The effect of formative testing, prescribed remediation, and retesting on student performance in calculus
by David Allen Thomas

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Education
Montana State University

Abstract:
During the Fall Quarter of 1982, the researcher conducted an experiment at Montana State University involving six sections of Math 121, freshman calculus. The purpose of the study was to determine the effect of formative testing, diagnostic and prescriptive remediation, and retesting on student performance in calculus in a conventional classroom instructional setting. The experimental treatments were limited to the use of weekly formative quizzes and did not intrude into lecture time or homework practices.

The students in the six sections were divided randomly into two groups for the study, an experimental group and a control group. Experimental and control group students all took the same quizzes and examinations. The control group students received conventional forms of written feedback on their quizzes and remedial assistance based on their own perceptions of need and initiative in seeking help. Experimental group students received a detailed error analysis for each formative quiz and remediation individualized to address the diagnosed errors.

At the end of the quarter, student performance was evaluated as the sum of the three one hour examination scores plus the final examination score. Of 500 possible points, the control group had a mean score of 375.9 and the experimental group had a mean of 402.3 for students finishing the course. The analysis of variance showed significant main effect differences for the experimental and control groups. The p-value for the difference in mean performance for entering freshmen was .0186. No aptitude-treatment interaction was found.

On the basis of this analysis, the researcher concluded that student performance in calculus can be significantly improved by using formative testing, prescribed remediation, and retesting, procedures which do not require extensive intrusion into the conventional lecture- discussion format of college calculus.

THE EFFECT OF FORMATIVE TESTING, PRESCRIBED REMEDIATION, AND RETESTING ON STUDENT PERFORMANCE IN CALCULUS

by

David Allen Thomas

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Education

Montana State University
Bozeman, Montana

July, 1983

APPROVAL

of a thesis submitted by

David Allen Thomas

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

26 JULY 1983
Date

Leroy J. Casagranda
Chairperson, Graduate Committee

Approved for the Major Department

July 26, 1983
Date

Henry W. Warner
Head, Major Department

Approved for the College of Graduate Studies

7-28-83
Date

Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan, 48106, to whom I have granted "the exclusive right to reproduce and distribute copies of the dissertation in and from microfilm and the right to reproduce and distribute by abstract in any format."

Signature David A Thomas

Date 7/26/83

ACKNOWLEDGMENTS

The writer is grateful for the help provided by many people during the completion of this study. He wishes to pay special thanks to his wife and daughter for their continuing support and encouragement.

The writer is especially grateful to Dr. Leroy Casagranda, the chairman of his graduate committee. No other individual has contributed so much to the writer's professional development. His help and friendship have been greatly appreciated. He also expresses gratitude to the other members of his graduate committee.

The writer also acknowledges the cooperation of the staff of the Department of Mathematical Sciences at Montana State University. In particular, he is grateful to Dr. Kenneth Tiahrt, department chairman, for his assistance and cooperation in this study.

TABLE OF CONTENTS

TABLE OF CONTENTS --- CONTINUED

TABLE OF CONTENTS --- CONTINUED

LIST OF TABLES

LIST OF TABLES --- CONTINUED

LIST OF TABLES --- CONTINUED

ABSTRACT

During the Fall Quarter of 1982, the researcher conducted an experiment at Montana State University involving six sections of Math 121, freshman calculus. The purpose of the study was to determine the effect of formative testing, diagnostic and prescriptive remediation, and retesting on student performance in calculus in a conventional classroom instructional setting. The experimental treatments were limited to the use of weekly formative quizzes and did not intrude into lecture time or homework practices.

The students in the six sections were divided randomly into two groups for the study, an experimental group and a control group. Experimental and control group students all took the same quizzes and examinations. The control group students received conventional forms of written feedback on their quizzes and remedial assistance based on their own perceptions of need and initiative in seeking help. Experimental group students received a detailed error analysis for each formative quiz and remediation individualized to address the diagnosed errors.

At the end of the quarter, student performance was evaluated as the sum of the three one hour examination scores plus the final examination score. Of 500 possible points, the control group had a mean score of 375.9 and the experimental group had a mean of 402.3 for students finishing the course. The analysis of variance showed significant main effect differences for the experimental and control groups. The p-value for the difference in mean performance for entering freshmen was .0186. No aptitude-treatment interaction was found.

On the basis of this analysis, the researcher concluded that student performance in calculus can be significantly improved by using formative testing, prescribed remediation, and retesting, procedures which do not require extensive intrusion into the conventional lecture- discussion format of college calculus.

CHAPTER ONE

INTRODUCTION

Introduction

Mathematics, the physical sciences, engineering in all its forms, and many other disciplines share a characteristic orientation toward sequential, task-oriented development. In these disciplines, a student progresses from the basic to the more complex by mastering increasingly more sophisticated concepts and skills in a well defined order. Bloom (1976) argues that failure at any point in such a sequence to master the task at hand not only weakens the student's understanding at the point of failure, but also inhibits his ability to master subsequent tasks. In light of this characteristic, the curriculum must provide some way of dealing with student errors and misunderstanding.

At Montana State University, the first quarter of calculus was Math 121. The course briefly reviewed algebraic skills before beginning differential calculus. Approximately two-thirds of the quarter was spent on this topic, which deals primarily with a study of the rate of change of functional values and associated applications to science and engineering. The last topic introduced in the course was integral calculus, which is often applied to the problem of finding the area under a given curve. A thorough investigation of integration was presented in Math 122 and subsequent course offerings. The problem of dealing with student errors and misunderstandings in Math 121 was the focus of this study.

The Mastery Learning paradigm proposed by Bloom (1976) stresses large group instruction, use of formative tasks, and remediation to achieve mastery. The classroom teacher's role is extended beyond conventional patterns of presentation and evaluation to include diagnosis of student characteristics and prescription of individualized instruction and remediation.

Another systematic approach to individualization is the Keller, or PSI, Plan. PSI (Personalized System of Instruction) programs stress the use of the written word as the primary vehicle of communication for course content. Tutors are used extensively in PSI programs to respond to student questions and administer unit quizzes, which must be repeated until mastery is achieved (Keller, 1968). The primary task of course lectures is to motivate the students and direct the course of study. Research conducted on PSI courses in psychology (Cooper and Greiner, 1971), child development (Semb, 1974), and pre-calculus mathematics (Haver, 1978) indicates that the PSI method often results in higher student performance than conventional lecture-recitation methods of instruction.

Use of PSI as an instructional format in college calculus has resulted in seemingly contradictory findings by different researchers. Pascarella (1977a, 1977b, 1978), Peluso and Baranchik (1977), and Kulik, Kulik and Smith (1976) found that students in PSI sections of calculus scored higher on exams than students in conventionally taught sections. Klopfenstein (1977) and Thompson (1980) found no significant differences between exam scores of PSI and conventionally taught students.

In recent years, research on the nature and importance of individual learning styles (Dunn and Dunn, 1978; Gregorc, 1979) has led to the use of these concepts in the individualization of instruction (Kusler, 1979). This study makes use of learning styles research as a vehicle for identifying appropriate modes of remediation for individual students.

In introductory calculus, the problem of dealing with student misunderstanding is complicated by the wide range of abilities and degrees of preparation in the student population. In partial answer to this problem, the Montana State University Department of Mathematical Sciences operates a Learning Center every school day from 8 am to 5 pm. At the learning center, students have access to tutorial help from qualified staff members and graduate students. Students may seek assistance on homework problems, clarification of topics presented in class, or review in preparation for an examination. This service has been offered since 1977 and was originally provided to provide students with assistance when their instructors were unavailable (Tiahrt, 1982).

Statement of the Problem

Student failure in introductory calculus has created problems both for the student personally and for college departments faced with scheduling trailer sections for students who have delayed taking course work in their major area because of failure in prerequisite calculus courses. The purpose of this study was to determine the effect of formative testing, prescribed remediation, and retesting on student performance in freshman calculus when remediation is adjusted to accommodate individual learning styles.

Need for the Study

In 1981, approximately 30% of the students enrolled in introductory calculus at Montana State University received a grade of D or F. This was typical of failure rates in comparable courses across the country (Struik and Flexer, 1977). Additionally, this failure rate created of considerable personal disappointment to both students and parents. When the effect of such failures was seen in the broader perspective of the university disciplines for which mathematics is a foundation, the failures took on further significance. Student failures created a need for trailer sections of Math 121 and other courses sequenced with Math 121 in other departments. As an example, consider the impact of student failure in Math 121 on course scheduling in the Engineering Mechanics Department. With enrollments in engineering growing every year, the university was facing shortages in qualified staffing and facilities. Ultimately, engineering students who failed calculus and had to repeat the course added to the burden of the engineering departments by creating scheduling problems when mathematics prerequisites were not completed ahead of specified engineering courses as planned (Williams, 1982). The same type of scheduling problems occurred in other departments as well. Thus, in order to deal with student failures both as a personal problem and as a university problem, there was a need for research addressing the improvement of instruction in introductory calculus.

A review of the literature revealed studies which support the use of Mastery Learning and PSI strategies as effective approaches to improving student performance in calculus (Pascarella, 1977a and Peluso and Baranchik, 1977). The review also revealed studies which concluded that the Mastery Learning and PSI strategies made no difference in student performance when compared to conventional methods of instruction (Klopfenstein, 1977). The variety of instructional formats and procedures employed in the studies cited above and in other studies discussed in Chapter Two of this paper ranged from moderate modifications of conventional classroom procedure to major diversions from traditional procedures. This study tested the effectiveness of certain Mastery Learning and PSI strategies in an instructional setting that was close to conventional classroom procedures.

Finally, the solution to the problem of failures in calculus does not lie in college pre-calculus remedial courses. Whitesitt (1980) showed that pre-calculus courses such as college algebra and trigonometry offered at Montana State University were not effective in preparing students for calculus. Thus, the problem of student errors and misunderstanding in calculus must be dealt with as a part of the calculus curriculum.

Questions Answered

This study answered nine questions.

1. Did the random assignment of students to treatment groups result in an experimental and control group of equal ability?

2. Did the random assignment of students to instructors result in the instructors teaching groups of equal ability?

3. For students of known ability, did the experimental group students perform at a different level than the control group students?

4. Was there an aptitude-treatment interaction? In particular, when the difference in exam performance of control group and experimental group students of relatively low ability was compared to the difference in performance of control and experimental group students of relatively high ability, were the differences constant or did one treatment benefit students of one ability level more than students of another ability level?

5. For students of unknown ability, did the experimental group students perform at a different level than the control group students?

6. Was the standard deviation of the experimental group scores different than the standard deviation of the control group scores for

each of the four examinations?

7. Was there any relationship between the amount of time a student spent in remediation with his/her instructor and his/her overall performance on examinations?

8. Was there any relationship between a student's expected GPA and his/her overall examination performance?

9. Was there any relationship between a student's learning style and his/her overall examination performance?

These nine questions were restated operationally in Chapters 3 and 4.

General Procedures

Arrangements were made through the chairman of the Department of Mathematical Sciences at Montana State University for four Math 121 instructors to participate in the study. These individuals taught a total of six sections of introductory calculus. One of the instructors was a visiting professor with many years teaching experience at the university level. The other three instructors were Graduate Teaching Assistants in the Mathematics Department with varying degrees of experience at the secondary and college level.

To control for differences between instructors, each instructor taught roughly the same number of experimental and control group students. Two instructors taught only one section each of Math 121. The students in each of these sections were randomly assigned membership in either the experimental group or control group, thus splitting the students in each section. Two instructors taught two sections each of Math 121. In both cases, these sections met at consecutive hours. For each of these instructors, one section was assigned to the experimental group and one section to the control group. The assignments were all made the second week of class after the last day to add a class had passed.

The researcher used the expected GPA, a statistic generated by the university for entering freshmen, as a measure of aptitude or ability to do college work. All the entering freshmen in the Math 121 sections in the study were ranked by the researcher on the basis of their expected GPA scores. The top third of the ranking was labeled the high ability group, the low third the low ability group, and the middle of the list was the middle ability group. This information was used in the data analysis to look for interactions of treatment with ability, but it was not made available to the instructors or students. Thus, this classification had no bearing on the treatments given.

During the third week of class, Gregorc's Learning Styles Delineator (Gregorc, 1982) was administered to the students. At the end of the third week of class, students in the six sections began a series of weekly formative tasks (Appendix A). These tasks took the form of quizzes taken in class. Instructors graded the quizzes of the experimental group students and marked a table of specifications (Appendix B) for each student characterizing the student's errors both by subject matter and level of difficulty. When the student reported to the instructor's office the next day to pick up his quiz, the instructor discussed the student's error analysis and remediated (Appendix D) the student on the indicated objectives, taking into account the student's preferred learning style (Appendix E). Following the remediation, a retest (Appendix C) was given to the student to complete within a few days and return to the instructor. For the experimental group, the instructor directed the student's attention to the diagnosed areas of concern.

The control group students received identical quizzesat the same time as the experimental group students. The classroom teachers evaluated these papers in the conventional manner and returned them to the students with written feedback but without a systematic error analysis. Students reported to the instructor in the same manner as the experimental group students to claim their quizzes. At all times, students in the control group were free to seek help from their instructors and initiate any discussion concerning their progress or perceived problems. For the control group, the student initiated remediation by requesting help.

For both groups, the measure of performance was a series of three one hour exams and a two hour final exam (Appendix F). All students took the same exams at the same time and had their papers scored by the same committee of instructors. All students had equal access to help in the Math Learning Center. The treatments and procedures described here applied only to the use of remedial procedures relevant to the formative tasks. All questions arising from classroom discussions, homework assignments, etc. were handled by the instructors in a conventional manner without reference to the strategies outlined in this study.

Analysis of the data was accomplished using computer facilities at Montana State University. The statistical software package MSUSTAT (Lund, 1978) was used . For each of the three hour exams and the two hour final, the mean performance levels for the control and experimental groups were compared using a 2X3 ANOVA in which the interaction of treatment with ability was examined. Mendenhall and Reinmuth (1982) state that the ANOVA is an appropriate technique for identifying important independent variables in a study and how they interact and affect the dependent variable. For identifying the strength of a relationship between variables, Linder (1979) states that the correlation coefficient is most useful. Thus, Pearson correlation coefficients were calculated for the relationships between overall performance on examinations and student ability level as measured by the expected GPA and between overall examination performance and the amount of time spent in remediation. The effect of student learning style to overall performance was examined using ANOVA.

Limitations and Delimitations

This study was limited to students enrolled in Math 121 at Montana State University during Fall quarter, 1982.

The study was delimited to addressing the instructional value of formative testing, prescribed remediation, and retesting in calculus with regard to examination performance. No questions were asked regarding affective aspects of the instructional procedures or other possible dependent variables or outcomes.

Definitions

Terms defined for this study:

1. Ability or Ability Level referred to a student's relative position with regard to all the other students in the study in a ranking based on expected MSU grade point average. Students ranked in the top third of this listing were designated as being of "high" ability. Students ranked in the bottom third of this listing were designated as being of "low" ability. The remaining third was designated "average".

2. Formative Testing or Evaluation is a form of criterion based testing used to diagnose student errors and misunderstandings for the purpose of identifying instructional strategies appropriate to the student's needs. In this study, students receiving formative testing were remediated on the basis of their individual learning styles. Thus, two students who answered the same problems incorrectly on a formative task were remediated over the same content material. However, if the students had different learning styles, the format for the remediation was different for the two students.

3. A mathematics Learning Center was a room or group of rooms where students went for tutorial help or evaluation. At Montana

State University, students enrolled in pre-calculus classes went to a different room than students enrolled in calculus or more advanced courses. The MSU learning center was staffed by instructors, graduate teaching assistants, and advanced undergraduates with a strong background in mathematics.

4. Learning Style in this study refers to the range of cognitive behaviors by which individuals perceive and process information in a learning environment. The four orientations towards information processing used in this study are identified as concrete/sequential, abstract/random, abstract/sequential, and concrete/random (Gregorc, 1982).

5. Mastery Learning refers to a theory of instruction developed by Bloom (1976). The essential features of mastery learning require the instructor to provide the student several specific services. First, directions, demonstrations, and explanations are provided in large group instruction. Second, the student's time is utilized in such a way as to provide sufficient time on task practice of each topic to be mastered. Third, individualized corrective feedback and reinforcement are used to develop both mastery of each topic and a positive attitude toward learning.

6. In this study, Prescribed Remediation refers to a process whereby the tutor uses the diagnosis provided by the formative test to determine the content and difficulty level of the materials used for remediation.

7. PSI or Personalized System of Instruction is a theory of instruction developed by Keller (1968). The most distinctive features of the system are as follows: self-pacing, a unit perfection requirement, use of lectures as vehicles of motivation rather than sources of critical information, dependence upon written student-teacher communication, and the use of proctors in evaluation, feedback, and reinforcement.

8. Retesting refers to the practice of evaluating a student's knowledge after remediation using a parallel form of the unit test.

9. A Summative Evaluation was used to establish a grade or ranking and was not used in a formative sense.

10. A Table of Specifications is a diagnostic report form used to classify errors on a test. In this study, the table listed all the learning objectives for a unit down the left margin and the headings " notation/definition, direct application, synthesis, and abstract" across the top. Inserted in this matrix was a list of the problem numbers from the formative quiz. By inserting the problem number in the correct row and column, the content of the problem and its level of difficulty were characterized.

11. A Tutor in this study was the general term used to denote learning center personnel. These individuals were either instructors, graduate teaching assistants, or advanced undergraduates with a strong background in mathematics.

12. Unstructured Help refers to informal student-tutor interaction in which the student determines the substance of the discussion or problems to be addressed.

Summary

Student failure (grade of D or F) rates in college and university introductory calculus courses in the United States are typically in the range of 30% - 35% of course enrollment. The percentage of failures in freshman calculus at Montana State University has also typically fallen in this range. A study of the pre-calculus courses offered at MSU has shown that they are generally not effective in preparing a student for success in calculus (Whitesitt, 1980). The purpose of this study was to examine the effect of formative testing, prescribed remediation, and retesting on student performance in introductory calculus when remediation is adjusted to accommodate a student's preferred learning style.

The data for this study was collected during the Fall quarter, 1982 at Montana State University. A sample of approximately 200 students was randomly selected from the students enrolled in the course (700 - 800). These subjects were divided into a control and an experimental group for treatment. The experimental group received formative testing, prescribed remediation, and retesting with remediation individualized to accommodate the student's learningstyle as indicated by Gregorc's Learning Styles Delineator (Appendix F).

The control group students took the same quizzes but did not receive the diagnostic treatments offered the experimental group. Feedback was in conventional form, written comments on the quizzes, etc.

At the end of the quarter, the data was analyzed using the software statistical package MSUSTAT (Lund, 1978) on the CP6 computer at Montana State University. The analysis made use of one and two way analysis of variance to compare examination performance for the experimental and control groups, Pearson correlations to measure the strengths of relationships between variables, and F tests to compare examination variance scores.

CHAPTER TWO

REVIEW OF THE LITERATURE

Introduction

For the purpose of this study, the literature was reviewed with regard to the following major topics: the theoretical frameworks of Mastery Learning and Personalized Systems of Instruction (PSI), Learning Styles, feedback, remediation, and retesting as instructional practices, affective aspects of testing and evaluation procedures, interaction effects of various instructional practices with aptitude, preparation or readiness, and motivation, "no-difference" studies, and organizational and managerial practices in PSI and Learning Center environments.

Mastery Learning and Personalized Systems of Instruction

At the foundation of the Mastery Learning paradigm is a rejection of the notion that student achievement should be normally distributed. Bloom (1981) states

> There is nothing sacred about the normal curve. It is the distribution most appropriate to chance and random activity. Education is a purposeful activity and we seek to have the students learn what we have to teach. If we are effective in our instruction, the distribution of achievement should be very different from the normal curve (p. 155).

This statement is based in part on Carroll's (1963) view that aptitude is the amount of time required by a learner to master a learning task. Given a group of learners normally distributed with respect to aptitude for some subject, and given an instructional strategy that treats all students alike without regard for personal differences, then the end result will be a normal distribution on an appropriate measure of achievement. Bloom (1981) concludes, however, that if

> ... the kind and quality of instruction and the amount of time available for learning are made appropriate to the characteristics and needs of each student, the majority of students may be expected to achieve mastery of the subject (p. 156).

Bloom (1976) suggests three independent variables that effect learning: Cognitive Entry Behaviors, Affective Entry Characteristics, and Quality of Instruction. Cognitive Entry Behaviors are prerequisite learnings or behaviors necessary for the completion of a learning task. Affective Entry Characteristics are a "compound of interests and attitudes towards the subject" (p. 168). Quality of Instruction refers to the extent to which a teacher effectively communicates information to an individual student, engages the student in practice of the learning task, and reinforces (positively or negatively) the student via feedback. The use of formative evaluation over short units lasting one or two weeks is recommended (Bloom) as an effective means of providing the necessary reinforcement.

In brief, the Mastery Learning model suggests that the educational process should be highly individualized. When a teacher faces a class having a wide range of Cognitive Entry Behaviors, the instructional process may call for several alternative introductions to the learning task, each introduction bridging the gap between some students' entry behaviors and the prerequisites for the learning task. Various methods may be used to motivate the class and prepare the students affectively to deal with the learning task. Also, feedback can be given frequently via formative testing and individualized remediation strategies. Bloom (1981) suggests that, given such a program, a much higher percentage of students will perform at a mastery level (B+ or A) than in a conventional program of instruction.

The Personalized System of Instruction (PSI) is characterized by Keller (1968) as differing from conventional teaching procedures in the following ways:

1. The go-at-your-own-pace feature, which permits a student to move through the course at a speed commensurate with his ability and other demands upon his time.
2. The unit-perfection requirement for advance, which lets the student go ahead to new material only after demonstrating mastery of that which preceeded.
3. The use of lectures and demonstrations as vehicles of motivation, rather than sources of critical information.
4. The related stress upon the written word in teacher-student communications; and, finally:
5. The use of proctors, which permits repeated testing, immediate scoring, almost unavoidable tutoring, and a marked enhancement of the personal-social aspect of the educational process (p. 83).

The most essential feature in the PSI paradigm is the constant use of reinforcement. Keller (1968) states

> The kind of change needed in education today is not one that will be evaluated in terms of the percentage of A's in a grade distribution or of differences at the .01 level of confidence. It is one that will produce a reinforcing state of affairs for everyone involved--- a state of affairs that has heretofore been reached so rarely as to be the subject of eulogy in the world's literature, and which, unfortunately, has led to the mystique of the "great teacher" rather than a sober analysis of the critical contingencies in operation (p. 86).

Learning Styles

Keefe (1979) defines learning style as follows:

> Learning styles are characteristic cognitive, affective, and physiological behaviors that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment (p. 4).

Cognitive styles are habits representing the learner's preferred or typical mode of perceiving, thinking, problem solving, and remembering. Gregorc's (1979) investigations revealed that there is a duality in learning preference.

> People learn both through concrete experiences and through abstraction. Further, both of these modes have two subdivisions, sequential and random preference (p. 20).

Gregorc (1979) discovered that these sets of dualities merged to form four distinct cognitive styles: concrete-sequential, concrete- random, abstract-sequential, and abstract-random. Most people exhibit all four patterns or styles to some degree. It is common for individuals to prefer one or two of these styles and to use those preferred styles most of the time.

A concrete-sequential learner prefers step by step directions and will follow them. Such individuals like clearly ordered presentations and a quiet atmosphere.

A concrete-random learner is intuitive and often does not "show his work". They use trial and error approaches but dislike cut and dried procedures. Such individuals work well independently and may not respond well to teacher intervention.

An abstract-sequential learner is skillful with symbolic systems and makes use of diagrams or pictures extensively. Such individuals work well with their teachers and other authority figures and prefer logical, sequential presentations. Main ideas or principles are readily accepted and used.

Abstract-random learners are sensitive to human interactions and function well in group activities and busy environments. They like to receive information in an unstructured manner and dislike rules and guidelines. Such learners organize information through a process of reflection.

Dunn and Dunn (1978) have identified a set of environmental and social factors influencing an individual's ability to learn in a given situation. For instance, some people most readily learn in a well ordered room free of distractions. Other individuals would be made quite uncomfortable by such an environment and would consequently have difficulty learning.

The intent of the identification of an individual's learning style is to provide the educator with information useful in the individualization of instruction for that student. The extent to which a learning process recognizes and accommodates individual differences in learning styles may be seen as one measure of the quality of instruction offered by the instructional program.

<u>Feedback, Remediation, and Retesting</u>

Many questions remain unanswered, if not unaddressed, concerning the nature of feedback best suited to a given instructional task. The timing of the feedback, relative effectiveness of verbal vs. written feedback, the length of the unit being tested, and the criteria for passing each unit all have a bearing on this study.

Sturges (1978) compared the effect of immediate vs. delayed feedback on retention of learning. Various delays in feedback treatments following computer managed multiple choice tests were employed in a study of 112 students enrolled in a child psychology course. Results confirmed previous studies indicating that retention following delayed feedback is not degraded by the delay. In fact, feedback delays of 20 minutes to 24 hours had a greater positive effect on student confidence on subsequent retention tests than did immediate feedback. As a result of this study, Sturges hypothesized that with a longer delay students "engage in a more thorough semantic analysis of information presented at feedback" (1978).

In another experiment involving 98 introductory psychology students, Cooper and Greiner (1971) compared student performance in a traditionally taught lecture section with student performance in a PSI section. Feedback in the PSI section was provided in the following manner: after collecting a weekly quiz, the papers were redistributed randomly to the class, no student receiving his own paper. The quizzes were then corrected in class. This provided each student with fast feedback on the correct solutions. The quizzes were then turned in for double-checking and recording by TA's. Within two hours, these papers were available to the students for study. Tutors could then assist them in remediation and evaluation. Students failing to meet the mastery standard retested on equivalent forms of the quiz later in the week. At the conclusion of the study, it was found that the PSI students earned higher course grades and had better long-term retention than students in the conventionally taught section. Cooper and Greiner hypothesized that PSI reinforces regularly spaced practice, while conventional methods reinforce massed practice prior to exams and that this is the reason for the differences in performance.

With regard to the effectiveness of written corrective feedback on quizzes, Belanger (1976) found in a study of 51 undergraduates enrolled in a PSI psychology course that only 23% took advantage of the written feedback on their quizzes. Observations made during the study led Belanger to suggest further research into the effect of verbal interaction between instructors or TA's and students.

Semb (1974) investigated the effect of unit length on student performance in a study of 193 students enrolled in an introductory child development course. The Mastery Learning method of instruction was used. When a student completed a quiz he or she took it to the TA or proctor for immediate grading. Answers were graded either correct, partially correct or incorrect. If the student could justify or complete a partially correct answer the item was marked correct. Otherwise, it was scored incorrect. The effect of various levels of mastery on final exam performance was also studied. The results of this study indicate that use of a high mastery criterion with short assignments produces better test performance than use of a low mastery criterion on long assignments.

Summarizing research on the PSI method, Kulik, Kulik, and Smith (1976) emphasize the findings of several studies examining the reasons for the reported effectiveness of the PSI method. There appear to be 3 key features: small units of instruction, effective feedback, and a unit-mastery requirement. Other aspects of the PSI approach seem to offer less benefit than has been previously assumed. In citing research by Calhoun (1976) the authors reported that immediate verbal feedback from proctors or TA's was superior to written feedback.

In the Mastery Learning paradigm, the follow-up to feedback is remediation. Hassett and McCoy (1979) investigated the effect of post-quiz prescribed remediation on student peformance in an introductory college algebra course involving 100 students. In this study, the students in the experimental group received remediation keyed to their errors on quizzes. The controls were told to go over their mistakes, but received no additional drill, as did the experimental group students. The results of this study showed a significantly smaller variance in final exam scores for the experimental group than for the control group. Performance was higher also for the experimental group.

In a study of 235 calculus students, Struik and Flexer (1977) compared the traditional lecture method with a self-paced mastery approach. In the mastery approach students had the option of retesting up to four times, with only the best score counting. Feedback occurred within a day. The results of this study showed that students in the self-paced, mastery learning course did substantially better than those in the traditional course.

Affective Aspects of Testing and Evaluation Procedures

It is not uncommon for students to experience anxiety in anticipation of and during an exam. How does this anxiety affect the way students feel about a given course or mathematics in general? Douthitt (1978) reported in a study of 47 students enrolled in a course in finite mathematics that frequent use of informal evaluation procedures (homework, boardwork, etc.) instead of traditional unit examinations resulted in achievement levels on the final exam which were not significantly different from a control group taking the conventional unit exams. However, mean scores on the Aiken-Dreger Mathematics Attitude Test were higher for the experimental group. In his conclusion, Douthitt recommended that instructors should give more attention to finding other less anxiety producing methods of evaluation.

In the previously cited study by Cooper and Greiner (1971) an analysis was also made of student anxiety levels. The PSI-taught students experienced essentially the same levels of anxiety over quizzes as did the conventionally taught students overunit exams. However, at the end of the course the PSI students rated the value of their course significantly higher than did the students in the conventionally taught sections.

A comparison of student attitudes was conducted by Haver (1978). In the study, 1200 students enrolled in a college intermediate algebra course were divided into various groups, each of which received a different instructional format. Of the various formats tested, the Mastery approach produced the highest grades, highest completion rate, and the highest student opinion rating. The Mastery Learning format was also significantly more successful in improving initially negative attitudes towards mathematics during the course than was any other approach in the experiment. This reshaping of affective behavior is consistent with Bloom's theoretical framework for Mastery Learning.

Interaction Effects

The application of the same treatment, x, to subjects in an experiment often produces varying results or degrees of variability in the dependent variable, y. If this variability is attributable to some other variable, z, active in the experiment, then it is said that there is an interaction effect between variables x and z on variable y. Such is the case in a study by Born, Gledhill, and Davis (1972). Sixty students in a Psychology of Learning class were divided into two groups, one taught by PSI methods and the other by a conventional approach. At the end of the course student performance on the final exam was higher on all test item types from PSI-taught students. This effect was most pronounced for the students with "poor" to "good" academic records. The greatest difference in scores occurred on essay and fill-in item types rather than on recognition type questions. Thus, the level of student preparation or ability interacted with the type of instruction to effect the level of student performance.

Similar results are reported by Pascarella (1977a) in a study of 248 students enrolled in calculus. The results indicated that students with the lowest levels of ability and preparation were most likely to perform better in the PSI sections than in the conventionally taught sections. Students of high ability and preparation did not perform at significantly different levels when taught by different methods. In addition, it was found that level of preparation is a more significant factor in predicting student performance when the student is enrolled in a conventionally-taught section than when in a PSI section.

Pascarella (1977b) also compared the effect of motivation on student performance in calculus in a PSI section with performance in a conventionally taught course. This study involved 94 students. It was found that the most dramatic differences in both achievement and attitude were indicated at the highest levels of motivation. This seems to indicate that the greatest benefit of PSI in mathematics instruction may go to the most highly motivated students.

A study by Peluso and Baranchik (1977) involving 395 calculus students reported similar findings. Students enrolled in a Learning Center (PSI) course did better than students taught by the traditional approach. Based on final exam performance and drop rates of the "top" students as measured by a calculus readiness test, it was also suggested that the PSI method may not be appropriate for the best students.

The study by Kulik, Kulik, and Smith (1976) places some restraints on conclusions concerning aptitude-teaching method interactions. The study points out the large differences between Bloom's theoretical relationships concerning aptitude and performance under different teaching methods and cites research in which the interaction is the opposite of what Bloom's Mastery Learning theory would predict. Clearly, more work needs to be done in the area of interaction effects in Mastery Learning or PSI environments.

"No-Difference" Studies

As an indication of the mixed results reported regarding PSI, Klopfenstein (1977) reported that in a study involving 57 students in a PSI calculus section and an unspecified number of students in a conventionally taught section, that no difference was observed either in student performance or attitude. In this version of PSI, no lectures were given. All time was spent in the PSI section working with the study guides and proctors.

A study by Harris and Liquori (1974) comparing the PSI and conventional approaches in the teaching of a course in mathematics for business students found no significant differences in student learning between groups. The study involved 128 students.

Organizational and Managerial Practices in PSI and Learning Center Environments

The proctor's or tutors's role was discussed by Romizowski, Bajpai, and Lewis (1976). In particular,self-marking of quizzes by students followed by an analysis of student errors by the proctor and remediation based on that analysis was preferred by students over grading procedures in which the student's paper was graded by a tutor. The reasons given by the students for this preference were three. Grading was smoother and less time was lost waiting for the tutor to correct a paper. Once graded, students could receive immediate feedback and remediation, rather than wait while the tutor graded someone's paper. Quizzes could be graded and somefeedback given regardless of whether or not a tutor was available. The main disadvantage seen by students was a worry that in grading their own papers a student might pass himself on a problem that a tutor would not accept and that this would lead to trouble on the summative or final exams where students do not grade their own papers. The need for detailed marking keys was discussed as a partial remedy for inexperienced student graders.

Summary

In summarizing the literature, the following points stand out: 1) in the Mastery Learning and PSI approaches, the most significant features are the use of short units, frequentquizzes followed by undelayed feedback and remediation, and the use of a high performance criterion, 2) in comparisons of the PSI and conventional approaches, student attitudes were better in PSI courses, 3) significant interaction effects between motivation and instructional method on performance and between aptitude and instructional method on performance were noted, and 4) mixed results are reported in studies dealing with various subject areas at the college level. In light of these findings, the need for research in the area of individualized instructional methods involving Mastery Learning or PSI in college mathematics is apparent.

CHAPTER THREE

PROCEDURES

Introduction

The specific assumptions and procedures for this study are discussed under the following headings: population description and sampling procedures, treatments, content validity and departmental examinations, examination reliability, statistical hypotheses, statistical procedures, precautions taken for accuracy, and summary.

Population Description and Sampling Procedure

The population under study was Montana State University students enrolled in Math 121, introductory calculus, during Fall Quarter, 1982. Nearly 700 students enrolled in the course, requiring 21 sections of Math 121. The inferred population was students at the college level enrolled in introductory calculus courses using the Swokowski (1979) text and a large group instructional format.

Based on a sample of 100 Math 121 students' final percentage scores in 1980, the researcher estimated the 1980 Math 121 population performance standard deviation to be 20.45% and the mean to be 70.4% . Using the 1980 standard deviation as an estimate of the 1982 value, the following formula (Snedecor and Cochran, 1980, p103) yielded a recommended sample size for each group, control and experimental:

$$n = (Z_{2\alpha} + Z_{\beta})^2 \sigma^2_D / \delta^2$$

where n = sample size

α = significance level

β = probability of type 2 error

σ_D = population standard deviation

δ = true difference to be detected by experiment

Using tables provided by Snedecor and Cochran (1980, p104) to simplify calculations, if the power of a one-tailed t-test is set at .95, the significance level at .05, and the percentage difference to be detected between the two groups' exam scores at 6%, then the recommended sample size per group was 125. Thus, if the true difference between the exam scores of the two groups was as large as 6%, there was a 95% probability that the difference would be detected statistically.

Initial class enrollments for Math 121 sections were set between 33 and 38 students. On the basis of the researcher's estimates of the sample size needed to meet statistical requirements in the interpretation of the experimental data, the researcher decided to seek the participation of a minimum of six sections of Math 121 in this study. The selection of the six sections took place in a meeting of all Math 121 instructors shortly after the start of the Fall Quarter. The researcher explained the purpose of the research and experimental procedures to the group and solicited volunteers to participate in the study. From those instructors offering to participate, four were selected by the researcher to take part in the study. Between them, these instructors taught 6 sections of calculus. For purposes of identification, these four individuals will henceforth be referred to as instructors M1, M2, F1, and F2.

Instructor M1 wasan experienced, male, college mathematics instructor with a PhD and an interest in research on instructional questions. Instructor M2 was a male graduate student working on a Masters in mathematics with a teaching assistantship. M2 participated in the study but did not manifest any personal interest in the questions addressed by the study. Instructor F1 was a female graduate student working on a Masters degree in mathematics with a teaching assistantship. F1 was a foreign student, having been born, raised and educated through college in Germany. Instructor F2 was a female graduate student working on a Masters degree in mathematics with a teaching assistantship and had a genuine professional interest in the study.

After the last day to add a class passed in the second week of the quarter, each instructor randomly assigned each of his students to one of two groups of roughly equal size, a control group and experimental group. Instructors M2 and F1 each taught two sections of Math 121. For both instructors, these two sections met at consecutive hours of the day. Each of these two instructors randomly assigned one entire section to the control group and one entire section to the experimental group. Instructors M1 and F2 each taught one section of Math 121. Each of these two instructors assigned half of his/her section to the control group and the other half to the experimental group using a random process.

The selection procedures outlined above were used to control for the effect of teacher differences on group performance by having each teacher make equivalent contributions instructionally to each group. In this way, neither group received the benefits or liabilities of a given teacher's instruction differently than the other group. Thus, teachers, students, and treatments were randomly assigned.

Within both the experimental and control groups, three subgroups were identified by partitioning each group on the basis of general academic readiness. The university generates an expected GPA for all incoming freshmen. This study used that statistic to rank all the students in each group. Students in the top third of the list in each group were designated as being of high ability for the purpose of this study. Students in the lowest third of the list were designated as being of lowest ability. The middle third were designated as being of average ability. This partitioning of the groups lead to roughly equal subgroup sizes, a useful design element for statistical purposes. However, it did not partition each Math 121 section equally. Also, because of transfer students and returning students that were enrolled in the sections involved in the study, no expected GPA was available for some individuals. The data for these individuals was analyzed in a fourth category for students of unknown