall 1978, 1979, and 1980 who reported ACT information who were interested in special Math programs.

Category	Fall 1978	Fall 1979	Fall 1980
Expecting to need special assistance in Math	28%	27%	26%
Interested in advanced placement in Math	25%	27%	26%
Wishing to receive credit by exam in Math	36%	36%	36%

Measures

Two instruments were used in this study: The Michigan State University Math placement test (MSUpt) and the American College Test Math subscore (ACTms). Both instruments are described below.

Michigan State University Math Placement Test All new freshmen entering MSU take a series of examinations which are referred to as the Orientation Tests. The results of the tests are used for academic advising, placement, admissions decisions, and determining ability patterns in groups of students. Tests are given in the areas of reading, arithmetic, Mathematics, chemistry, and foreign language.

The Mathematics placement test consists of 30 algebra type items and is administered with a 40 item basic arithmetic test. Forty minutes is allowed for the test. The Department of Mathematics designed the test specifically for placement of Michigan State University students into MSU Math courses in the early 1960's. Juola (1973) discusses the conditions leading to the development of the test.

> The initial rationale for establishing many of these diverse placement areas and for developing tests for each was governed by the high incidence of failure in several basic courses. The proportion of students getting F and D grades at one time ranged from 30-40% in many classes. When tested entry skills were considered, this ratio ranged up to 75% and more for students with low scores on relevant tests. (p. 1)

The KR20 reliability for the Mathematics test is approximately .89. The KR20 for the combined Math and arithmetic scores is .92. Validity coefficients in the form of correlations between Math placement scores and course grades are low to moderate. The coefficients are displayed in Table 8.

Scores for the orientation examinations are reported on each student's orientation record which is compiled and produced by computer. Students take their orientation records to their academic advisors when selecting their first term schedules. Scores are provided both in raw form and in percentiles which are reported graphically. Graphic reporting enables the student to visually perceive his ability relative to other entering students.

Table 8

Correlation coefficients relating MSU Math placement test scores to course grades for freshmen entering MSU fall 1979.

Course	r	Se	N
Math 081/103	.15	1.11	211
Math 082/104	.38	1.02	785
Math 108	.26	1.19	1653
Math 111	.33	.92	1258

<u>The American College Test</u> The American College Test (ACT) is a college entrance examination which consists of four parts: (1) English usage, (2) Mathematics usage, (3) Social Studies reading, and (4) Natural Science reading. An Interest Inventory and Personality Profile is administered in conjunction with the ACT.

The Math test which is considered in this investigation is 40 items long and has a 50 minute time limit. All items are multiple choice with 5 alternatives available. The test consists of algebra, arithmetic and plane geometry items. Very little "modern Mathematics" is included. Scale scores from 1-36 are reported and used in this study. The standard deviation of the score distribution is set at approximately 5.

Reliabilities for the ACT are calculated using odd/even procedures and range from the .70's through the .90's for the subscores and composite scores (Hills, 1978, in Buros _Eighth Mental Measurements Yearbook_, p 617). Since many students do not finish sections of the test within the time limits the reliability estimates may be spuriously high. Wallace (1972) reports that the test's validity seems appropriate.

> Validation of the ACT has been very extensive with consistently good results. There is a vigorous and ongoing program of checking the validity of the ACT for predicting college grades...Furthermore the predictive effectiveness of ACT scores is analyzed by course or subject-matter groupings within schools as well as by programs such as business, engineering, education, and art. This wide variety of situations, of course, yields a broad range of validity coefficients but it is estimated that the central tendency of the distribution of correlations between ACT composite scores and overall grade point averages is about .50. (Wallace in Buros _Seventh Mental Measurements Yearbook_, p. 614-616)

The median ACT Math score for entering MSU students is in the approximate range of 23-24. Students in the 90th percentile have scores slightly better than 29 while students in the 10th percentile tend to have scores slightly less than 14. ACT Math scores corresponding to selected percentiles are presented in Table 9.

Table 9

ACT Math scores corresponding to selected percentiles for students entering MSU fall 1975-1981.

%point	1975	1976	1977	1978	1979	1980	1981
p90	30.0	29.9	29.1	29.7	29.3	29.4	29.3
p75	27.6	27.6	27.0	27.3	27.3	27.1	26.8
p50	24.7	24.7	23.8	24.1	23.5	23.3	23.1
p25	18.5	18.2	17.9	18.4	18.6	18.9	18.6
p10	14.6	13.8	13.1	13.8	13.8	13.8	13.9

Design

This investigation was designed to assess the value of the MSU Math placement test and the currently used cutting scores for maximizing student outcomes as reflected by grades. The study involved the examination of an operating Mathematics placement procedure and an educational process. In order to answer the research hypothesis it was necessary to follow a series of branching steps. The results of intermediate steps at some points determined the direction that subsequent steps took.

Four course sequences served as treatments in this study. The sequences were: long calculus (Math 108, 109, 112), short calculus (Math 111, 112), long algebra (Math 081/103, 082/104, 108) and, short algebra (Math 082/104, 108). The general procedures which were followed for each hypothesis are described below.

Appropriateness of the current Mathematics placement test cutting scores for sectioning students into either the long or short calculus sequence.

1. Select students in the 1978 entering group who completed either the long calculus sequence or the short calculus sequence.

2. Develop regression lines for grades on placement test scores for students in each treatment.

3. Determine if the regression lines were linear or curvilinear.

4. Determine if there was an interaction between the regression lines.

5. If there was a disordinal interaction, determine if it occurred at the point where the current cutting score for sectioning students into either of the two treatments is set. If an interaction occurred within the range of the current cutting score then the current cutting score was the most appropriate score when the decision maker wished to maximize course grade outcomes.

6. If there was a disordinal interaction which did not occur at the point where the current cutting score was set determine the corresponding score on the placement test where the interaction took place. Select students in an independent sample, (1979 entering group) who enrolled in the treatment consistent with currently suggested policy and determine the proportion passing their

first course with a 2.00 or better. Select students in the independent sample who enrolled in Math consistent with the score suggested by the point of interaction and determine the proportion of students passing with a 2.00 or better. Determine if there is a significant difference in the proportion of students passing each sequence for the two groups.

7. If there was an ordinal interaction which did not take place within the range of the current cutting score or if there is a disordinal interaction, adjust the regression lines to determine the discrepancy between the course grade criterion and the actual utility value the institution placed on student outcomes.

8. If either regression line is curvilinear, determine the implications, if any, for placement procedures. If the curvilinear relationship is caused by a test ceiling determine if it affects the appropriateness of the current cutting scores.

Appropriateness of the current Mathematics placement test cutting scores for sectioning students into either the long or short algebra sequence.

1. Select students in the 1978 entering group who completed either the long or short algebra sequence.

2. Develop regression lines for grades on placement test scores for students in each treatment.

3. Determine if the regression lines were linear or curvilinear.

4. Determine if there was an interaction between the regression lines.

5. If there was a disordinal interaction, determine if it occurred at the point where the current cutting score for sectioning students into either of the two treatments was set. If an interaction occurred within the range of the current cutting score then the current cutting score was the most appropriate score when the decision maker wished to maximize course grade outcomes.

6. If there was a disordinal interaction which did not occur at the point where the current cutting score is set determine the corresponding score on the placement test where the interaction took place. Select students in an independent sample, (1979 entering group) who enrolled in the treatment consistent with currently suggested policy and determine the proportion passing their first course with a 2.00 or better. Select students in the independent sample who enrolled in Math consistent with the score suggested by the

point of interaction and determine the proportion of students passing with a 2.00 or better. Determine if there was a significant difference in the proportions of students passing each sequence for the two groups.

7. If there was an ordinal interaction which did not take place within the range of the current cutting score or if there was a disordinal interaction, adjust the regression lines to determine the discrepancy between the course grade criterion and the actual utility value the institution placed on student outcomes.

8. If either regression line was curvilinear, determine the implications, if any, for placement procedures. If the curvilinear relationship was caused by a test ceiling determine if it affects the appropriateness of the current cutting scores.

Usefulness of the American College Test Math subscore for sectioning students into either the long calculus or the short calculus sequence.

1. Select students from the 1978 entering class who completed either the long or short calculus sequence.

2. Determine the regression lines relating course grades and American College Test Math subscores for students in each treatment.

3. Determine if there was no interaction, a disordinal interaction, or an ordinal interaction between the regression lines for the two treatments.

4. If there is no interaction or a ordinal interaction then the ACTms is not appropriate for sectioning students into the alternate treatments when common course grades are used as the utility criterion.

5. If a disordinal interaction occurred determine the score on the ACTms corresponding to the point of interaction. The point corresponding to the interaction of the regression lines will be the optimal cutting score when the ACTms is used for sectioning students and course grades are the criterion of interest. Apply the ACTms cutting score to an independent sample of students (1979 entering group) and determine the proportion of students completing the sequence successfully when the cutting score indicated by the interaction is used. Determine the proportion of students who successfully completed the sequence when they enrolled consistent with the currently used MSUpt. Determine if both tests and corresponding cutting scores result in equal proportions of successful treatments.

6. If the use of the ACTms subscore resulted in an equal or greater proportion of successful placement than does the MSUpt cross validate the ACT cutting score using the 1980 entering group.

<u>Students who enroll in a sequence consistent with their placement test scores and placement policy will not perform differently as a group than will students who enroll in a course higher than that suggested for them by current placement policy.</u>

1. Divide students who entered during the fall of 1978 into groups on the basis of the Math course they completed during their first term at Michigan State University (Math 081, 082, 108, 109, 111, or 112)
2. Within each Math course select students who enrolled consistent with placement policy and students who were enrolled in the course at a higher level than suggested by placement policy.
3. Determine if there is a significant difference in mean outcomes for individuals enrolling consistent with placement policy and individuals enrolling in a course higher than suggested by placement policy.

<u>Students completing either the long calculus sequence or the short calculus sequence will display the same level of achievement as expressed by Math 112 final grades.</u>

1. Select students from the 1978 entering group who completed either long calculus sequence or the short calculus sequence.

2. Determine if there is a significant difference in final Math 112 course grades for the two groups.

Research Hypotheses

Question 1 Are the current cutting scores used with the Michigan State University Math placement test for sectioning students into either the long calculus sequence or the short calculus sequence the scores which will maximize student outcomes as measured by final Math 112 grades?

> Null Hypothesis: The regression lines for Math 112 course grades on Math placement test scores for students in either treatment (long calculus or short calculus) will intersect in the range of the current cutting score of 17 (range 16.5-17.5).

> Alternative Hypothesis The regression lines for Math 112 course grades on Math placement test scores for students in either treatment (long calculus or short calculus) will not intersect in the range of the current cutting score of 17 (range 16.5-17.5).

Question 2 Are the current cutting scores used with the Michigan State University Math placement test for sectioning students into either the long algebra sequence or the short algebra sequence the scores which will maximize student outcomes as measured by final Math 108 grades?

> Null Hypothesis: The regression lines for Math 108 course grades on Math placement test scores for students in either treatment (long algebra or short algebra) will intersect in the range of the current cutting score of 7 (range 6.5-7.5).

Alternative Hypothesis The regression lines for Math 108 course grades on Math placement test scores for students in either treatment (long algebra or short algebra) will not intersect in the range of the current cutting score of 7 (range 6.5-7.5).

Question 3 Is there an aptitude treatment interaction between the ACT Math subscore and the treatments (long calculus or short calculus) which can be used to set a placement cutting score?

Null Hypothesis: There will be no significant interaction between the regression lines for each treatment (long calculus or short calculus).

Alternate Hypothesis: There will be a significant interaction between the regression lines for each treatment (long calculus or short calculus).

Question 4 Will students who enroll consistent with MSU placement policy in relation to their initial placement test score perform at the same level as students who enroll in a course at a higher level than recommended for their placement test score?

Null Hypothesis: The mean course grade for students enrolling in a Math course consistent with their placement test score will not be significantly different from the mean course grade of students enrolling in a higher level course than is recommended by placement policy for their placement test score (each course will be analyzed separately: Math 081/103, 082/104, 108, 111, 112).

Alternate Hypothesis: The mean course grade for students enrolling in a Math course consistent with their placement test score will be significantly higher than the mean course grade of students enrolling in a higher level course than is recommended by placement policy for their placement test score (each course will be analyzed separately: Math 081/103, 082/104, 108, 111, 112).

Question 5 Do students completing the long calculus sequence perform the same in Math 112 as students completing the short calculus sequence?

Null Hypothesis: The mean Math 112 course grade for students taking the long calculus sequence will not be significantly different than the mean Math 112 course grade for students taking the short calculus sequence.

Alternative Hypothesis: The mean Math 112 course grade for students taking the long calculus sequence will be significantly different than the mean Math 112 course grade for students taking the short calculus sequence.

Analysis

Three models are used in the analysis of the data: a linear regression model, an analysis of covariance model, and a Z test for the equality of two means. The Johnson-Neyman Technique is applied in order to determine where regression lines intersect and where regions of significant differences between regression lines occur.

Linear regression is a technique commonly used for the prediction of one variable from another variable or from several other variables and to demonstrate the degree of relationship between two variables. A predictor or independent variable is used to predict a criterion or dependent variable. The regression line provides the best prediction of the dependent variable from knowledge of the independent variable and results in the smallest possible sum of squared deviations of predicted scores from obtained scores. The assumptions of the model are that the predictor and criterion variables are continuous and have a joint bivariate normal distribution.

In this study the Michigan State University Math placement test scores and the American College Test Math subscores were treated as independent or predictor variables while Math course grades served as dependent or criterion variables. Common course grades were predicted from test scores for groups given alternate treatments. If the predictor variables are useful for differential placement into alternate treatments the regression lines for comparison groups can be expected to intersect. The point of intersection will indicate the division point which when used to assign students to treatments will result in the greatest average benefit.

Although regression lines may intersect they are not likely to be significantly different for all values of the predictor variable. The Johnson-Neyman technique (Walker and Lev, 1953, p. 401-404) was used to determine the point at which two regression lines intersect and to determine for which values of the independent variable the predicted criterion values are significantly different. The technique was used in this investigation to determine the range within which cutting scores may legitimately fall.

Analysis of covariance is used to test for significant differences between groups. It differs from analysis of variance in that it takes into account possible effects of an independent variable or covariate on the dependent variable. Analysis of covariance is useful for determining if there are significant outcome differences after initial differences in the groups have been taken into account. The basic question answered by the technique is: do initial differences account for criterion differences?

Three types of hypothesis under the analysis of covariance model are considered in this study. The first hypothesis concerns the equality of slopes of the regression lines for comparative treatments. If the slopes of the regression lines are equal then no aptitude treatment interaction occurs. The second hypothesis concerns whether the common slope is significantly different from 0. The third hypothesis concerns whether both treatments share a common regression line. If they do share a common regression line it indicates that all outcome differences can be explained by initial covariate differences. If the treatments do not share a common regression line than it is accepted that treatment differences affected the outcome differences.

The Z test for the equality of two means is used to determine if individuals in this study who complete different Math sequences which end in a common course perform the same in the common course. Alternate sequences which are intended to culminate in comparable outcomes are considered. The assumptions underlying the Z test are random sampling and normal populations. The students in this study were not randomly selected. However, examination of their characteristics indicate that they are fair representatives of similar students entering Michigan State University.

Summary

This study encompassed fitting the placement decision model developed by Cronbach and Gleser to the Mathematics placement process at Michigan State University, determining the usefulness of the American College Test Math subscore for setting cutting scores, and comparing the common outcomes of individuals assigned to different educational sequences. In order to accomplish these ends, regression analysis was applied to predictor and criterion data for groups of students completing the respective treatments. Aptitude treatment interactions would indicate the predictor scores which would be most beneficial for sectioning students into alternate treatments.

The usefulness of the American College Test Math subscore for placement was examined in terms of the decision model. If the ACTms were useful for placement students who differ in ability as measured by the ACTms should perform differently in the alternate treatments which culminate in a common criterion. If appropriate cutting scores can be set for the ACTms in relation to decision theory the ACTms may be more beneficial for placement than the MSUpt since ACTms scores are available from the admissions record for most students.

One of the difficulties of applying psychometric decision theory is that it may be difficult or impossible to construct utility scales on which all relevant factors may be considered concurrently. In aptitude treatment interaction studies grades are often used as the utility measure. Grades do not reflect the cost of testing or the relative costs of different treatments.

If aptitude treatment interactions do not occur at the point at which cutting scores are set it is likely that either the utility scale does not reflect the true values of the decision maker or that the decision maker is not acting consistent with his stated beliefs. By adjusting the regression lines for a pair of treatments so that they intersect at the point at which the cutting score is set (assuming that they do intersect at some point) it is possible to determine the magnitude of differences between policy and practice, in terms of the criterion, which is being considered.

Students who are assigned to educational treatment on the basis of their placement scores but are expected to reach common educational objectives should perform in a similar fashion on the relevant common criterion measure. In this study the mean common course grades for students taking alternate treatments which are intended to have a common end point are compared.

Chapter 4: Analysis of Results

Hypothesis 1: The regression lines for Math 112 course grades on Math placement test scores for students in either the long calculus sequence (Math 108, 109, 112) or the short calculus sequence (Math 111, 112) will intersect in the range of the current cutting score of 17 (range = 16.5-17.5).

The regression lines for the long calculus sequence and the short calculus sequence for the 1978 entering group did not intersect in the range of the of the current cutting score. The sample lines for the 1978 group intersected at 27.17 but there was no significant difference between the slopes of the regression lines for the two treatments. The slopes of the regression lines were significantly different from 0. The two treatment groups did not share a common regression line and the two treatments produced significant differences in the mean outcome for the groups after adjustments for initial ability had been made. Illustrations of the regression lines for all pairs of treatments are provided in Figures 14-21 in Appendix A.

A similar analysis was performed on data collected on the 1979 entering group to determine if a significant interaction would be found. The sample regression lines intersected at 3.01, however, no significant differences in the slopes of the treatment regression lines were found. Both treatment regression lines were significantly different from 0. The treatments resulted in significant differences in outcome after adjustments for initial differences in ability were made. Table 10 displays the analysis of covariance significance tests for both the 1978 and the 1979 entering groups.

Table 10

Analysis of covariance results comparing the long calculus sequence and the short calculus sequence for both the 1978 and 1979 entering groups with common terminal course grades regressed on Math placement test scores.

Long vs short calculus	N	F common slope	F slope = 0	F single line	F.95
1978	820	.591	27.474*	6.497*	3.84
1979	559	.219	27.501*	4.113*	3.84

*p<.5

The means of the independent variable, the dependent variable, and the dependent variable adjusted for initial Math placement test score differences are displayed in Table 11. Treatment differences were found for both years.

There was a reversal in the treatment which produced the superior result between the 1978-79 and the

1979-80 groups. During the 1978-79 academic year students taking the calculus sequence did better in the short course

Table 11

Comparison of the adjusted means for individuals taking either the short calculus sequence or the long calculus sequence 1978-79 or 1979-80.

	N	Mean X	Mean Y	Adjusted Mean Y
Short calculus 78 vs	645	18.934	24.868	24.577*
Long calculus 78	175	15.709	21.171	22.246*
Short calculus 79 vs	394	20.201	20.393	19.145*
Long calculus 79	165	15.909	18.970	21.951*

*$p < .05$ for mean differences between pairs
Note. Mean criterion grades (Y) are reported on a 0-40 scale or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

than in the long course after criterion measures were adjusted for differences in Math placement test scores. Students taking the long calculus course outperformed those taking the short calculus course during the 1979-80 year after adjustments for placement test differences were made. Table 12 displays the analysis of covariance results when the same sequences are compared over years. In this case sequence is constant and the year constitutes the treatment.

Table 13 displays the mean criterion scores for students receiving the same calculus treatment during different years adjusted for initial placement test differences. Students in the short sequence in 1978 outperformed students in the short sequence 1979 after placement test adjustments. No significant treatment differences were found for the long calculus sequence over years.

Table 12

Analysis of covariance results for groups receiving the same calculus treatment during a different year with terminal course grades regressed on Math placement test scores.

	N	F common slope	F slope = 0	F single line	F.95
Short cal. 78 Short cal. 79	1039	6.127*	36.983*	53.067*	3.84
Long cal. 78 Long cal. 79	340	1.250	12.911*	3.376	3.84

*$p<.05$

Table 13

Comparison of the adjusted means for individuals taking each calculus sequence across years.

	N	Mean X	Mean Y	Adjusted Mean Y
Short calculus 78	645	18.935	24.868	25.106*
Short calculus 79	394	20.201	20.393	20.004*
Long calculus 78	175	15.709	21.171	21.252
Long calculus 79	165	15.909	18.970	19.007

*$p<.05$ for mean differences between pairs.
Note. Mean criterion grades (Y) are reported on a 0-40 or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

Hypothesis 2: The regression lines for Math 108 course grades on Math placement test scores for students in either the long algebra sequence (Math 081/103, 082/104, 108) or the short algebra sequence (Math 082/104, 108) will intersect in the range of the current cutting score of 7 (range = 6.5-7.5).

The regression lines for the long algebra sequence and the short algebra sequence for the 1978 entering group intersected at 14.50. However, the slopes of the two regression lines were not significantly different indicating that the intersection was due to sampling error (see Table 14). The common slope of the regression lines were not significantly different from 0 indicating that the Math placement test was not useful for predicting within treatment outcome. The criterion means of the two regression lines were significantly different after adjustments for initial placement test differences were made.

The analysis was replicated on data collected from the 1979 entering group. The sample regression lines intersected at -2.88. However, there was no significant difference between the slopes of the regression lines, thus indicating that the interaction was not significant. The slopes of the 1979 group regressions, like the 1978 group, were not significantly different from 0. The treatments resulted in significant differences in mean outcome after adjustments for initial differences in ability were undertaken.

The means of the independent variable, the dependent variable and the dependent variable adjusted for initial Math placement test differences are displayed in Table 15.

Table 14

Analysis of covariance results comparing the long algebra sequence with the short algebra sequence for both the 1978 and 1979 entering groups with common terminal course grades regressed on Math placement test scores.

Long vs short algebra	N	F common slope	F slope = 0	F single line	F.95
1978	455	.783	.715	16.042*	3.84
1979	616	.340	1.000	6.587*	3.84

*$p < .05$

Table 15

Comparison of the adjusted means for individuals taking either the long algebra sequence or the short algebra sequence 1978-79 or 1979-80.

	N	Mean X	Mean Y	Adjusted Mean Y
Short algebra 78 vs	423	10.683	22.754	22.679*
Long algebra 78	32	5.875	11.095	12.082*
Short algebra 79 vs	579	10.675	19.076	19.011*
Long algebra 79	37	6.541	11.892	12.900*

*$p < .05$ for mean differences between pairs
Note. Mean criterion grades (Y) are reported on a 0-40 point scale or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

There was no reversal in the algebra treatment which produced the superior result over the two years of the investigation. During both years students taking the short algebra sequence outperformed the students taking the long algebra sequence after corrections for initial Math placement test differences were made. Adjustments using the placement test as a covariate were minor since the slopes of the regression lines were not significantly different from 0. Table 16 displays the analysis of covariance results when the same algebra sequences are compared over years. In this case sequence is constant and the year administered constitutes the treatment.

Table 16

Analysis of covariance results for groups receiving the same algebra treatment during a different year with terminal course grades regressed on Math placement test scores.

	N	F common slope	F slope = 0	F single line	F.95
Short alg. 78 Short alg. 79	1002	.102	1.555	20.523*	3.84
Long alg. 78 Long alg. 79	69	1.245	.168	.043	3.99

*p<.05

Table 17 displays mean criterion scores for students receiving the same algebra sequence during different years adjusted for initial Math placement test differences. Students in the short algebra sequence in 1978 outperformed students in the short algebra sequence in 1979 after adjustment for placement test differences. No significant differences were found for the long algebra sequence over years.

Table 17

Comparison of the adjusted means across years for individuals taking each algebra sequence using the Math placement test score as a covariate.

	N	Mean X	Mean Y	Adjusted Mean Y
Short alg. 78	423	10.683	22.754	22.753*
Short alg. 79	579	10.675	19.076	19.077*
Long alg. 78	32	5.875	11.093	11.206
Long alg. 79	37	6.541	11.892	11.794

*$p<.05$ for mean differences between pairs
Note. Mean criterion grades (Y) are reported on a 0-40 scale or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

Hypothesis 3: The regression lines for Math 112 course grades on the American College Test Math subscore for students in either the long calculus sequence (Math 108, 109, 112) or the short calculus sequence (Math 111, 112) will intersect at a point which is appropriate for sectioning students into the two treatments.

The sample regression lines for Math 112 grades on the American College Test Math subscore, for the 1978 group, intersected outside the score range at the extrapolated score of 44.40. However, the slopes of the regression lines were not significantly different indicating that the interaction was not significant. The common slope was significantly different from 0. The two regressions did not share a common line indicating that mean criterion differences existed (see Table 18).

The analysis was replicated on the 1979 entering group. The sample regression lines intersected at a point corresponding to a Math placement test score of 3.01 but the slopes of the lines were not significantly different. The common slope was significantly different from 0. The hypothesis that the two regression lines shared a common line was accepted indicating that there was no difference in mean outcome for the two treatments after adjustments for initial placement test differences (see Table 18).

Table 18

Analysis of covariance results comparing the long calculus sequence and the short calculus sequence for both the 1978 and 1979 entering groups with common terminal course grades regressed on ACT Math scores.

Long vs short calculus	N	F common slope	F slope = 0	F single line	F.95
1978	820	.454	24.595*	9.803*	3.84
1979	559	.354	7.537*	.223	3.84

*p<.05

When ACT Math scores were used as a covariate, treatment differences were found between the long and short calculus sequences for the 1978 group. Similar differences were not found for the 1979 group. The adjusted criterion means using ACT Math scores as the covariate are displayed in Table 19.

Table 19

Comparison of the adjusted means for individuals taking either the long calculus sequence or the short calculus sequence 1978-79 or 1979-80 using ACT Math scores as a covariate.

	N	Mean X	Mean Y	Adjusted Mean Y
Short calculus 78 vs	645	26.012	24.868	24.678*
Long calculus 78	175	23.800	21.171	21.872*
Short calculus 79 vs	394	26.091	20.393	20.139
Long calculus 79	165	23.879	18.970	19.578

*$p < .05$ for mean differences between pairs
Note. Mean criterion grades (Y) are reported on a 0-40 point scale or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

Analysis of covariance was performed comparing each individual sequence across years. In this case, year served as the treatment and the sequence was held constant. The regression lines for the short calculus sequence over treatment years did not intersect but the lines were significantly different indicating that the short calculus sequence produced different results for the two years (see Table 20). Students taking the short calculus sequence in 1978 did better than students taking the short calculus sequence in 1979 after adjustments were made for initial ACT Math score differences.

There were no treatment by year effects for students taking the long calculus sequence either during 1978-79 or 1979-80 (see Table 21). There were no differences in the regression lines for the long calculus sequence across years. The criterion scores were the same for individuals taking the long calculus sequence either year after adjustments were made for initial ACT Math score differences.

Table 20

Analysis of covariance results for groups receiving the same calculus treatment during a different year with terminal course grades regressed on ACT Math scores.

	N	F common slope	F slope = 0	F single line	F.95
Short cal. 78 Short cal. 79	1039	.062	20.048*	41.599*	3.84
Long cal. 78 Long cal. 79	340	.000	10.303*	3.010	3.84

*$p<.05$

Table 21

Comparison of the adjusted means for individuals taking each calculus sequence across years using the ACT Math score as a covariate.

	N	Mean X	Mean Y	Adjusted Mean Y
Short cal. 78	645	26.012	24.868	24.879*
Short cal. 79	394	26.091	20.393	20.376*
Long cal. 78	175	23.800	21.171	21.191
Long cal. 79	165	23.879	18.970	18.948

*$p<.05$ for mean differences between pairs
Note. Mean criterion grades (Y) are reported on a 0-40 scale or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

Analysis of covariance was performed comparing the short and long algebra sequences for the entering groups of

1978 and 1979. In 1978 the regression lines for the long and short algebra sequences shared a common line and had a common slope (see Table 22). The slopes were significantly different from 0. The means of the regression lines were significantly different indicating treatment differences across years.

For the 1979 entering groups the long and short algebra treatment group regression lines did not share a common slope and did not have equal criterion means. The lines intersected at the point corresponding to an ACT Math score of 11.13. The significant disordinal interaction indicates an aptitude treatment interaction for the 1979 entering group (see Table 22).

Table 22

Analysis of covariance results comparing the long algebra sequence with the short algebra sequence for both the 1978 and 1979 entering groups with common terminal course grades regressed on ACT Math scores.

	N	F common slope	F slope = 0	F single line	F.95
Short alg. 78 Long alg. 78	456	2.099	4.991*	18.237*	3.84
Short alg. 79 Long alg. 79	617	7.880*	18.185*	4.838*	3.84

*$p < .05$

The long and short algebra criterion means were compared for each entering group using the ACT Math score as a covariate (see Table 23). The treatments resulted in significantly different mean outcomes, after adjustment for differences in initial ACT Math scores, for the entering groups during both years. Students taking the short algebra sequence consistently outperformed students taking the long algebra sequence after adjustments were made.

Table 23

Comparison of the adjusted means for individuals taking either the long algebra sequence or the short algebra sequence 1978-79 or 1979-80 using ACT Math scores as a covariate.

	N	Mean X	Mean Y	Adjusted Mean Y
Short alg. 78 vs	424	19.457	22.700	22.606*
Long alg. 78	32	14.594	11.094	12.351*
Short alg. 79 vs	580	19.524	19.043	18.901*
Long alg. 79	37	14.595	11.892	14.120*

*$p<.05$ for mean differences between pairs
Note. Mean criterion grades (Y) are reported on a 0-40 scale or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

Each sequence was compared individually across years. Sequence was held constant and the year that it was administered served as the treatment. Regression lines for the short algebra sequence across years had common slopes but did not result in common criterion means. The short algebra sequence produced different treatment effects depending upon the year it was administered (see Table 24).

Table 24

Analysis of covariance results for groups receiving the same algebra treatment during a different year with terminal course grades regressed on ACT Math scores.

	N	F common slope	F slope = 0	F single line	F.95
Short alg. 78 Short alg. 79	1004	1.834	27.664*	21.161*	3.84
Long alg. 78 Long alg. 79	69	.361	3.443	.087	3.99

*$p < .05$

The criterion means for each individual algebra sequence were adjusted for initial ACT Math score differences and compared across years (see Table 25).

Table 25

Comparison of the adjusted means for individuals taking each algebra sequence across years using the ACT Math score as a covariate.

	N	Mean X	Mean Y	Adjusted Mean Y
Short alg. 78	424	19.458	22.700	22.718*
Short alg. 79	580	19.524	19.043	19.030*
Long alg. 78	32	14.594	11.094	11.094
Long alg. 79	37	14.595	11.891	11.892

*$p < .05$ for mean differences between pairs
Note. Mean criterion grades (Y) are reported on a 0-40 scale or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

The 1978 short algebra group outperformed the 1979 short algebra group after adjustments for initial ACT Math score differences were made. Students taking the long algebra sequence achieved the same criterion mean scores across years after adjustments for initial ACT Math score differences were made.

Hypothesis 4: The mean course grade for students enrolling in a Math course consistent with their placement test score will not be significantly different from the mean course grade of students enrolling in a higher level course than is recommended by placement policy for their placement test score.

Students who entered the university fall term 1981 were selected and grouped on the basis of the Math course (within the placement policy group) in which they enrolled during their first term at the institution. Within Math classes students were sorted by whether the class was higher than that recommended for their Math placement test score or appropriate for that recommended for their placement score. The mean outcomes of students enrolling in a higher than recommended class was compared with the mean outcomes of students enrolling in the class at the proper level.

The variances of the grade outcomes were compared for the two groups in each class (see Table 26). No significant differences in variances were found. The T test for large samples, assuming equality of variance, was performed on the subgroups within each class to determine if there were mean grade differences. Data for Math

081/103 were not analyzed since no students can enroll in the class at a level higher than is recommended for their placement test score (i.e. the score range is 0-7).

Table 26

Comparison of the equality of course grade variance for freshmen Fall 1981 enrolling in a Math course higher than recommended for their placement test score (incorrect) and students enrolling in the recommended course (correct).

Course	N	Variance	F	F.99
Math 082/104				
Incorrect	32	1.659		
Correct	545	1.503	1.103	1.47
Math 108				
Incorrect	216	1.381		
Correct	847	1.409	1.021	1.28
Math 111				
Incorrect	116	1.309		
Correct	827	1.339	1.023	1.43
Math 112				
Incorrect	167	1.570		
Correct	431	1.454	1.079	1.42

*$p < .01$

Students enrolling in Math 082/104 did not perform significantly different regardless of whether they enrolled above their placement test level or enrolled at their placement test level (see Table 27). Students who enrolled correctly in Math 108, 111, or 112 consistent with their placement test score did do significantly better than students who enrolled in the same classes but had placement test scores which indicated they should be in a lower class. In each case the mean grade was higher for the group enrolling properly than it was for the group which enrolled in a course higher than recommended for their placement test score.

Table 27

Results of the T significance test for mean course grade differences for Fall 1981 freshmen enrolling in a higher course than recommended for their placement test score (Incorrect) and enrolling the course recommended for their placement test score (Correct).

Course	N	Average GPA	T	T.95
Math 082/104				
Incorrect	32	1.703		
Correct	545	1.957	-1.136	1.96
Math 108				
Incorrect	216	1.431		
Correct	847	1.720	-3.201*	1.96
Math 111				
Incorrect	116	1.431		
Correct	827	2.051	-5.412*	1.96
Math 112				
Incorrect	167	2.000		
Correct	431	2.564	-5.075*	1.96

*$p < .01$

Hypothesis 5: The mean Math 112 course grade for students taking the long calculus sequence will be equal to the mean Math 112 course grade for students taking the short calculus sequence.

Freshmen students entering the university Fall 1978 who completed either the long calculus sequence or the short calculus sequence were selected and compared in terms of their mean outcomes in the terminal Math 112 class. A T test was used to compare the mean outcomes of the groups taking the alternate sequences which were intended to culminate in the common terminal Math 112 course. There was a significant difference in the Math 112 means for the groups (see Table 28). Students taking the short calculus sequence had a mean Math 112 grade of 2.49 while students taking the long sequence had a mean Math 112 grade of 2.12. A Similar analysis was performed for 1979 entering students who completed either the long or short calculus sequence. No significant differences in Math 112 means were found for the 1979 entering group.

Table 28

Comparison of the mean terminal course grade outcomes for students completing different sequences which had a common terminal course.

Sequence	N	Mean Grade	T	T.95
1978				
Short Calculus	645	2.487		
Long Calculus	175	2.117	-4.155*	1.96
Short Algebra	423	2.275		
Long Algebra	32	1.109	-5.009*	1.96
1979				
Short Calculus	394	2.039		
Long Calculus	165	1.897	-1.232	1.96
Short Algebra	579	1.905		
Long Algebra	37	1.189	-3.369*	1.96

*$p < .05$

The mean outcomes for students taking either the long algebra sequence or the short algebra sequence were compared in relation to the common terminal Math 108 grade. For both entering groups the algebra treatments resulted in significant mean differences in the common terminal Math 108 grade (see Table 28). The mean Math 108 grades for students taking short algebra sequence were higher than the mean grades for students taking the long sequence during both years of the study.

Additional Analyses

Linearity of Regression The linearity of regression of terminal course grades on Math placement test scores was examined (see Table 29). In the majority of cases there was no evidence of significant curvilinearity. The two exceptions were the short calculus sequence during the 1978 school year and the short algebra sequence during the 1979 school year. The plots of the mean grade for each placement test score are displayed in Figures 6-13 in Appendix A.

Restricted Ranges The regression lines for students enrolling in alternate treatments consistent with placement policy were compared to determine if discontinuities in regression lines occurred at the current cutting scores. Only students who completed the sequence recommended for their placement test scores were considered. No discontinuities were found for either the calculus or algebra sequences. The regression lines for corresponding treatments had common slopes (see Table 30).

Table 29

Tests for linearity of the regression of terminal course grades on placement test scores.

	1978		1979	
	N	F	N	F
Long Algebra	31	0.23	36	1.97
Short Algebra	423	0.94	579	5.51*
Long Calculus	174	0.13	164	1.28
Short Calculus	645	8.28*	393	0.06

*$p < .05$

Table 30

Analysis of covariance results for groups enrolled consistent with placement policy using Math placement test scores as the covariate.

	N	F common slope	F slope = 0	F single line	F.95
Long cal. vs Short cal.					
1978	613	.231	27.693*	.442	3.84
1979	474	2.494	20.282*	5.382*	3.84
Long alg. vs Short alg.					
1978	425	.183	.630	14.186*	3.84
1979	591	2.035	.766	5.226*	3.84

*$p < .05$

There were treatment differences between the long and short calculus sequences in 1979 and differences between the long and short algebra sequences during both years of the study. Students in the long calculus sequence outperformed students in the short calculus sequence in 1979 after adjustments for initial placement differences were made (see Table 31). Students in the short algebra sequence outperformed students in the long algebra sequence after adjustments for placement test differences during both years.

Table 31

Comparison of the adjusted means for individuals enrolled consistent with placement policy in corresponding treatments using the Math placement test as a covariate.

	N	Mean X	Mean Y	Adjusted Mean Y
Short calculus 78 vs	472	21.229	25.567	24.242
Long calculus 78	141	14.624	20.532	25.269
Short calculus 79 vs	340	20.924	21.029	18.981*
Long calculus 79	134	15.022	18.507	23.705*
Short algebra 78 vs	398	10.980	22.827	22.742*
Long algebra 78	27	5.444	10.185	11.432*
Short algebra 79 vs	561	10.815	19.135	19.071*
Long algebra 79	30	5.800	11.500	12.713*

*$p < .05$ for mean differences between pairs
Note. Mean criterion grades (Y) are reported on a 0-40 point scale or 10 times the normal 1-4 point scale (where 1=D, 2=C, 3=B, and 4=A).

Predictive Validity The correlations between the Math placement test and individual class grades are moderate to low (see Table 32). Variation in Math Placement Test scores only predict from 2% to 14% of the variation in class grades for the Math classes in the placement policy domain.

The terminal course grades of fall 1979 freshmen completing one of the course sequences were correlated with the MSU Math placement test and the ACT Math subscore. The Math placement test was more positively correlated with terminal course grades than was the ACT Math subscore for the long algebra sequence, the long calculus sequence, and the short calculus sequence (see Table 33).

Table 32

The predictive validity coefficients for the Math Placement Test and Math classes.

Class	r	r sq	Standard Error
Math 081/103	.15	.02	1.11
Math 082/104	.38	.14	1.02
Math 108	.26	.07	1.19
Math 111	.33	.11	.92
Math 112	.29	.08	1.00

Table 33

Comparison of the correlations between the final grade in each course sequence for both the Math Placement Test score and the ACT Math Score for freshmen 1979.

Sequence	MSUpt r	MSUpt r sq	ACTms r	ACTms r sq
Long Algebra	-.06	.00	-.29	.08
Short Algebra	.04	.00	.20	.04
Long Calculus	.23	.05	.17	.03
Short Calculus	.21	.04	.10	.01

The correlation between terminal grades and ACT Math subscores were higher than the correlation between placement test scores and terminal grades for students completing the short algebra sequence. The variation in placement test scores or the ACT Math subscores did not explain over 8% of the variation in outcomes within course sequence. The placement test correlations for within course sequences tended to be lower than the correlations for all students within in a particular class.

The grade in the first class in a sequence was correlated with the terminal grade in the sequence for both years of the study. The correlations between grades within a sequence were moderate. The variation in the first grade in the sequences predicted between 10% and 35% of the variation in the terminal grade (see Table 34).

Table 34

The correlation between the first grade and the terminal grade in each Math sequence for entering freshmen fall 1978-79 and 1979-80.

	1978		1979	
Sequence	r	r sq	r	r sq
Long Algebra	.44	.19	.31	.10
Short Algebra	.54	.29	.54	.29
Long Calculus	.50	.25	.56	.31
Short Calculus	.47	.22	.59	.35

Students Following Placement Policy Not all students follow placement policy. The proportion of students who followed placement policy and completed the sequences under study are provided in Table 35. Placement policy was more rigidly enforced during the 1979 year than it was during the 1978 year. In 1978 13% of the freshmen students completing the short calculus sequence would have enrolled in a lower level sequence had they followed placement policy. In 1979 only 1.5% of the students completing the short calculus sequence had begun at a point higher than recommended by placement policy.

Table 35

The percentages of entering freshmen completing Math sequences who enrolled in a sequence other than the one recommended for their placement test score by placement policy.

	Enrolling Above Recommendation		Enrolling Below Recommendation	
Sequence	1978	1979	1978	1979
Long Algebra	*	*	15.6%	18.9%
Short Algebra	5.9%	2.9%	7.1%	2.8%
Long Calculus	21.1%	17.0%	19.4%	18.8%
Short Calculus	13.0%	1.5%	12.1%	8.9%

*not applicable

The proportion of students enrolling in their first course consistent with placement policy, regardless whether or not they completed the sequence their first year was also considered. Data concerning the proportions of 1981 entering students beginning in the appropriate course during their first term are provided in Table 36. Many students do not complete their Math sequence during their first year.

Table 36

The proportion of entering freshmen fall 1981 who enrolled in a Math course consistent with placement policy during their first term.

	Enrolled Too High		Enrolled Properly		Enrolled Too Low	
	N	%	N	%	N	%
081/103	*	---	138	84.1	26	15.9
082/104	32	5.4	545	92.1	15	2.5
108	216	15.2	847	59.7	355	25.0
111	116	10.7	827	75.9	146	13.4
112	167	27.9	431	72.1	*	---

*not applicable.

Math Grade Distributions   The mean Math grades for entering freshmen during the two years of the study are presented in Table 37. During the 1978 academic year there was a tendency for the average grade to increase as the level of the Math class increased for enrolled freshmen. The tendency did not consistently occur during the 1979 academic year. In general, the mean grades for freshmen in each class were lower during the second year of the study than during the first year of the study.

The mean Math grades for all students regardless of class level are displayed in Table 38. The decline in the mean grades from 1978 to 1979 in the freshman only grades are not reflected in the mean grades for all students. Grades for all students include many repeats of the classes and grades from upper classmen beginning their Math studies.

Table 37

The mean Math grades for freshmen during the two years of the study.

	1978			1979		
Math Class	N	Mean	Sx	N	Mean	Sx
081/103	214	2.11	11.25	263	2.20	11.19
082/104	790	2.14	11.02	947	1.94	11.53
108	1656	2.29	12.31	1903	2.10	12.65
111	1333	2.57	9.17	1165	2.35	11.83
112	1259	2.55	10.52	1026	2.28	12.17

Table 38

Mean Math grades for each term 1978-79 and 1979-80 for all students at all class levels.

Math Class	F78	W79	S79	F79	W80	S80
081/103	2.07	1.92	1.65	2.07	2.20	1.68
082/104	2.02	1.66	1.64	1.94	1.79	1.83
108	2.14	2.40	1.93	2.08	2.06	1.34
109	1.97	2.26	2.24	1.94	2.32	1.92
111	2.28	2.18	2.21	2.34	2.11	2.07
112	2.17	2.06	2.04	2.19	2.05	2.00

Note. F=fall, W=winter, S=spring

A considerable proportion of students who enrolled in Math classes received below a 2.00. In some cases over 50% of all of the students taking any one Math class during a single term received a grade below 2.00. The proportion of all students who were enrolled in a Math class during the 1978 or 1979 academic years who received a final grade below a 2.00 are presented in Table 39.

Table 39

The percentage of all enrollees receiving below a 2.00 in each Math class during each term of the 1978 and 1979 academic years.

Math Class	F78	W79	S79	F79	W80	S80
081/103	34.1%	42.9%	49.5%	33.6%	15.7%	50.0%
082/104	40.5%	51.2%	53.2%	40.5%	47.0%	44.5%
108	34.3%	25.8%	41.6%	38.4%	37.4%	61.0%
109	39.9%	31.9%	31.1%	15.2%	30.5%	40.9%
111	29.6%	34.6%	34.1%	26.2%	33.2%	37.1%
112	33.6%	37.5%	33.7%	33.7%	32.8%	40.7%

Note. F=fall, S=spring, W=winter

Summary

The hypotheses considered in this study are summarized in Table 40. Hypotheses were tested using the 1978 entering group and replication was performed using the 1979 entering group. In general, the outcomes that would ideally be expected under the psychometric decision model did not appear. Significant aptitude treatment interactions were seldom found and when they were discovered they did not remain constant over treatment years. Students who enrolled in the course recommended for their placement test score tended to perform better than students who enrolled in a class which was at a higher level than recommended for their placement test score. The alternate treatments which ended in a common terminal class tended not to equalize initial differences. Students taking the short sequences did better in the terminal class than did students taking

Table 40

Summary of the hypotheses under study and whether they were accepted or rejected.

Hypothesis	Accept/Reject
Hypothesis 1: The regression lines for terminal course grades on placement test scores for the long and short calculus sequences will intersect in the range of the current cutting score (16.5-17.5).	reject
Hypothesis 2: The regression lines for terminal course grades on placement test scores for the long and short algebra sequences will intersect in the range of the current cutting score (13.5-14.5).	reject
Hypothesis 3: The regression lines for terminal course grades on the American College Test Math Subscore for the long and short calculus sequences will intersect at a point which is appropriate for sectioning students into treatments.	reject
Hypothesis 4: The mean class grade for students enrolling in a Math class at a level recommended for their placement test score will not be significantly different from the mean grade of students enrolling in a higher level class than is recommended for their placement test score.	
Math 082/104	accept
Math 108	reject
Math 111	reject
Math 112	reject
Hypothesis 5: The mean Math 112 grade for students taking the long calculus sequence will not be significantly different from the mean Math 112 grade for students taking the short calculus sequence.	reject

the long sequences except in calculus during the 1979 year where outcomes were equal.

Aptitude treatment interactions did not occur within the current Math placement test score cutting score range for either the long or the short calculus sequences. Interactions outside of the current cutting score range, which would indicate appropriate cutting scores, did not appear. The long and short calculus sequences produced different mean grade outcomes even after adjustments for initial Math placement differences were made. The treatment which produced the superior result reversed between the 1978 and the 1979 treatment years. In 1978 the short calculus sequence produced the highest mean outcome while in 1979 the long calculus sequence resulted in the highest mean outcome. When sequence was held constant and the year that the sequence was administered was varied a treatment by year effect was found. Students in short calculus in 1978 outperformed students in short calculus in 1979 after adjustments were made of initial placement test score differences. Students completing long calculus performed at the same level both years.

No aptitude treatment interactions occurred within the current cutting score range when terminal course grades were regressed on placement test scores for either the long or short algebra sequences. The regression lines within treatment were not significantly different from 0 indicating that there was no evidence of predictive validity within the sequence.

Students completing the short algebra sequence outperformed students completing the long algebra sequence during both the 1978 and the 1979 treatment years. There was no reversal in the algebra treatment which produced the superior effect across years as there was with the calculus treatments.

Students in the short algebra sequence in 1978 outperformed students in the short algebra sequence in 1979 after adjustments for initial differences in placement test scores were made. There were no significant differences in treatment outcomes across years for individuals who completed the long algebra sequence.

No significant aptitude treatment interactions, which would be useful for setting cutting scores, were found when terminal course grades were regressed on American College Test Math scores for the long and short calculus sequences. Individuals completing the short calculus sequence outperformed students completing the long calculus sequence in 1978 when the ACT Math score was used as a covariate. Students taking the short calculus sequence in 1978 outperformed students taking the short calculus sequence in 1979 when the American College Test Math score was used as the covariate. No significant outcome differences were found for the long calculus treatment over years.

An aptitude treatment interaction was found for the long and short algebra sequences during the 1979 treatment year; however, no similar interaction appeared during the first year of the study. Students completing the short algebra sequence outperformed students completing the long algebra sequence after adjustments were made for initial ACT Math score differences during both years of the study. Students completing the short algebra sequence in 1978 outperformed students completing the short algebra sequence in 1979 after adjustments for initial ACTms differences were made. There were no differences in mean outcome between the two years for students completing the long algebra sequence after ACTms adjustments were made.

The mean grades of students who enrolled in a Math class that was recommended for their placement test scores were compared with the mean grades of students who enrolled in the same class but at a higher level than was recommended for their placement test score. There was no significant difference in mean outcomes for those students completing Math 082/104. Students enrolling correctly in Math 108, 111, or 112 outperformed students who had placement test scores below the recommended cutting points.

The long and short sequences for both the algebra and the calculus treatments end in common terminal classes.

Students completing either a long or a short sequence should be prepared to begin the next successive level of Math. In 1978 students completing either the short algebra sequence or the short calculus sequence did significantly better in the terminal class than students completing the corresponding long treatment. In 1979 there were no significant differences in outcome for students completing either calculus sequence but students in the short algebra sequence outperformed students in the long algebra sequence.

Regression analysis was performed on only students who followed placement policy in order to determine if there were discontinuities of slope between treatment groups which would indicate proper cutting points. No discontinuities were discovered, however, mean outcome differences did appear when adjustments for initial Math placement test differences were made. Students did better in short algebra than in long algebra during both study years. In 1978 the long and short calculus sequences produced similar results while in 1979 the long calculus treatment proved superior after adjustments were made.

The usefulness of both the MSU placement test and the American College Test Math score to predict Math grades was limited. Predictive validity coefficients ranged from -.06 to +.21 for the MSU placement test and from -.29 to +.20 for the American College Test Math score.

The majority of students completing Math sequences during their first academic year do follow placement policy. There is evidence that placement policy was more rigidly enforced during the second year of the study. Many students who begin a 2 or 3 term Math sequence during their freshman year do not complete the sequence within three terms.

A high proportion of students receive below a 2.00 in the Math classes that they complete. The percentage of students receiving below a 2.00 for any course during any term of the study ranged from 15.7% to 61.0%.

Chapter 5: Summary and Recommendations

The purpose of this study was to examine the fit between the MSU Math placement test and psychometric decision theory. The American College Test Math subscore was also examined within the decision framework to determine its usefulness for Math placement. Students following placement policy and students not following placement policy were examined in order to determine their relative performance. Alternate treatments which ended in a common terminal class were compared to determine if they actually resulted in similar learning outcomes.

Within the psychometric decision framework aptitude treatment interactions should exist if there is justifiable reason for assigning students with different characteristics to alternate treatments intended to produce similar learning outcomes.

A long algebra sequence was compared with a short algebra sequence and a long calculus sequence was compared with a short calculus sequence. Each sequence was compared across years in order to determine if the treatments were consistent across years.

Students who participated in the study included those who: entered the university in fall 1978 and 1979 and reported ACT scores, and students who entered the university in fall 1981 and took a Math class during their first term at Michigan State University. Analysis of covariance and the Johnson-Neyman technique were utilized to determine if outcome differences occurred and if aptitude treatment interactions existed. The 1978 entering group served as the primary group and cross validation was performed on the 1979 entering group. T tests were used to determine if students enrolling above placement recommendations performed differently than students enrolling within placement guidelines and, if alternate sequences with common terminal courses resulted in similar performance outcomes.

No aptitude treatment interactions were found for the MSU Math placement test and the two calculus sequences when grades in a common terminal class were used as the criterion measure. When outcomes were adjusted for initial MSU Math test score differences significant outcome differences were found. The direction of the differences was not consistent across years. In 1978 students completing the short calculus sequence outperformed students completing the long calculus sequence while in 1979 students completing the long calculus sequence outperformed those completing the short calculus sequence after adjustments for initial score differences.

Each sequence was compared across years to determine if they were consistent. It was found that the short calculus treatment produced significantly different results in 1978 than it did in 1979 after adjustments for initial MSU placement differences were made. Students taking the short calculus sequence in 1978 outperformed students taking short calculus in 1979. The long calculus treatment remained consistent across study years.

No aptitude treatment interactions were found for the two algebra sequences and the MSU Math placement test when grades in a common terminal class were used as the criterion measure. There was no reversal in the algebra treatment which produced the best results over years. Students in both the 1978 group and the 1979 group who completed the short algebra sequence outperformed students who completed the long algebra sequence after adjustments for initial MSU Math placement score differences were made.

Each algebra sequence was examined to determine if it was consistent across years. The short algebra sequence was not consistent across years while the long algebra sequence was consistent across years. Students taking the short algebra sequence in 1978 outperformed students taking the short algebra sequence in 1979 after adjustments for initial MSU Math placement test scores were made.

The American College Test Math subscore was examined for usefulness in placement in relation to psychometric decision theory. Both the calculus sequences and the algebra sequences were considered.

No aptitude treatment interactions were found for the calculus sequences when the American College Test Math subscore was regressed on the common terminal course grade for the long and short sequences. In 1978 students taking the short calculus sequence outperformed students taking the long calculus sequence after adjustments for initial American College Test Math subscores were made. Outcome differences did not appear for the 1979 groups.

The calculus treatments were compared across years for consistency using the American College Test Math subscore as the covariate. Students taking the short calculus sequence in 1978 outperformed students taking the short calculus sequence in 1979 after adjustments for initial differences were made. Students completing the long calculus sequence performed at the same level both years.

No aptitude treatment interactions were found for the algebra sequences when common terminal course grades were regressed on American College Test Math subscores for the 1978 group. However, an aptitude treatment interaction between the American College Test Math subscore and the algebra treatments did appear in the 1979 group. The regression lines intersected at a point corresponding to an American College Test Math subscore of 11.13. The treatments produced significantly different results for individuals with Math subscores of 14 or above.

Students taking the short algebra sequence outperformed students taking the long algebra sequence during both 1978 and 1979 after adjustments for initial differences in American College Test Math subscores were made.

The algebra sequences were compared across years using the American College Test Math subscore as a covariate. The long algebra treatments were consistent across years but the short algebra treatments were not consistent across years. Students in the short algebra treatment in 1978 outperformed students in the short algebra treatment in 1979 after adjustments for initial American College Test Math scores were made while students in the long algebra treatment performed at a similar level both years. The patterns discovered when the American College Test Math subscore and the MSU Math placement test were used as covariates are highly similar.

The entering class of 1980 was examined to determine if students who enrolled in a Math class at the level suggested for their MSU Math placement test score performed the same as students who enrolled in a class at a higher level than was recommended for their Math placement test score. The mean grade in each class was compared for both groups.

Students who enrolled correctly in Math 082/104 (Intermediate Algebra) performed at the same level as students who enrolled incorrectly in Math 082/104.

For Math 108 (College Algebra and Trigonometry), Math 111 (College Algebra with Trigonometry), and Math 112 (Calculus and Analytic Geometry) students who enrolled consistent with placement policy performed significantly better than students who enrolled in a class higher than recommended by placement policy for their Math placement test score.

The long calculus sequence and the short calculus sequence end in a common terminal class and the short algebra sequence and the long algebra sequence end in a common terminal class. More able students are intended to take the short sequences while less able students are intended to take the long sequences. The corresponding sequences are designed for students with different levels of ability but end with a common class. Both calculus treatments and both algebra treatments were compared to determine if they resulted in similar performance outcomes.

Students completing the short calculus sequence outperformed students completing the long calculus sequence in 1978. In 1979 students completing either calculus sequence performed equally. Students completing the short algebra sequence outperformed students completing the long algebra sequence during both treatment years. In general, the alternate treatments did not appear to result in similar outcomes.

The regression lines for class grades on MSU Math placement test scores were examined for linearity. Only two instances of curvilinearity were found. These may have been due to ceiling effects, however, examination of the criterion means plotted on placement test scores suggest that variation in the extremes of the distributions produced the curves. Students in classes too difficult for their skill level may perform erratically.

An analysis seeking identification of aptitude treatment interactions was performed on only those students who enrolled consistent with placement policy in order to determine if significant discontinuities in regression lines might appear when placement policy was rigidly followed. No significant discontinuities between treatments were found.

The predictive validity coefficients for the MSU Math placement test and individual Math classes range from .15 to .38. These correlations are moderate to low. The placement test was only good for predicting between 2% and 14% of the variability in class grades.

The predictive validity for both the MSU Math placement test and the American College Test Math subscore were examined for Math sequences. For the MSU Math placement test predictive validity for sequences ranged from -.06 to +.23 while for the American College Test Math subscore predictive validity for sequences ranged from -.29

to +.20. In general, the MSU test had better predictive validity for sequences than did the ACT score. The ACT was superior for the short algebra sequence though. The relationship is negative between Math grades and either the MSU test or the ACT score for the long algebra sequence. Neither test has very high predictive validity for sequences.

The first grade in a Math sequence was correlated with the last grade in the Math sequence for each treatment. Correlations ranged between .31 and .59 for the treatments in 1978 and 1979 independently. The first grade in a sequence then only accounted for between 10% and 35% of the variation in the last grade in the sequence.

Up to 40% of the students in the study enrolled in a sequence above or below that recommended for their Math placement test score. Placement guidelines were more rigidly enforced during the 1979 year than they were during the 1978 year. In 1978 13.0% of the students enrolled in the short calculus sequence above recommendation while in 1979 only 1.5% did. Students completing the short calculus sequence in 1978 outperformed students completing the short calculus sequence in 1979.

Many students who begin a Math sequence their first term do not complete the sequence during their first three terms. A high proportion of all students including upper classmen received below a 2.00 in Math classes. The percentage of students receiving below a 2.00 for any class during any term 1978 or 1979 ranged from 15.7% to 61.0%. The mean grades in specific classes for all students were fairly consistent over the years of the study. The grades for freshmen completing sequence varied significantly across years. Examination of the mean grades for all students including individuals repeating classes may mask changes in treatments or grading scales which actually occur.

<u>Conclusions</u>

1. The Michigan State University Math placement test does not fit the psychometric decision model. No aptitude treatment interactions appear and no discontinuities in regression lines appear which would indicate aptitude treatment interactions. This does not preclude that aptitude treatment interactions would occur if random assignment to treatments was undertaken. The low predictive validities of the test suggest even if aptitude treatment interactions were found there would be a rather wide range of scores for which individuals could be expected to perform at the same level in alternate treatments.

2. No consistent aptitude treatment interactions were found for the American College Test Math subscore and the treatments. The one interaction which did appear was not replicated across treatment years. Predictive validity

estimates for the Michigan State University Math placement test tended to be higher than corresponding predictive validity estimates for the American College Test Math subscore. However, predictive validity for both instruments are low. The American College Test may be as useful for placement as the Michigan State University Math placement test if testing costs to the institution are considered.

3. In general, students enrolling as recommended by the MSU Math placement policy perform better than students enrolling in a class higher than is recommended for their test score by placement policy.

4. The long Math sequences do not compensate for differences in initial ability levels. In general, students taking a long sequence do not reach the level of students taking a short sequence.

5. The short algebra and short calculus treatments did not remain consistent across years. Grades for the two treatments were lower during 1979 than they were during 1978. Also, placement became more restrictive during the 1979 treatment year. Changes in grading distributions constitute treatment changes.

6. The comparison of grade distributions across years for all students in a Math class may not reflect treatment changes. Treatment differences for students completing a sequence of classes within a given time frame

may be masked when the grades for other students are included in the distribution. Grades received by upper classmen and students repeating classes may make total distributions remain stable across years while within treatment grades do not remain stable.

7. Treatments may not be coherent units. The first grade in a sequence may not predict very much of the variation in the final grade in the sequence. If the first grade in a sequence is not very good for predicting the final grade in the sequence it may be unrealistic to expect even moderate predictive validity for a placement test.

Discussion

The Mathematics educational process is complex. Many students have difficulty with the material. The primary purpose of placement is to improve student outcomes by assigning students to treatments in which they are likely to be successful while avoiding the presentation of material which the student already understands. There are secondary purposes for placement such as filling quotas in order to make institutional functioning smoother. Placement should be designed and executed in order to maximize student outcomes while institutional resources should be used in the fashion which best serves this goal.

A relatively high proportion of students who complete Math classes at MSU receive below a 2.00 in those classes. When too high of a proportion of students fail classes there are detrimental results. The students may be forced to change their majors or may even choose to leave the institution.

If the proportion of students who receive below a 2.00 is too high then the placement of students into the treatments is inadequate. If the problem is that students are not properly prepared, then the treatment is inappropriate for them. A placement procedure is not independent of the instructional method, the content of the treatment, or the motivation of students as a group. If too many students fail, then either placement procedures, instructional conditions, or both should be modified.

Within the framework of psychometric decision theory, aptitude treatment interactions are necessary in order to justify the placement of students into alternate treatments. An important assumption of the model is that outcomes are measured on a utility scale where all relevant variables are considered. In this study grades were used as the criterion scale. There are several problems associated with using grades as criterion measures: they may not be reliable, they may not accurately reflect the objectives of the course, they may vary from instructor to instructor, and they do not include other utility costs. Although grades do not include costs to the institution and costs to the student they are primary indicators of student success. Grades are treated as very important indicators in education and should have a major role in educational decision making, however, they should be interpreted with caution.

In this study no aptitude treatment interactions occurred when class grades were used as the criterion and placement test scores were used as the predictor variable. In general, the slopes of all of the regression lines tended to be low. A variable or variables other than ability as measured by the Math placement test is contributing to the variation which is observed in the grade distribution. In this case the Math placement test may be used as a guide for placement but other factors should be considered. It is important that the actual factors which affect grade variation be identified in order to better match students to instructional treatments. If the class content or the instructional method is such that it has only a minor relationship to measured ability then many adequately prepared students, who are adaquately prepared as measured by a placement test, will continue to fail. If the problem lies with student motivation then methods of motivating students must be found.

The first grade in each Math sequence considered was only moderately correlated with the final grade in the sequence. If the grades only have moderate predictability then it is not very realistic to expect a placement instrument to be highly correlated with final grades.

Initial grades within an instructional unit may be expected to be more highly related to final class grades than will be a preliminary measure of general achievement.

No consistent aptitude interactions appeared when the American College Test Math subscore was used as the predictor of final class grades. The only aptitude treatment interaction which occurred was not replicated across years of the study. Correlations between the American College Test Math subscore and final class grades tended to be even lower than correlations between the MSU Math placement test and final class grades. If only grades are considered then the MSU Math placement test is better than the American College Test Math subscore in this instance. However, if testing costs are considered then the American College Test Math subscore may be a viable placement instrument.

Although the MSU placement test is only moderately correlated with final class grades there is some indication that it is useful for placement. As a group, students who enrolled consistent with placement recommendations outperformed students who enrolled in classes above the level recommended for their placement test score. However, the low validity of the placement test suggests that there is little justification for enforcing it rigidly.

In this study, students were assigned to alternate treatments which ended in common terminal classes.

Students of lesser ability were given a longer treatment while students of higher ability were given a shorter treatment. If the alternate treatments compensate for initial differences in ability then students completing both treatments should perform at a similar level in the common terminal class.

More able students who completed a short treatment tended to perform better in the common terminal class than less able students who completed the corresponding longer treatment. The treatments were not successful in compensating for initial differences in ability as measured by the MSU placement test. If the treatments do not compensate for low initial ability then such students will continue at a disadvantage as they progress into higher levels of Math regardless of the instructional treatment to which they are assigned. Carried to its extreme this might indicate that differences in Math ability for entering students will limit Math learning in college. If the institution wishes to provide students with the opportunity to make gains beyond a level projected by their initial ability then treatments must be modified in order that low and marginally able students do not continue to be at a disadvantage.

Grades were used as the criterion measure in this study. The situation noted in the above paragraph may be even more extreme if the length of treatment were incorporated into the utility scale. If more able students do better in two terms than do less able students in three terms, then more able students will be at an even greater advantage when length of instruction is included in the outcome measure. Less able students not only take more time to reach the same terminal class but they tend to do less well in the terminal class than do more able students.

In the current situation where the MSU test is being used for placement many students are doing poorly. When the cost of testing is considered it is likely that the ACT would be as effective for placememt as is the MSU test. Neither the MSU test nor the ACT are adequate placement instruments in the existing situation. A new set of placement procedures which are more closely matched to instruction and outcomes are needed. The current procedures are not diagnosing student deficiencies with regard to the existing course seqeunces.

Recommendations

1. MSU Mathematics placement procedures should be studied in more detail. The factors leading to the relatively low correlations between the placement test and final class grades need to be identified. New placement procedures should be developed which are more highly related to success within the instructional sequences. Measures other than ability or aptitude should be considered. It may be necessary to modify both placement

procedures and instructional treatments in order to increase the rate of successful placement. Placement rules which result in increased success rates should be developed.

2. The consistency of grades within a sequence of Math classes needs to be studied. The sources of variation in grades other than subject matter ability needs to be identified.

3. The problems which students have in Math classes need to be considered. Successful and unsuccessful students in particular Math classes should be compared both in terms of their measured ability and in terms of their subjective reactions to the material and the instruction.

4. The impact of failure rates on the persistence of students should be investigated. Failure in Math is likely to lead students to change majors and may also lead some students to leave the institution.

5. The impact of letting students take a more active role in their placement decisions should be examined. Students may be able to make good placement decisions when they are provided with relevant information.

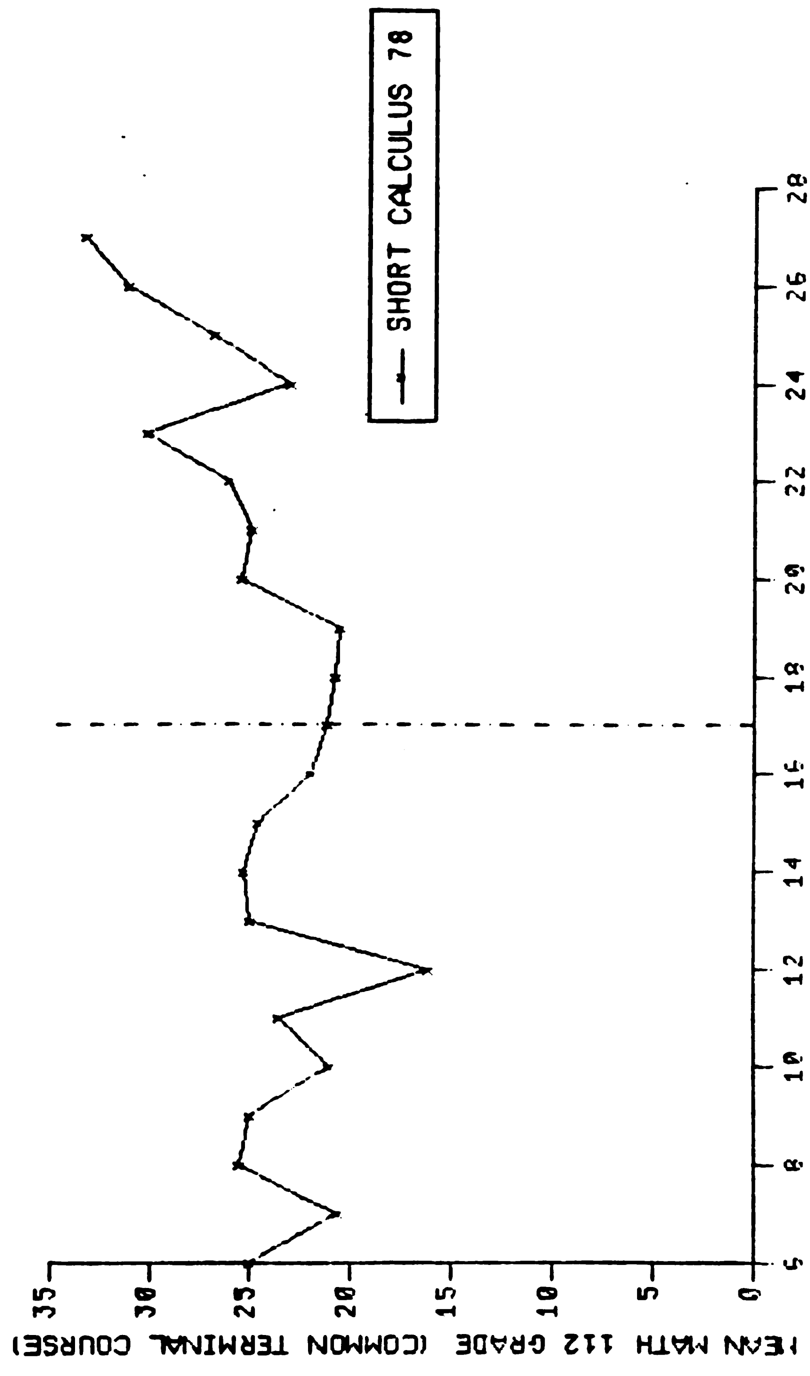
APPENDIX A

Appendix A

Figure 6

MEAN CRITERION SCORES FOR EACH PREDICTOR SCORE

SHORT CALCULUS 1978



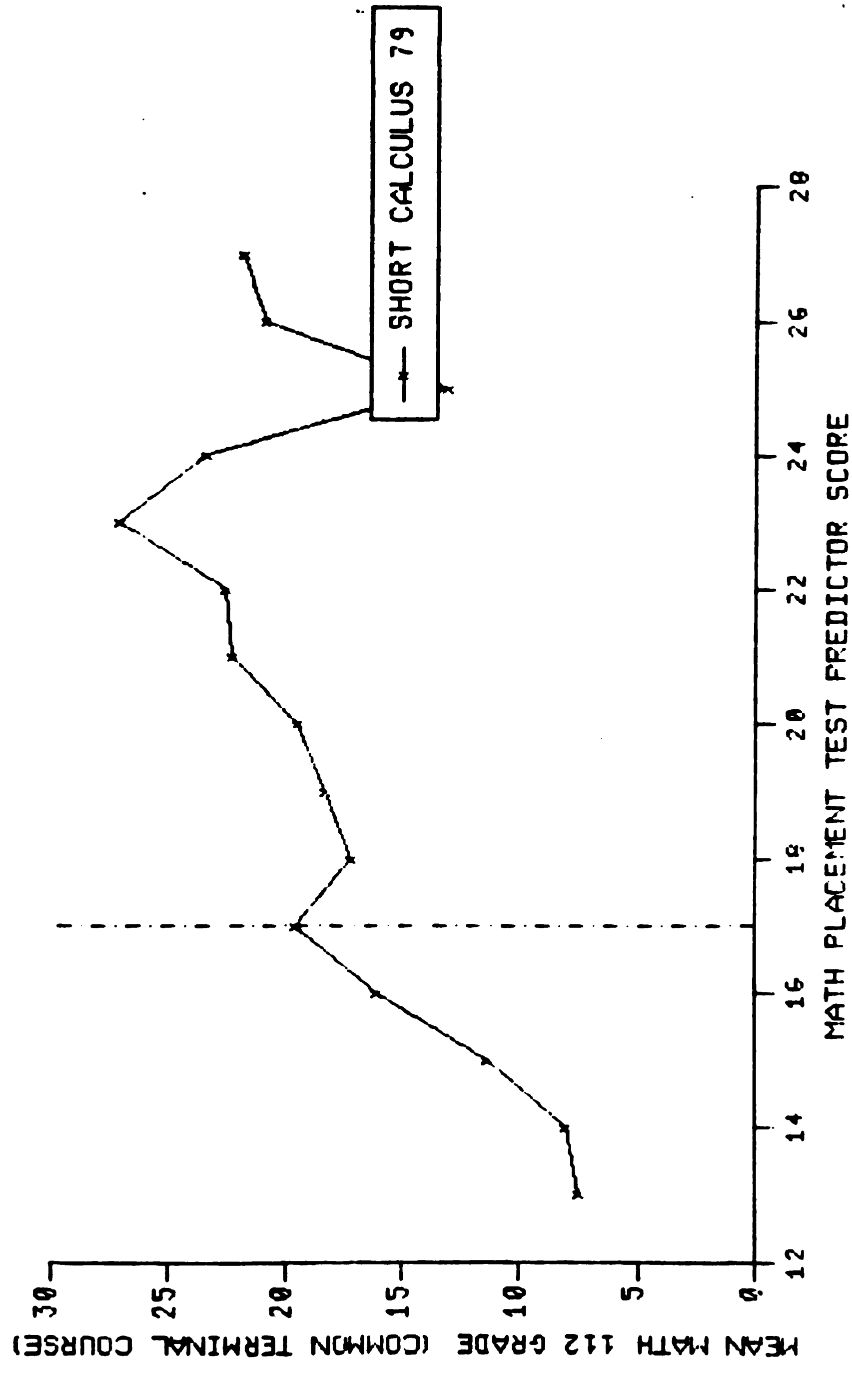
The dashed line is the current recommended cutting score

Appendix A

Figure 7

MEAN CRITERION SCORES FOR EACH PREDICTOR SCORE

SHORT CALCULUS 1979



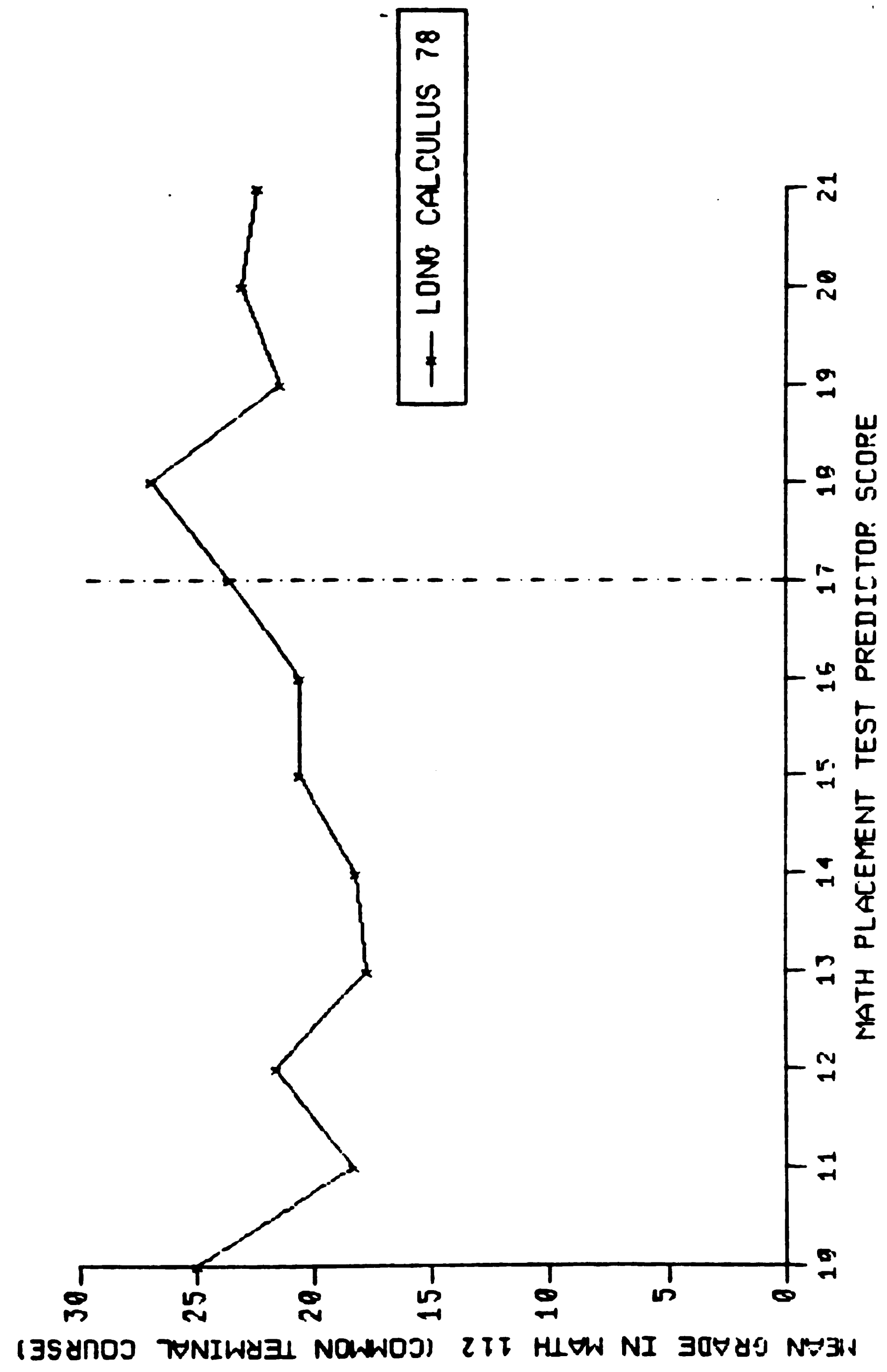
The dashed line is the current recommended cutting score

Appendix A

Figure 8

MEAN CRITERION SCORES FOR EACH PREDICTOR SCORE

LONG CALCULUS 1978



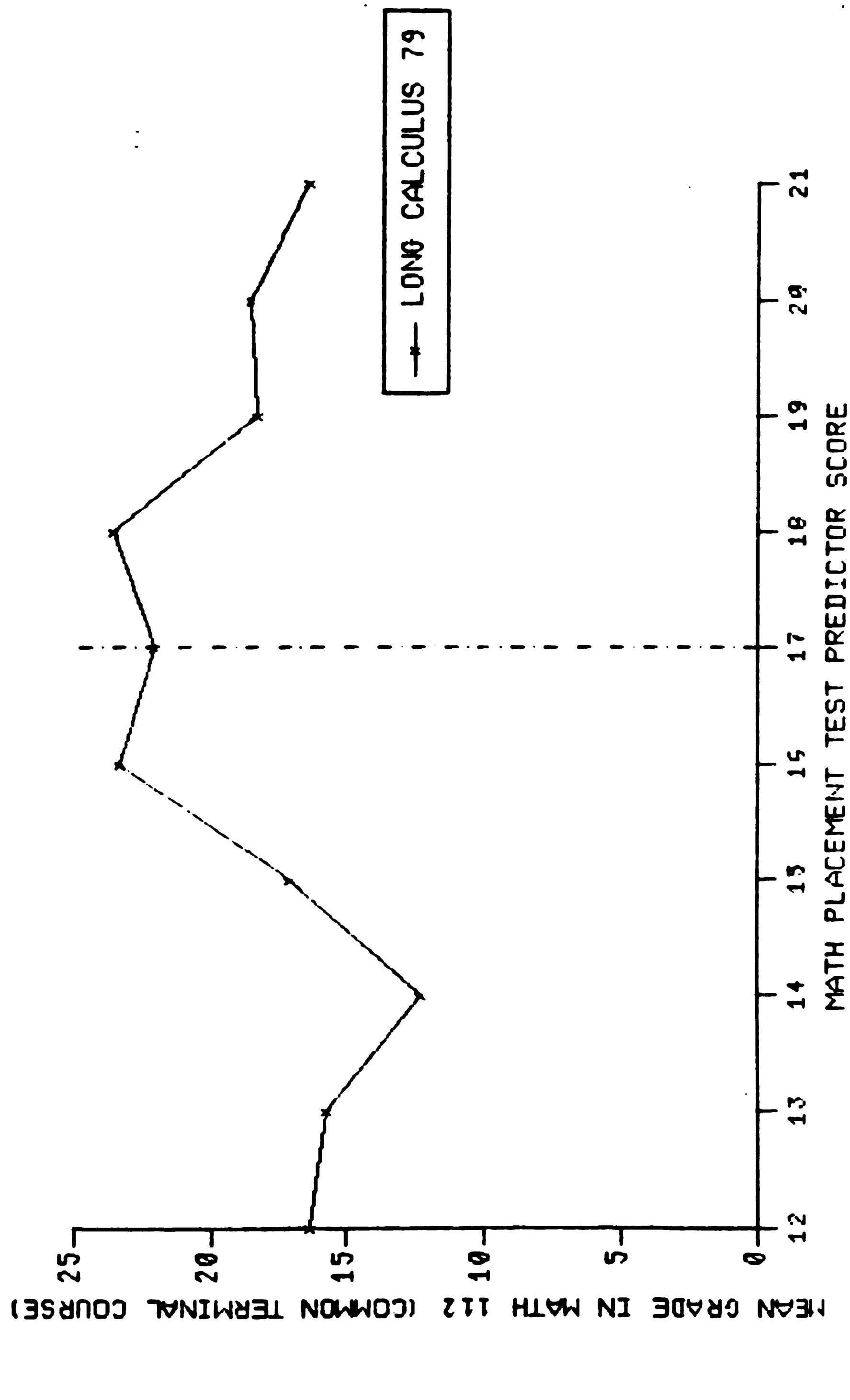
Dashed line corresponds to the suggested cutting score

Appendix A

Figure 9

MEAN CRITERION SCORES FOR EACH PREDICTOR SCORE

LONG CALCULUS 1979

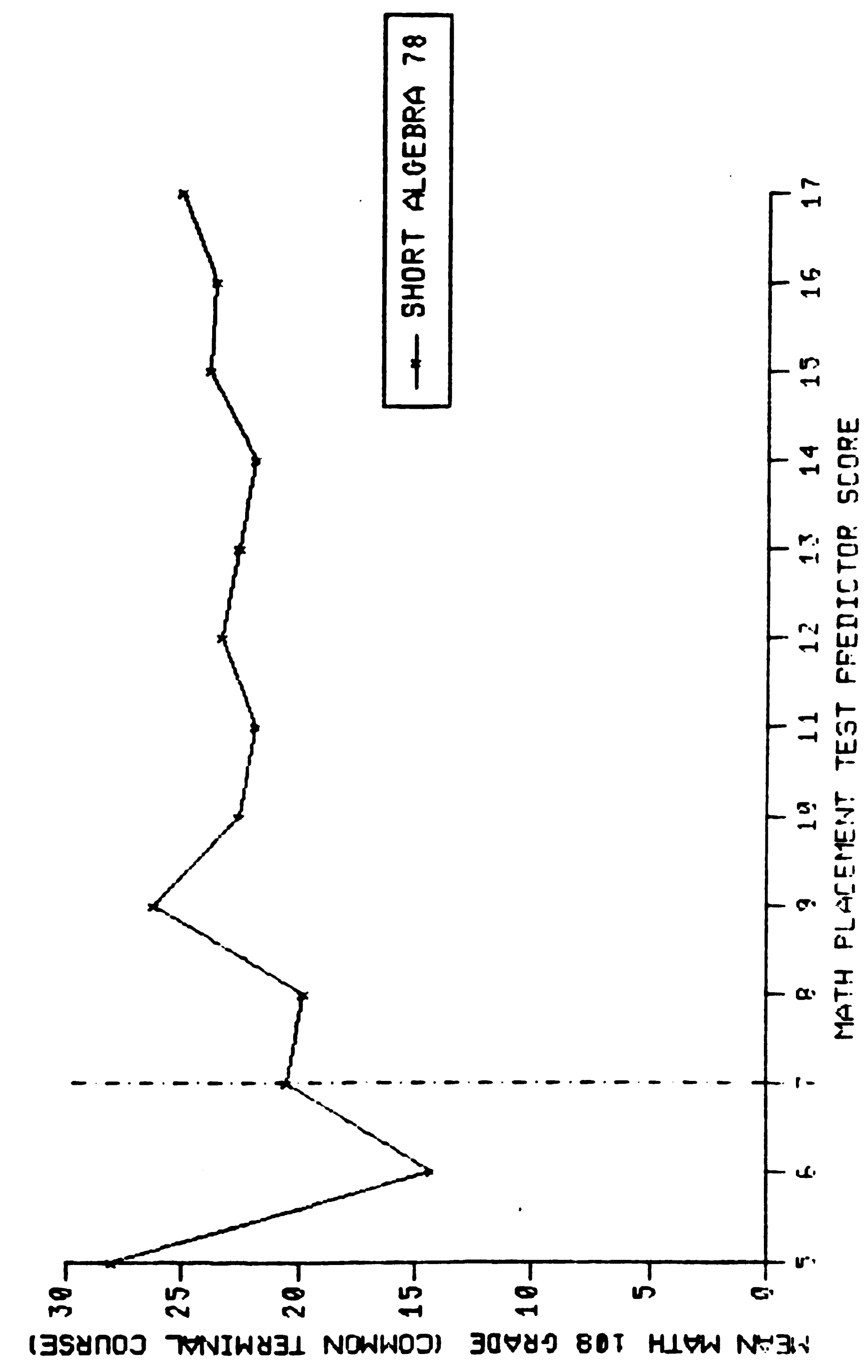


Dashed line corresponds to the suggested cutting score

Appendix A

Figure 10

MEAN CRITERION SCORES FOR EACH PREDICTOR SCORE
SHORT ALGEBRA 1978



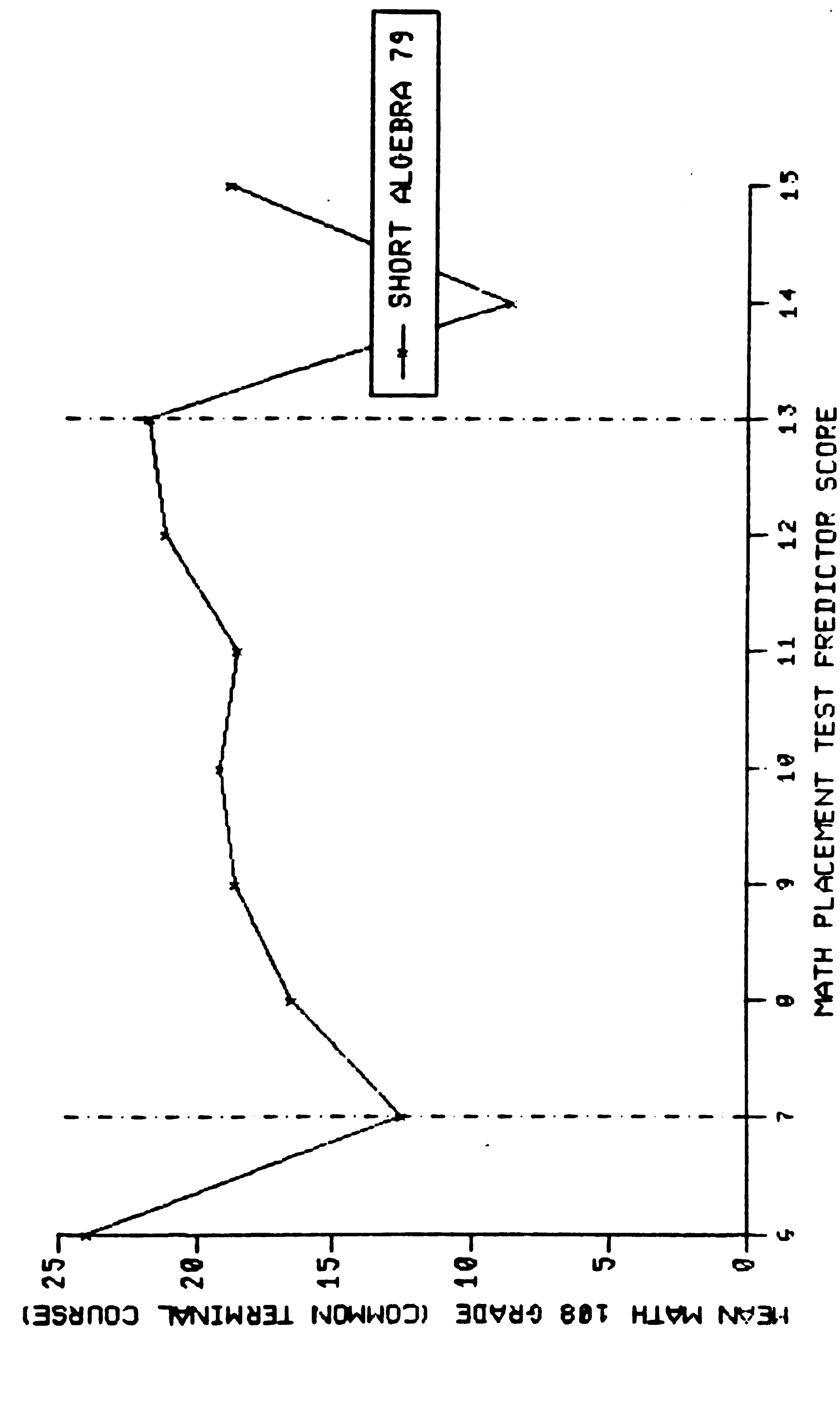
Dashed line represents current cutting score

Appendix A

Figure 11

MEAN CRITERION SCORES FOR EACH PREDICTOR SCORE

SHORT ALGEBRA 1979



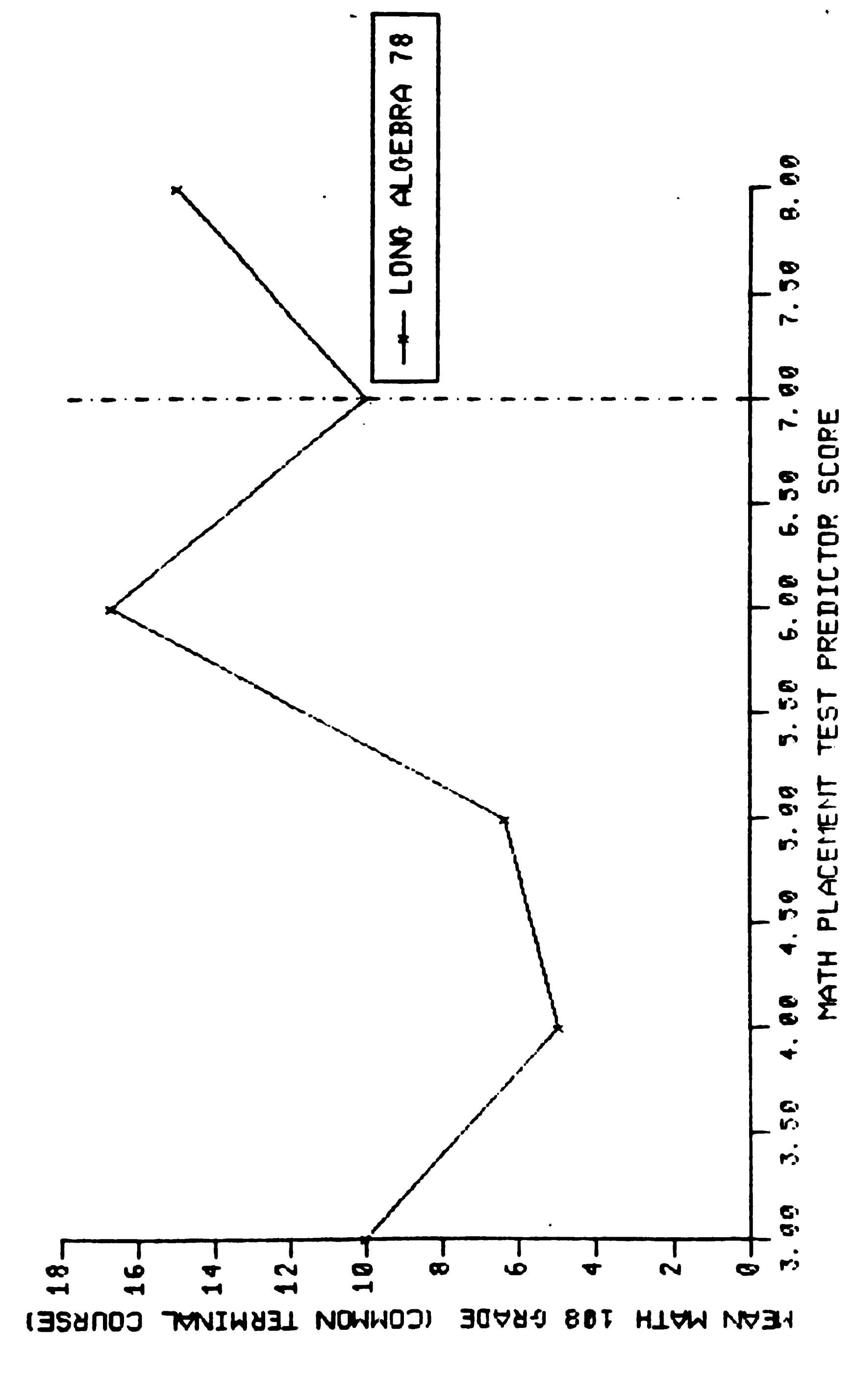
Dashed lines are current cutting scores

Appendix A

Figure 12

MEAN CRITERION SCORES FOR EACH PREDICTOR SCORE

LONG ALGEBRA 1978



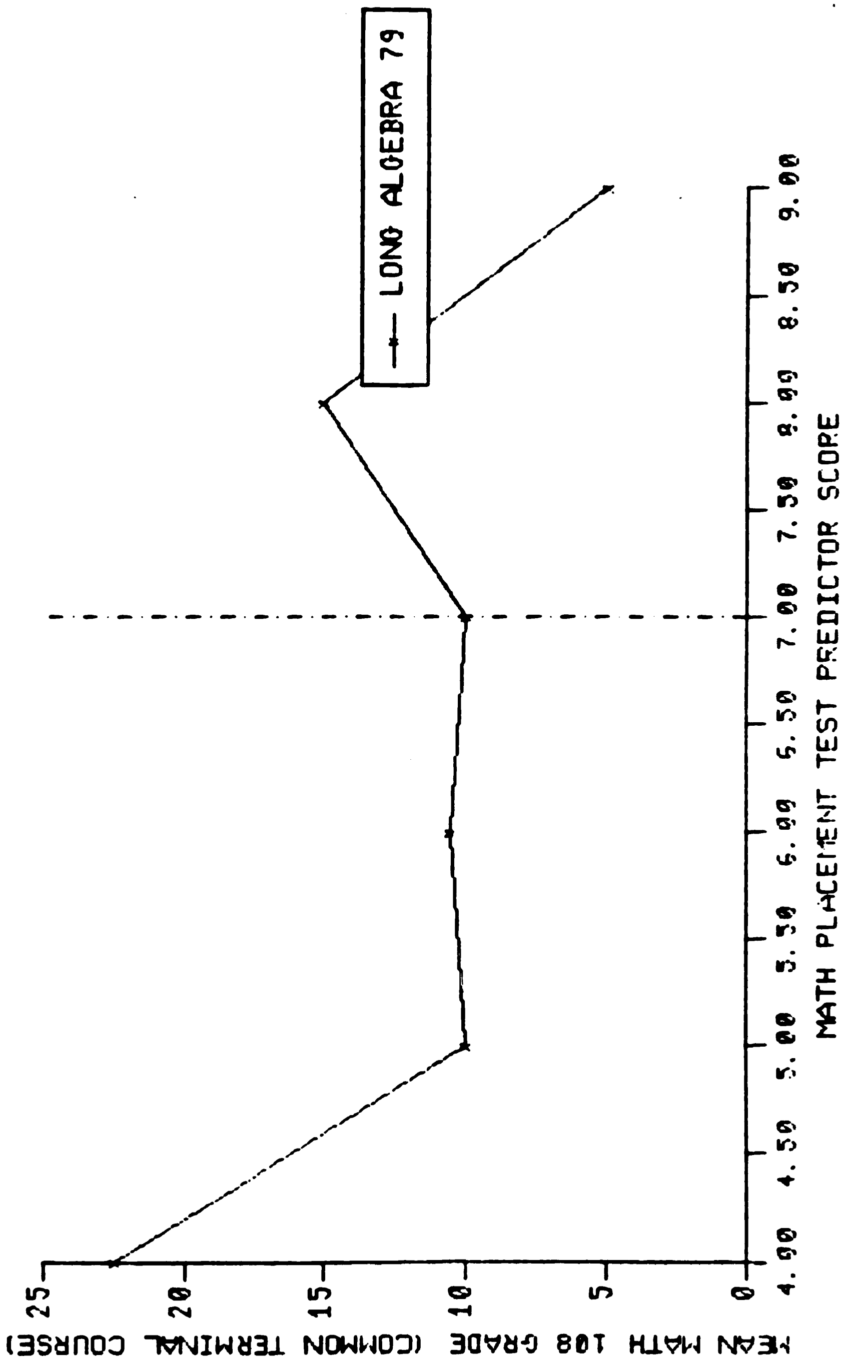
The dashed line is the current recommended cutting score

Appendix A

Figure 13

MEAN CRITERION SCORES FOR EACH PREDICTOR SCORE

LONG ALGEBRA 1979

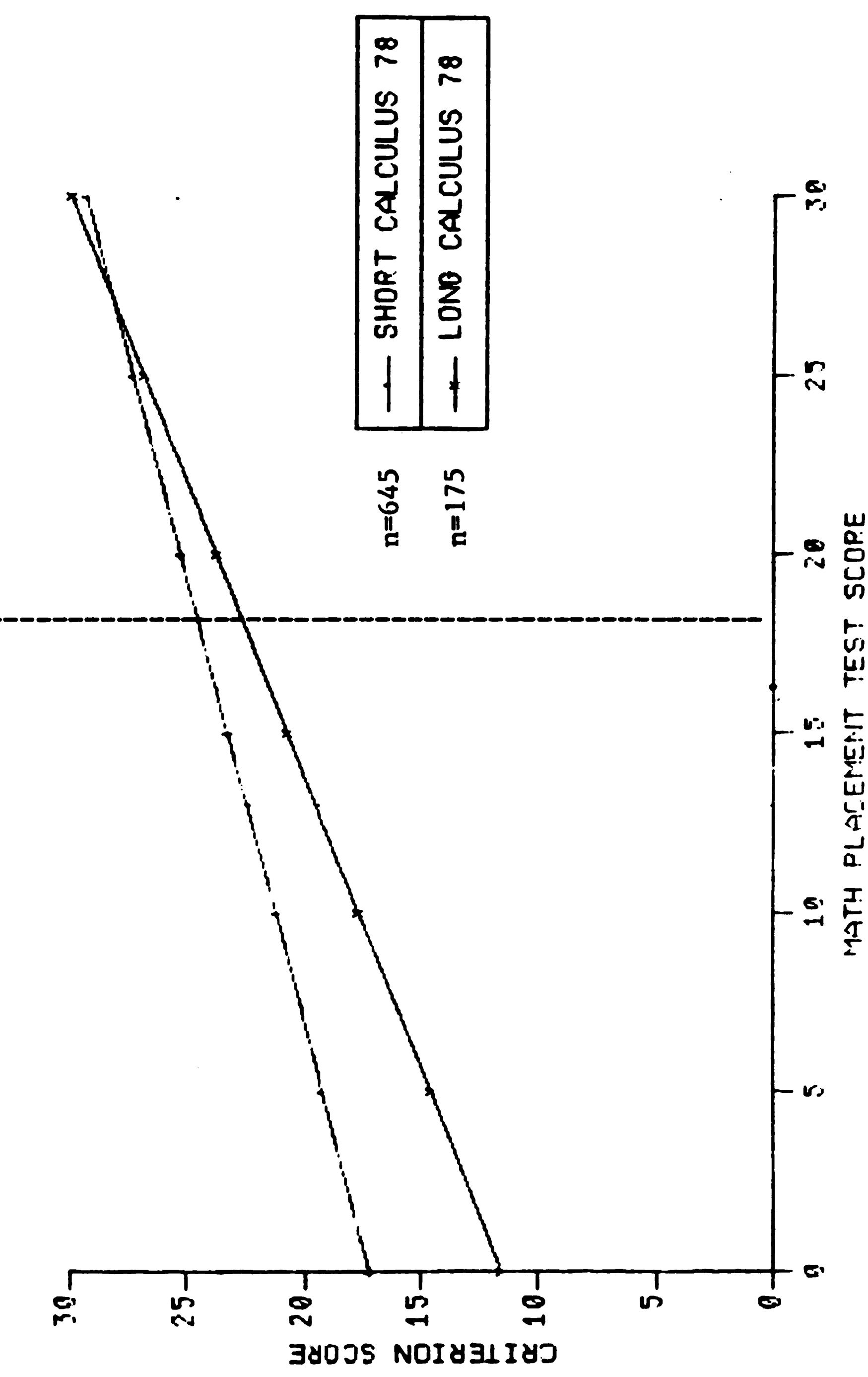


The dashed line is the current recommended cutting score

Appendix A

Figure 14

REGRESSION OF MATH 112 GRADES ON MATH PLACEMENT TEST SCORES
1978

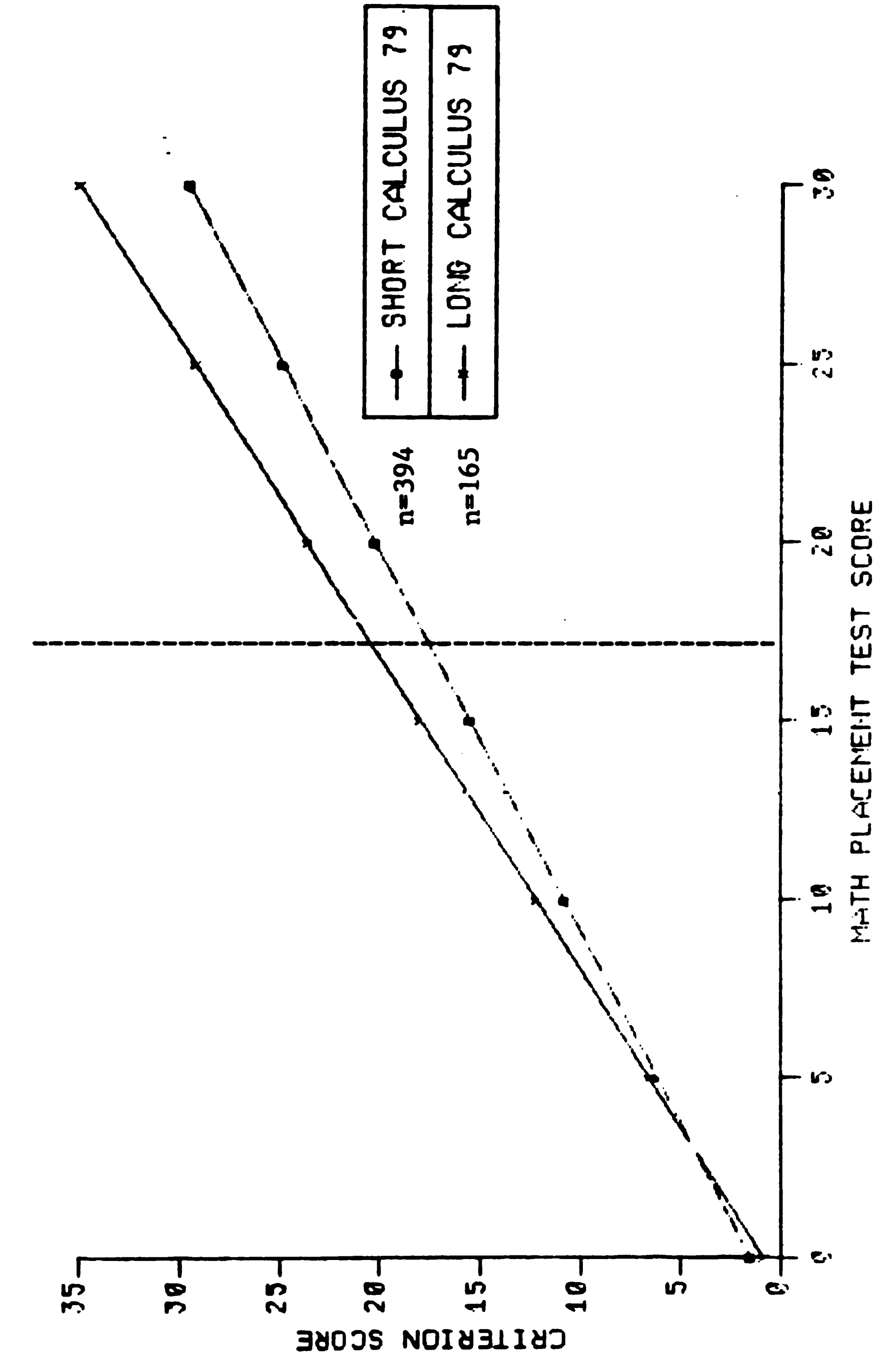


The dashed line represents the current cutting score

Apendix A

Figure 15

REGRESSION OF MATH 112 GRADES ON MATH PLACEMENT TEST SCORES
1979

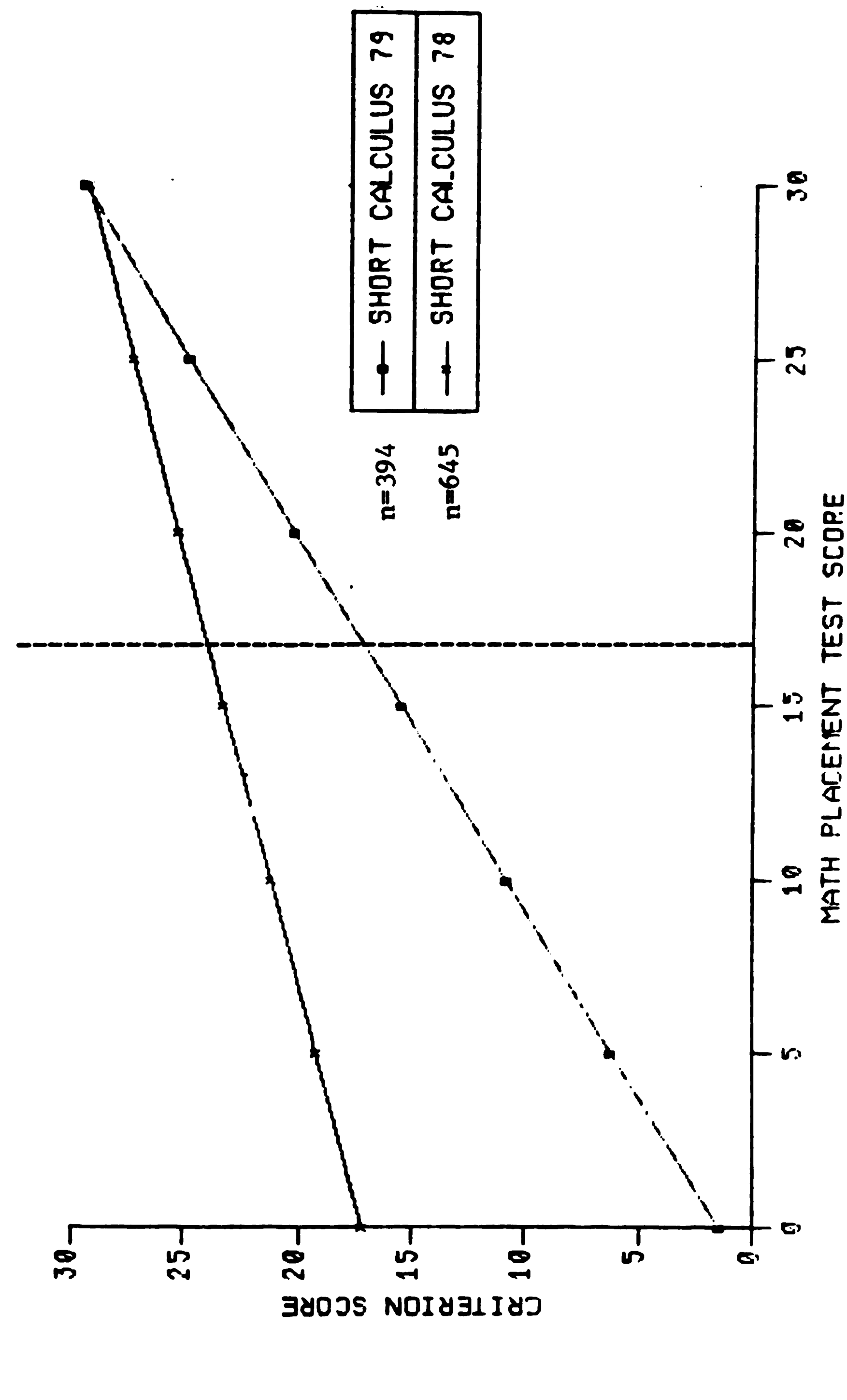


The dashed line represents the current cutting score

Appendix A

Figure 16

REGRESSION OF MATH 112 GRADES ON MATH PLACEMENT TEST SCORES
1978 AND 1979

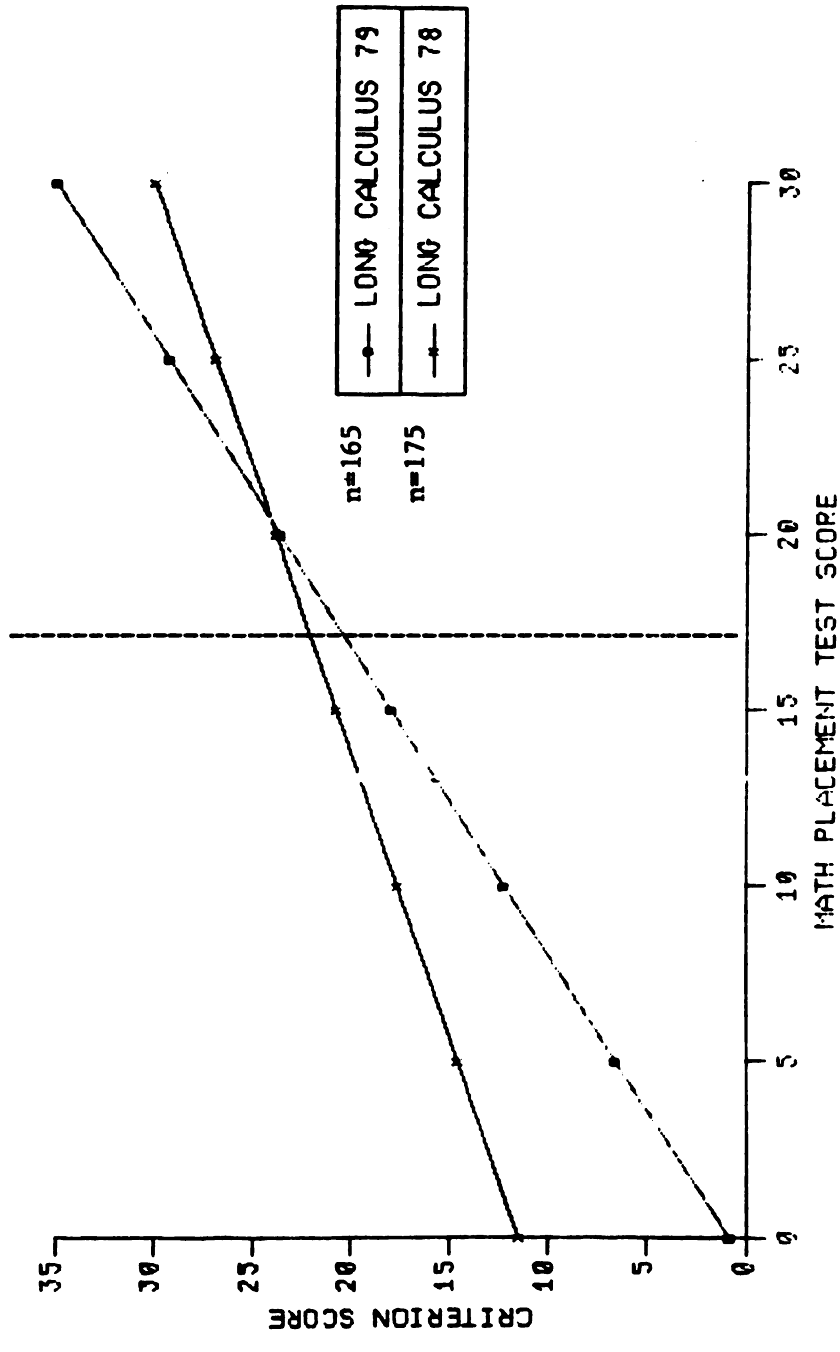


The dashed line represents the current cutting score

Appendix A

Figure 17

REGRESSION OF MATH 112 GRADES ON MATH PLACEMENT TEST SCORES
1978 AND 1979

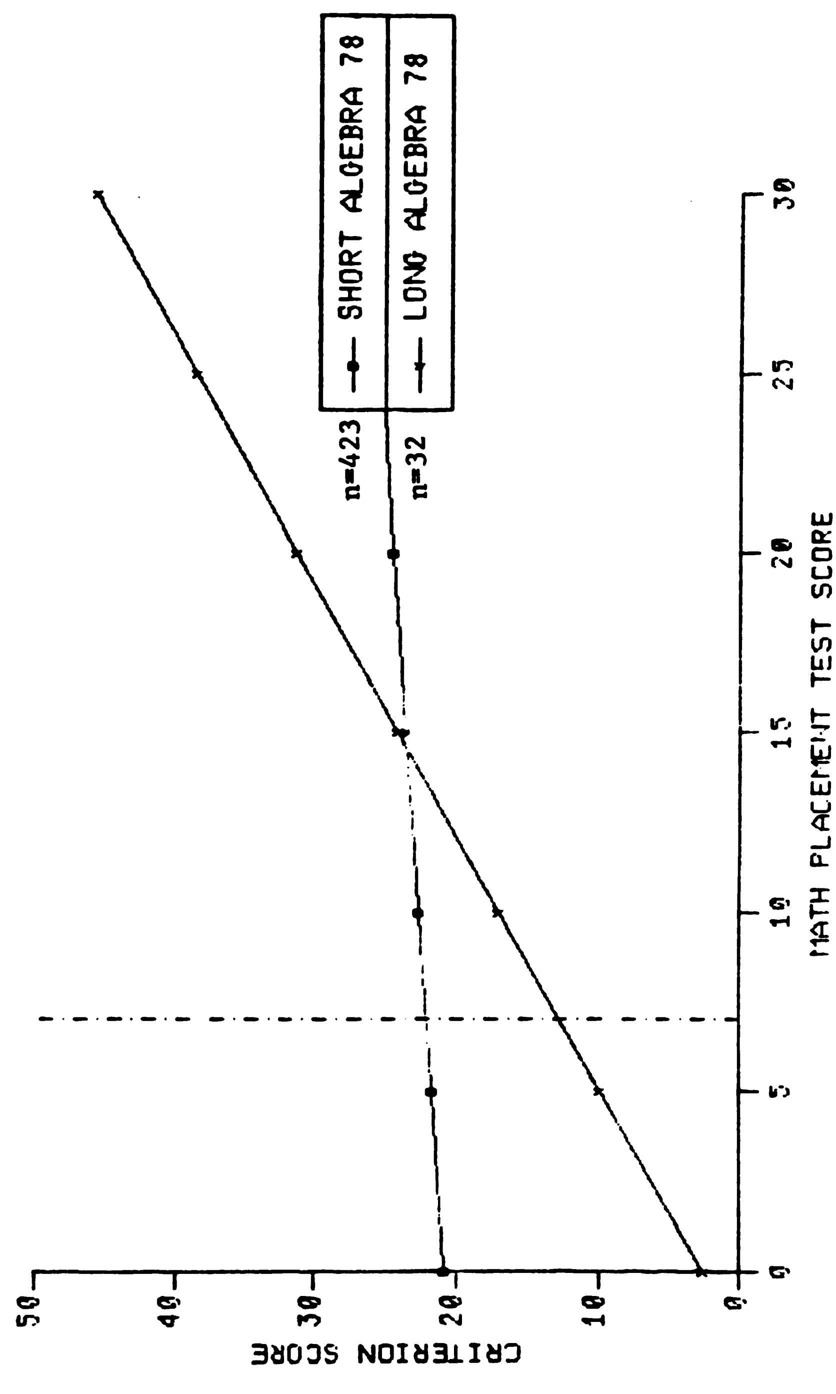


The dashed line represents the current cutting score

Appendix A

Figure 18

REGRESSION OF MATH 108 GRADES ON MATH PLACEMENT TEST SCORES
1978

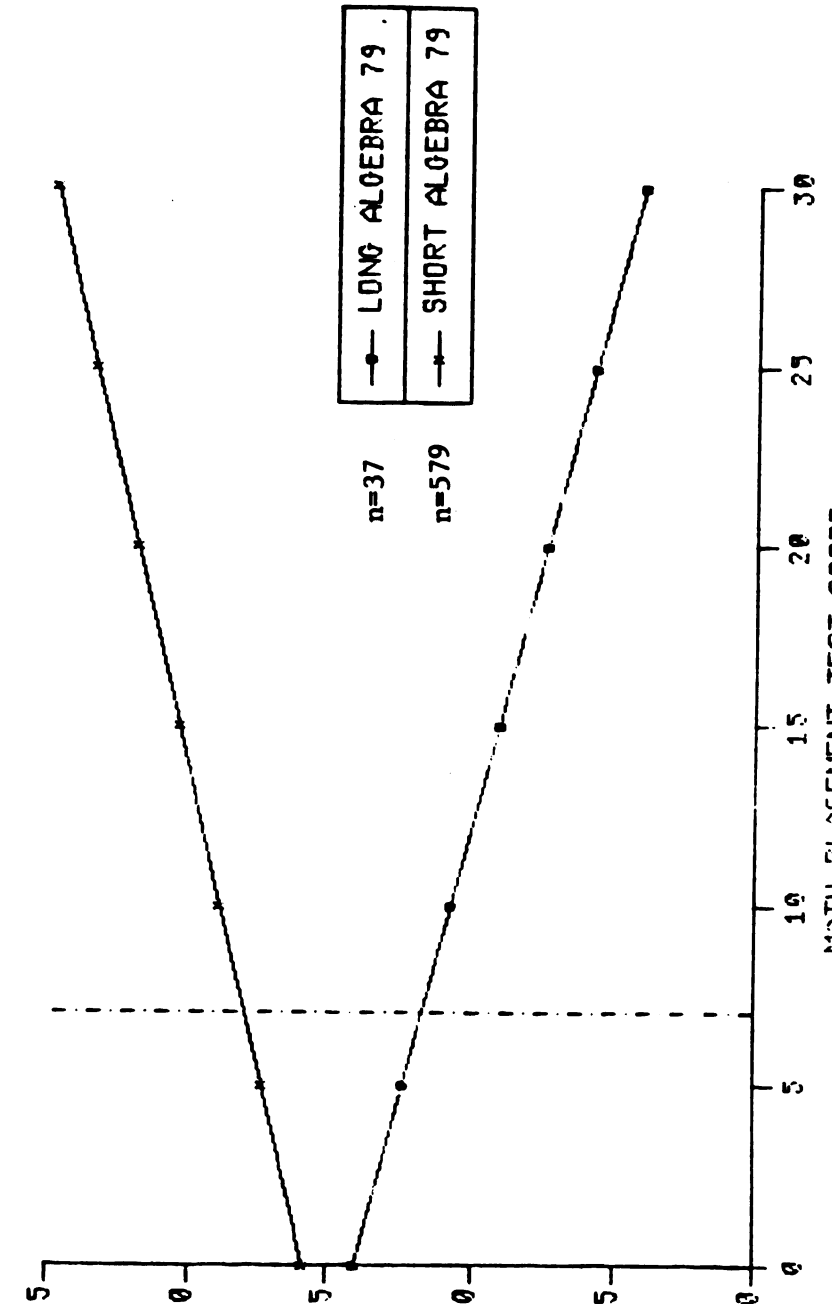


The dashed line represents the current cutting score

Appendix A

Figure 19

REGRESSION OF MATH 108 GRADES ON MATH PLACEMENT TEST SCORES
1979

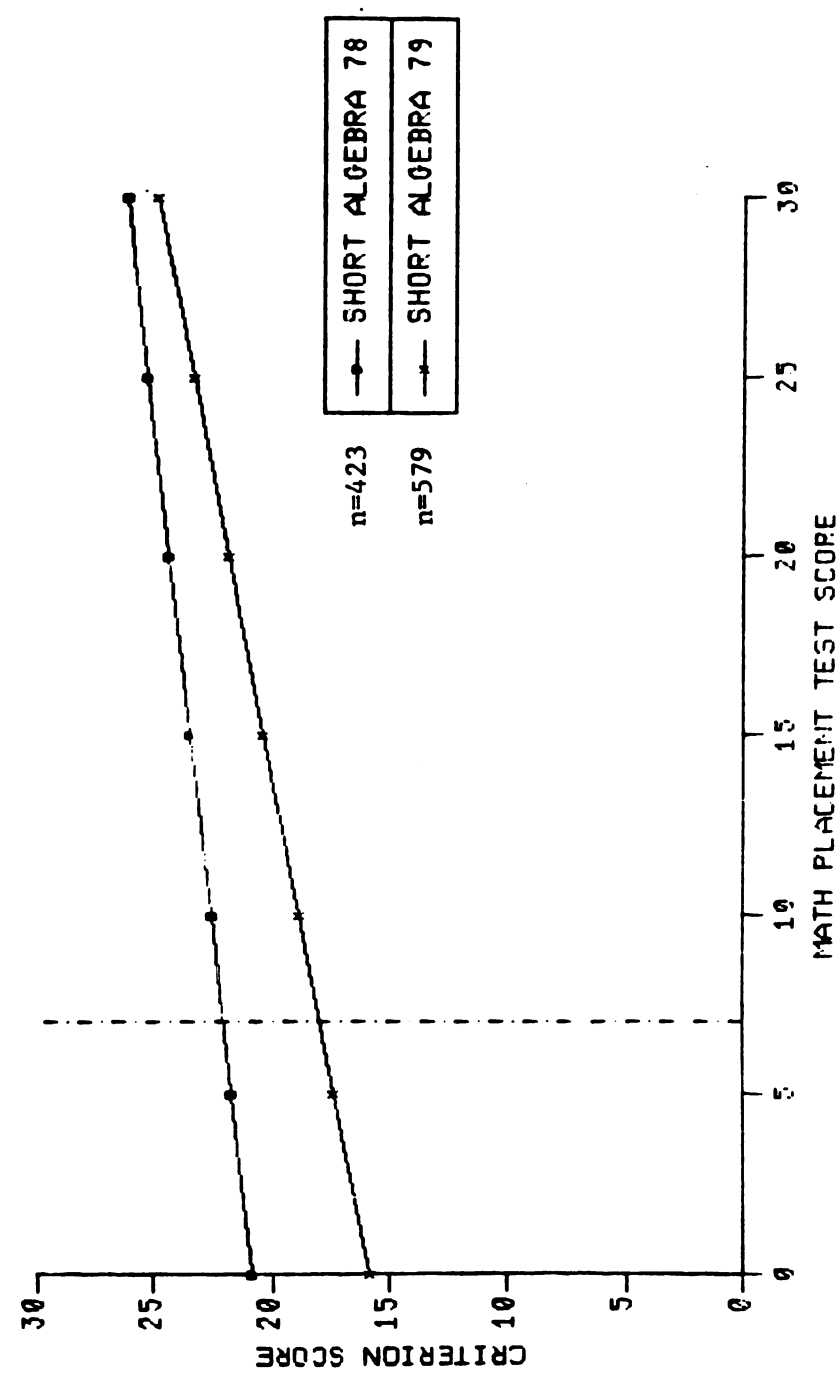


The dashed line represents the current cutting score

Appendix A

Figure 20

REGRESSION OF MATH 100 GRADES ON MATH PLACEMENT TEST SCORES
1978 AND 1979

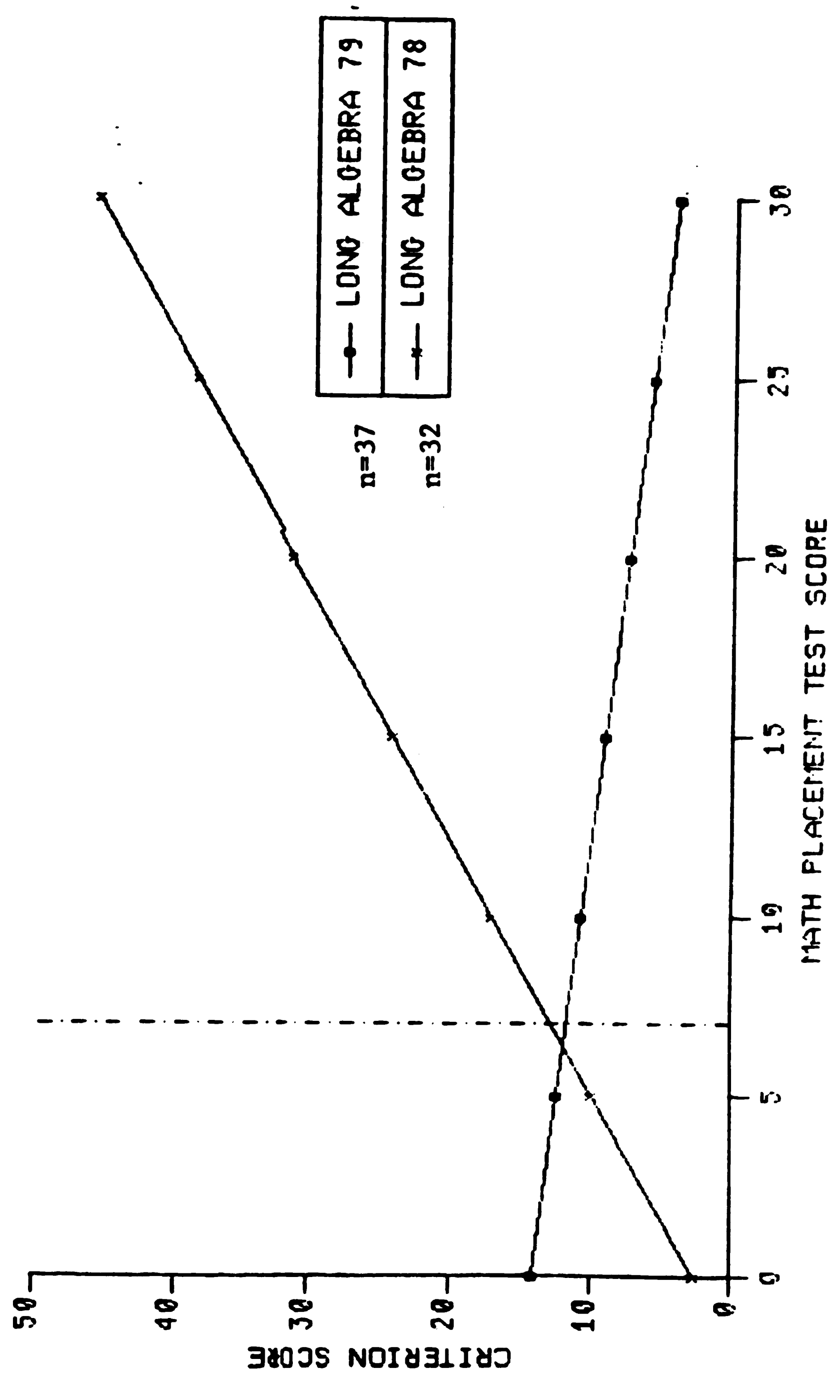


The dashed line represents the current cutting score

Appendix A

Figure 21

REGRESSION OF MATH 108 GRADES ON MATH PLACEMENT TEST SCORES
1978 AND 1979

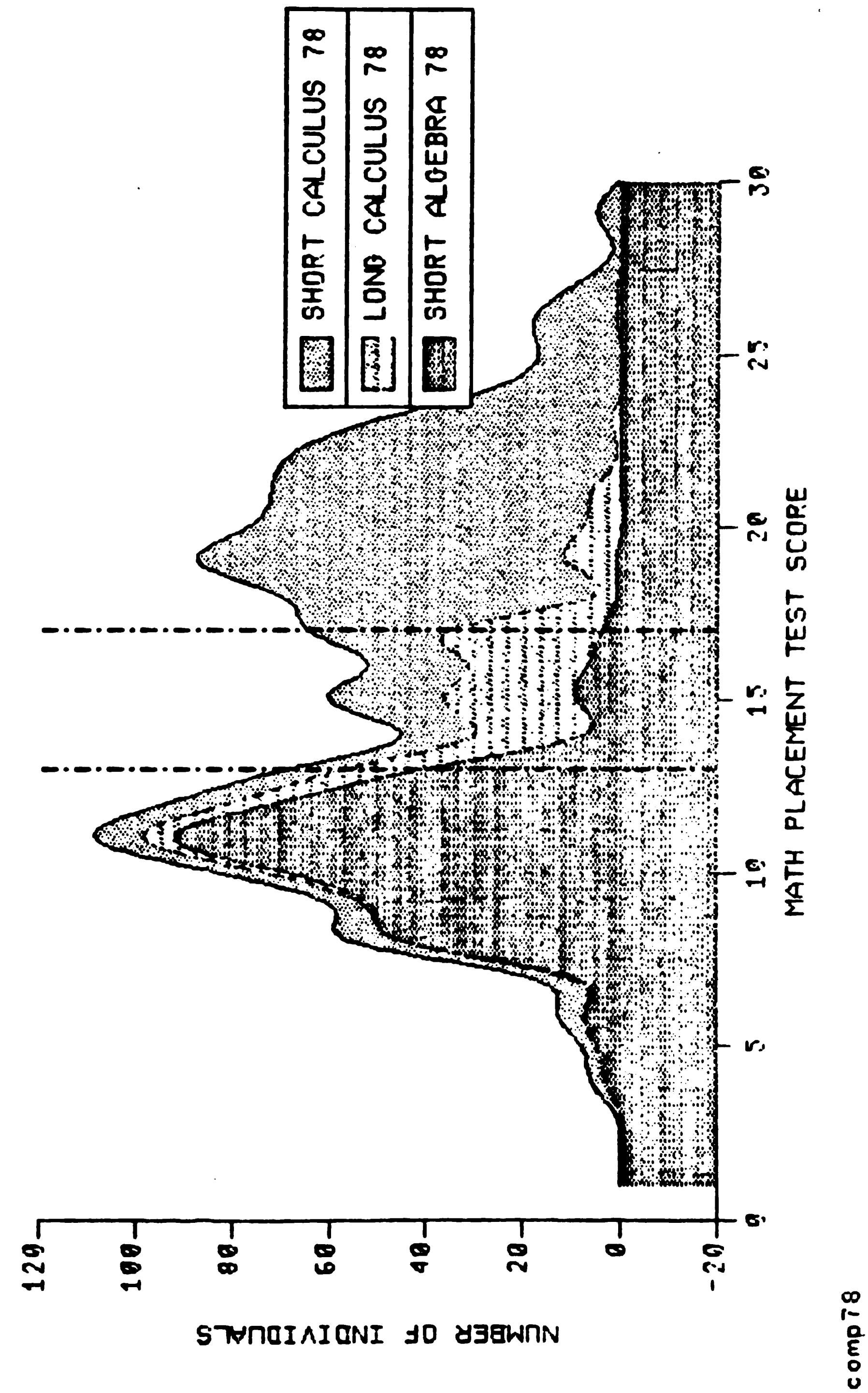


The dashed line represents the current cutting score

Appendix A

Figure 22

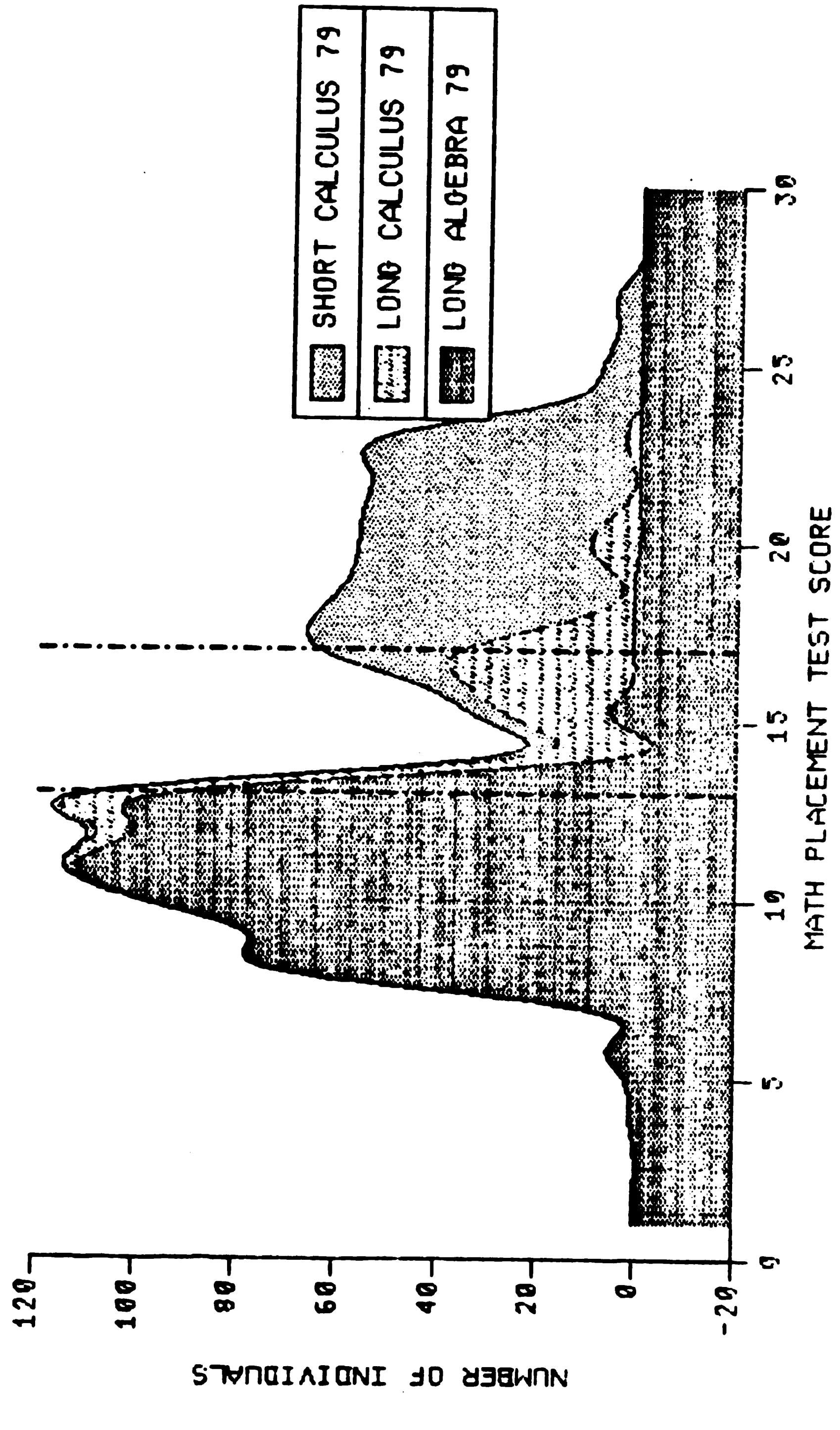
THE NUMBER OF STUDENTS ENROLLING IN EACH COURSE SEQUENCE
BY PLACEMENT TEST SCORE
1978



Appendix A

Figure 23

THE NUMBER OF STUDENTS ENROLLING IN EACH COURSE SEQUENCE
BY MATH PLACEMENT TEST SCORE
1979



comp78

REFERENCE LIST

REFERENCE LIST

Buros, O. K. (ed.) *The seventh mental measurements yearbook*. Highland Park, New Jersey: The Gryphon Press, 1972.

Bloom, B. S. *Human characteristics and school learning.* New York: McGraw-Hill Book Company, 1976.

Cronbach, L. J. Test validation. In R. Thorndike (ed.), *Educational measurement* (2nd ed.). Washington D.C.: American Council on Education, 1971.

Cronbach, L. J., & Gleser, G. C. *Psychological tests and Personnel decisions* (2nd ed.). Urbana, Illinois: University of Illinois Press, 1965.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods: a handbook for research on interactions*. New York: Irvington Publishers, Inc., 1981.

Hills, J. R. Use of measurement in selection and placement. In R. Thorndike (ed.), *Educational measurement* (2nd ed.). Washington D. C.: American Council on Education, 1971.

Joula, A. E. *The placement acceleration examination system at Michigan State University*. Paper presented at the national meeting of the National Council on Measurement in Education, New Orleans, Louisiana, February 1973.

Mehrens, W. A., & Lehmann, I. J. *Measurement and evaluation in education and psychology*. New York: Holt, Rinehart and Winston, Inc., 1973.

Thorndike, R. L. Educational measurement for the seventies. In R. L. Thorndike (ed.), *Educational measurement* (2nd. ed.). Washington D.C.: American Council on Education, 1971.

Wallace, W. L. ACT validity review. In O. Buros (ed.), *Seventh Mental Measurements Yearbook*. Highland Park, New Jersey: The Gryphon Press, 1972.

Willingham, W. W. *College placement and exemption*, New York: College Entrance Examination Board, 1974.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ablon, L. J., & Harvey, J. G., & Pownall, M. W. A modular approach to preparatory mathematics. <u>American Math Monthly</u>, December 1972, <u>79</u>, 1126-1131.

Ahmann, J. S., & Glock, M. D. An evaluation of the effectiveness of a freshman mathematics course. <u>Journal of Educational Psychology</u>, 1959, <u>50</u>(1), 41-45.

Aiken, L. R. Jr. The grading behavior of a college faculty. <u>Educational and Psychological Measurement</u>, 1963, <u>23</u>(2), 319-322.

Aleamoni, L. M. Can grade point average be more accurately predicted? <u>Journal of Educational Psychology</u>, 1977, <u>69</u>(3), 227-230.

Armstrong, R. A. <u>The development of freshman english, chemistry, and mathematics course placement procedures for fall, 1975 freshmen (University of Minnesota, Twin Cities Campus)</u>. (Tech. Rep. Vol. 17, No. 4). Minnesota: Univeristy of Minnesota, Office of Student Affairs, 1976.

Battista, M. The interaction between two instructional treatments of algebraic structures and spatial visualization ability. <u>Journal of Educational Research</u>, 1981, <u>74</u>, 337-341.

Bloom, B. S. <u>Human characteristics and school learning</u>. New York: McGraw-Hill Book Company, 1976.

Boehm, V. R. Differential prediction: a methodological artifact? <u>Journal of Applied Psychology</u>, 1977, <u>62</u>(2), 146-154.

Boyce, R. W. The prediction of achievement in college algebra. <u>Educational and Psychological Measurement</u>, 1964, <u>24</u>(2), 419-421.

Bracht, G. H. Experimental factors related to aptitude-treatment interactions. <u>Review of Educational Research</u>, 1970, <u>40</u>(5), 627-645.

Brooks, S., & Hartz, M. A. Predictive ability of a branching test. Educational and Psychological Measurement, 1978, 38(2), 415-419.

Brooks, S., & Hartz, M. Student placement - a comparison of traditional and computerized branching test administrations. Journal for Research in Mathematics Edcuation, May 1979, 10, 213-216.

Bullock, L. M., & Rigg, Jr., W. C. Relationship of individualized instruction to placment of exceptional children. Exceptional Children, November 1980, 47(1), 224-225.

Buros, O. K. (ed.) The seventh mental measurements yearbook. Highland Park, New Jersey: The Gryphon Press, 1972.

Buros, O. K. Fifty years in testing: some reminiscences, criticisms, and suggestions. Educational Researcher, July/August 77, 6, pp. 9-15.

Burnham, P. S., & Hewitt, B. A. Advanced placement scores: their predictive validity. Educational and Psychological Measurement, 1971, 31, 939-945.

Butzow, J. W., & Linz, L. W., & Drake, R. A. A study of the interrelations of attitude and achievement measures in an audio-tutorial college chemistry course. Journal of Research in Science Teaching, January 1977, 14, 45-49.

Cahen, L. S., & Linn, R. L. Regions of significant criterion differences in aptitude-treatment-interaction research. American Educational Research Journal, 1971, 8(3), 521-530.

Casserly, P. L. What's really happening in advanced placement? - II. College Board Review, Spring 1966, No. 59, 16-22.

Chuang, D. T., & Chen, J. J., & Novick, M. R. Theory and practice for the use of cut-scores for personnel decisions. Journal of Educational Statistics, Summer 1981, 6(2), 129-152.

Corotto, L. V. The prediction of success in initial college mathematics courses. The Journal of Educational Research, January 1963, 56(5), 268-271.

Cronbach, L. J. The two disciplines of scientific psychology. Address of the President at the Sixty-Fith Annual Convention of the American Psychological Association, New York, New York, September 2, 1957.

Cronbach, L. J. Test validation. In R. Thorndike (ed.), Educational measurement (2nd ed.). Washington D.C.: American Council on Education, 1971.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. American Psychologist, February, 1975, pp. 116-127.

Cronbach, L. J., & Gleser, G. C. Psychological tests and ersonnel decisions (2nd ed.). Urbana, Illinois: University of Illinois Press, 1965.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods: a handbook for research on interactions. New York: Irvington Publishers, Inc., 1981.

Daly, J. L., & Stahmann, R. F. The effect of time limits on a university placement test. The Journal of Educational Research, November 1968, 62(3), 103-104.

Dizney, H. F., & Gromen, L. Predictive validity and differential achievement on three MLA - Cooperative Foreign Language Tests. Educational and Psychological Measurement, 1967, 27, 1127-1130.

Domino, G. Differential prediction of academic achievement in conforming and independent settings. Journal of Educational Psychology, 59(4), 256-260.

Dunn, J. E. A study of the University of Arkansas mathematics entrance exam as a placement device. The Journal of Experimental Education, Spring 1966, 34(3), 62-68.

Domino, G. Interactive effects of achievement orientation and teaching style on academic achievement. Journal of Educational Psychology, 1971, 62(5), 427-431.

Eastman, W. M. Several proposals for sounder freshman placement. College Board Review, Summer 1969, No. 72, 27-29.

Ebel, R. L. Must all tests be valid? American Psychologist, 1961, 16, 640-647.

Edwards, R. R. The prediction of success in remedial mathematics courses in the public community and junior college. The Journal of Educational Research, December 1972, 66(4), 157-160.

Elbrink, L. C. A flexible and forgiving testing program a report of the current status of the testing program of the CRIMEL Project. School Science and Mathematics, 1973, 73(8), 686-690.

Francis, R. L., A placement study in analytic geometry and calculus. Educational and Psychological Measurement, 1966, 26, 1041-1046.

Frisbie, D. A. Some limitations of expectancy data used with college-level course placement. Measurement and Evaluation in Guidance, 1977, 9(4), 166-171.

Gilbert, C. D., & Cooper, D. The relationship between teacher/student attitudes and the competency levels of sixth grade students. School Science and Mathematics, October 1976, 76, 469-470.

Glaser, R. Instructional technology and the measurement of learning outcomes: some questions. American Psychologist, 1963, 18(8), 519-521.

Goldman, R. D., & Slaughter, R. E. Why college grade point average is hard to predict. Journal of Educatonal Psychology, 1976, 68(1), 9-14.

Goolsby Jr., T. M. The validity of a comprehensive college sophomore test battery for use in selection, placement, and advisement. Educational and Psychological Measurement, 1966, 26, 977-983.

Goolsby Jr., T. M. Differentiating between measures of different outcomes in the social studies. Journal of Educational Measurement, Fall 1966, 3(3), 219-222.

Hall, D. T. The effect of teacher-student congruence upon student learning in college classes. Journal of Educational Psychology, 1970, 61(3), 205-213.

Hills, J. R., & Gladney, M. B. Factors influencing college grading standards. Jounal of Educational Measurement, Spring 1968, 5(1), 31-39.

Hills, J. R. Factor-Analyzed abilities and success in college mathematics. Educational and Psychological Measurement, 1957, 17, 615-622.

Hills, J. R. Use of measurement in selection and placement. In R. Thorndike (ed.), Educational measurement (2nd ed.). Washington D. C.: American Council on Education, 1971.

Howlett, J. A study of placement methods for entering college freshmen in the proper mathematics sequence at Michigan Technological University. The Mathematics Teacher, December 1969, 651-659.

Humphreys, L. G. The phenomena are ubiquitous - but the investigator must look. Journal of Educational Psychology, 1976, 68(5), 521.

Hunter, J. E., & Schmidt, F. L. Differential and single-group validity of employment tests by race: a critical analysis of three recent studies. Journal of Applied Psychology, 1978, 68(1), 1-11.

Johnson, R. W., & Keochakian, S. V., & Morningstar, M., & Southworth, J. A., Validation of freshmen orientation test battery. Educational and Psychological Measurement, 1968, 28, 437-440.

Joula, A. E. The placement acceleration examination system at Michigan State University. Paper presented at the national meeting of the National Council on Measurement in Education, New Orleans, Louisiana, February 1973.

Kahneman, D. Validity and nonlinear heteroscedastic models. Personnel Psychology, 1962, 15, 1-11.

Koffler, S. L. A comparison of approaches for setting proficiency standards. Journal of Educational Measurement, Fall 1980, 17(3), 167-178.

Kohler, E. T. The relationship between the Cooperative Mathematics Test, Algebra III, Act Mathematics Usage Test, Act Composite, and grade point average in college algebra. Educational and Psychological Measurement, 1973, 33, 929-931.

Kolen, M. J. Expectency of success for new students in nine mathematics courses and application to mathematics placement. (Report No. 79-101), Evaluation and Examination Service: The University of Iowa, June 1979.

Kurtz, R. Who should place college freshmen in mathematics? The Mathematics Teacher, November 1969, 557-559.

Lindvall, C. M. A review of selected publications on behavioral objectives. Jounal of Educational Measurement, Spring 1972, 9(1), 75-81.

Lord, F. M. Cutting scores and errors of measurement. Psychometrika, 1962, 27(1), 19-30.

Lubin, A. The interpretation of significant interaction. Educational and Psychological Measurement, 1961, 21(4), 807-817.

Lunneborg, C. E., & Lunneborg, P. W. The prediction of different criteria of law school performance. Educational and Psychological Measurement, 1966, 26, 935-944.

Lunneborg, P. W., & Lunneborg, C. E. The differential prediction of college grades from biographic information. Educational and Psychological Measurement, 1966, 26, 917-925.

Mauger, P. A. Is the prediction of grades fleeting only in Illinois? Journal of Educational Psychology, 1976, 68(5), 520.

McKeachie, W. J., & Lin, Y., & Milholland, J., & Isaacson, R. Student affiliantion motives, teacher warmth, and academic achievement. Journal of Personality and Social Psychology, 1966, 4(4), 457-461.

Mcleod, D. B., & Adams, V. M. The interaction of field independence with discovery learning in mathematics. Journal of Experimental Education, Fall 1979, 48, 32-35.

McLeod, D. B., & Adams, V. M. The interaction of field independence with small-group instruction in mathematics. Journal of Expermental Education, Winter 1979-80, 48(2), 118-124.

McLeod, D. B., & Adams, V. M. Aptitude-treatment interaction in mathematics instruction using expository and discovery methods. Journal for Research in Math Education, May 1980, 225-234.

Mehrens, W. A., & Lehmann, I. J. Measurement and evaluation in education and psychology. New York: Holt, Rinehart and Winston, Inc., 1973.

Mezynski, K., & Stanley, J. C. Advanced placement oriented calculus for high school students. Journal of Research in Mathematics Education, November 1980, 11, 347-355.

Michael, J. J., & Michael, W. B., The validity of a pre-statistics survey test of basic mathematical skills in relation to achievment in the first course in psychological statistics. Educational and Psychological Measurement, 1969, 29, 915-919.

Mills, C. N. A comparison of three methods of establishing cut-off scores on criterion referenced tests. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, 1982.

Nazarro, J. R., & Todorov, J. C. Student ability and individualized intruction. Journal of College Science Teaching, 1972, 2(6), 29-30.

Neuringer, C. Cutoff: a Fortran program for the computation of classification efficiency probabilities at various cutting scores. Educational and Psychological Measurement, 1979, 39, 659-661.

Northup, L. L. Method of placing engineering students in mathematics and chemistry. Paper presented at the 88th Annual Convention of the American Society of Engineering Education, Univeristy of Massachusetts, June, 1980.

Osburn, H. G., & Melton, R. S. Prediction of proficiency in a modern and traditional course in begining algebra. Educational and Psychological Measurement, 1963, 23(2), 277-287.

Peterson, P. L. Aptitude X treatment interaction effects of teacher structuring and student participation in college instruction. Journal of Educational Psychology, 1979, 71(4), 521-533.

Reilly, R. R. Student placement by a maximum liklihood procedure. Measurement and Evaluation in Guidance, 1974, 7(1), 8-15.

Reiner, J. R. Differential effects of initial course placement as a function of ACT mathematics scores and high-school rank-in-class in predicting general performance in chemistry. Educational and Psychological Measurement, 1971, 31, 977-981.

Rulon, P. J., & Tiedeman, D. V., & Tatsuoka, M. M., & Langmuir, C. R. Multivariate statistics for personnel classification. New York: John Wiley & Sons, Inc., 1967.

Salomon, G. Heuristic models for the generation of aptitude treatment interaction hypotheses. Review of Educational Research, 1972, 42(3), 327-343.

Seigle, W. F. Prediction of success in college mathematics at Washburn University. Journal of Educational Research, 1954, 47, 577-588.

Serlin, R. C. & Levin, J. R. Identifying regions of significance in aptitude-by-treatment-interaction research. American Educational Research Journal, Fall 1980, 17(3), 389-399.

Shevel, L. R., & Whitney, D. R. Predictive validity of the mathematics placement examination. Educational and Psychological Measurement, 1969, 29, 895-901.

Skakun, E. N., & Kling, S. Comparability of methods for setting standards. Journal of Educational Measurement, Fall 1980, 17(3), 229-235.

Snow, R. E., & Tiffin, J., & Seibert, W. F. Individual differences and instructional film effects. Journal of Educational Psychology, 1965, 56(6), 315-326.

Snow, R. E. Aptitude, learner control, and adaptive instruction. Educational Psychologist, 1980, Vol. 15, No. 3, 151-158.

Solomon, D. Finding optimal instructional programs and procedures for individual students. Teaching Sociology, April 1980, 7(3), 202-326.

Sweet, P. R., & Nuttall, R. L. The effects of a tracking system on student satisfaction and achievement. American Educational Research Journal, 1971, 8(3), 511-520.

Tallmadge, G. K., & Shearer, J. W. Interactive relationships among learner characteristics, types of learning, instructional methods, and subject matter variables. Journal of Educational Psychology, 1971, 62(1), 31-38.

Tamir, P. The relationship between achievement in biology and cognitive preference styles in high school students. British Journal of Educational Psychology, 1976, 46, 57-67.

Tatsuoka, M. M. Multivariate analysis. New York: John Wiley & Sons, Inc., 1971.

Thorndike, R. L. Educational measurement for the seventies. In R. L. Thorndike (ed.), Educational measurement (2nd. ed.). Washington D.C.: American Council on Education, 1971.

Waits, B. Individualized instruction in large enrollment mathematics courses. American Math Monthly, 1973, 80, 307-310.

Walker, H. M., & Lev, J. Statistical Inference. New York: Holt, Rinehart and Winston, 1953.

Wallace, W. L. ACT validity review. In O. Buros (ed.), Seventh Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press, 1972.

Wheeler, R. G. Computer predicts student success as they register. College Management, April 1971, 36-38.

Wick, M. E. A study of the factors associated with success in first-year college mathematics. The Mathematics Teacher, November 1965, 642-648.

Willingham, W. W. College placement and exemption, New York: College Entrance Examination Board, 1974.

Wiseman, S. The effect of restriction of range upon correlation coefficients. British Journal of Educational Psychology, 1967, 37, 248-252.

Yoshida, H. Effects of drill practice on aptitude in the learning of mathematics. Journal of Educational Psychology, 1980, 72(5), 706-715.