8-1984

An investigation of the effects of tutoring behaviors and organizational structure on student performance in an individualized remedial algebra course at the college level

Betsy Darken Smith

To the Graduate Council:

I am submitting herewith a dissertation written by Betsy Darken Smith entitled "An investigation of the effects of tutoring behaviors and organizational structure on student performance in an individualized remedial algebra course at the college level." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

Henry Frandsen, Major Professor

We have read this dissertation and recommend its acceptance:

Lawrence Barker, Jan Handler, Jerry Bellon, Clint Allison

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Betsy Darken Smith entitled "An Investigation of the Effects of Tutoring Behaviors and Organizational Structure on Student Performance in an Individualized Remedial Algebra Course at the College Level." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

Henry Frandsen

Henry Frandsen, Major Professor

We have read this dissertation and recommend its acceptance:

Clinton Allison

Janet Handler

Jerry J Bellon

Lawrence Barber

Accepted for the Council:

Cw Minkel

The Graduate School

AN INVESTIGATION OF THE EFFECTS OF TUTORING BEHAVIORS AND ORGANIZATIONAL STRUCTURE ON STUDENT PERFORMANCE IN AN INDIVIDUALIZED REMEDIAL ALGEBRA COURSE AT THE COLLEGE LEVEL

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Betsy Darken Smith

August 1984

I dedicate this thesis to my long-suffering husband Larry, who has made great sacrifices so that I could accomplish my goal.

ACKNOWLEDGMENTS

I wish to extend my appreciation to each member of my doctoral committee, to Dr. Clint Allison for opening my eyes to the vagaries of educational history, to Dr. Lawrence Barker for the tremendous amount of time he devoted to teaching me mathematical statistics, to Dr. Jan. Handler and Dr. Jerry Bellon for giving me insight, and most of all for giving me moral support in this project.
Most especially, I thank Dr. Henry Frandsen, for whose patience, faithfulness, encouragement and insights I will be forever grateful.

ABSTRACT

This study investigated the effects of two variables, organizational structure and tutoring behaviors, on performance in an individualized algebra course. Three treatments were used:

1. Treatment I: large lectures with tutor-supervised workshops; tutors engaged in limited duties, mainly answering questions.
2. Treatment II: same lecture/workshop structure as Treatment I; tutors closely monitored and encouraged student progress, establishing supportive relationships with their students. By midsemester most students in Treatments I and II attended workshop instead of lecture because of slow progress.
3. Treatment III: class of 30 students supervised mainly by an instructor; same tutoring behaviors as Treatment II.

In Treatment I grades were based on course progress and attendance; in Treatments II and III grades were also based on points awarded for meeting test deadlines.

The experiment was conducted at one period with 186 students, and was replicated at a second period with 160 students. Performance was measured by the number of units completed by the end of the semester.

ANCOVAs were conducted using three covariates: arithmetic and algebra achievement, and attitude toward mathematics. Chi-square tests were conducted on success rates, partial completion rates and attrition rates. Results were:

1. Treatment II was superior to Treatment I;

2. Treatment II was not significantly different from Treatment III;
3. Treatment III was superior to Treatment I at one class period.

It was thus concluded that under the conditions of this study, differences in organizational structure (lecture/workshop versus regular class) did not have a significant effect on performance, while differences in tutoring behaviors and/or grading systems did. Tutors were found to be just as effective as instructors in an individualized classroom, and their effectiveness was increased either by high intensity tutoring behaviors or by an incentive-based grading system.

This study was limited by the fact that none of the treatments were very successful--the highest success rate was only 40%. These low rates may have been due to a heterogeneous population, the inclination of students to extend the course into a second semester, and the course emphasis on word problems.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	Statement of the Problem	1
	Purpose of the Study	6
	Importance of the Study	8
	Design of the Experiment	8
	Analyses	10
	Sources of Data	11
	Scope and Limitations	12
	Assumptions	13
	Definitions	14
	Organization of the Study	16
II.	REVIEW OF THE LITERATURE	17
	Preface	17
	Defining and Measuring Effectiveness	18
	General Reviews of Individualized Instruction	21
	Reviews of Studies on Individualized Programs in Mathematics	24
	Overview of Reports on Individualized Programs in Mathematics	27
	Pacing Contingencies	29
	Mastery Learning	32
	Introduction	32
	Completion Rates	35
	Mastery Learning in Mathematics	39
	The Value of Repeating Tests	41
	Small Units of Study and Frequent Testing	42
	Study Guides and Objectives	45
	Immediate Feedback	46
	Tutors	47
	Introduction	47
	Tutors' Role in Discussing Tests	48
	Methods of Assigning Tutors	50
	Tutors in Mathematics Courses	51
	Student/Tutor Ratios	55
	Organizational Structure	56
	Introduction	56
	Categories of Organizational Structures	58
	Studies of Organizational Structure	59
	Indirect Information Concerning Organizational Structure	61
	Programs without Scheduled Class Meetings	62
	Programs with Class Meetings Supervised by Tutors . . .	70

CHAPTER PAGE

II. (Continued)

Programs with Class Meetings Supervised by Faculty Members 71
Summary 76
Conclusion 77

III. THE INDIVIDUALIZED REMEDIAL MATHEMATICS PROGRAM AT THE UNIVERSITY OF TENNESSEE AT CHATTANOOGA 80

IV. METHODOLOGY AND DESIGN 84

Introduction 84
General Description of Mathematics 107 84
The Sample 86
The Treatments 87
General Description 87
Comparison of Treatments I and II 88
Description of Treatment III 94
Design of the Experiment 94
Analysis 95

V. PRESENTATION AND ANALYSIS OF THE DATA 99

Introduction 99
Data from the First Experiment 99
The Sample 99
Course Progress 100
Success Rates 126
Partial Completion Rates 130
Attrition Rates 133
Data from the Second Experiment 136
The Sample 136
Course Progress 136
Success Rates 151
Partial Completion Rates 151
Attrition Rates 155

VI. CONCLUSIONS AND RECOMMENDATIONS 156

Introduction 156
Summary of Findings 158
Findings from the First Experiment 158
Findings from the Second Experiment 160
Discussion 162
Preface 162
Implications about Organizational Structure 164
Implications about the Role of Tutors and Grading Systems 166

CHAPTER PAGE

VI. (Continued)

The Possible Effect of Grading Systems on Performance 166
The Possible Effect of High Intensity Tutor Behaviors on Performance 167
Suggestions for Teaching 169
Recommendations for Further Research 170

LIST OF REFERENCES 171

APPENDIXES 181

APPENDIX A. SYLLABUS FOR MATHEMATICS 107 182
APPENDIX B. STATISTICAL DATA 184

VITA 189

LIST OF TABLES

TABLE		PAGE
IV-1.	Comparison of Tutoring Behaviors in Treatment I and Treatment II	89
IV-2.	Grading System for Treatment I	91
IV-3.	Grading System for Treatments II and III	92
IV-4.	Point System for Treatments II and III	93
IV-5.	Information on Treatments, Instructors and Tutors by Section	96
V-1.	Means and Standard Errors for the Current Unit: First Experiment	102
V-2.	Number of Students in Each Unit for Each Instructor x Treatment Combination: First Experiment	103
V-3.	Percent of Students in Selected Unit Intervals for Each Instructor x Treatment Combination: First Experiment	104
V-4.	Analysis of Variance, Including Tests of Pre-Planned Comparisons: First Experiment	105
V-5.	Correlation Coefficients for the Covariates and the Current Unit: First Experiment	107
V-6.	Means, Standard Errors, Minimums and Maximums for the Three Covariates, by Treatment and by Instructor/Treatment Combination: First Experiment	108
V-7.	Subsample Sizes, Means and the Distribution of Missing Values for the Reduced ANCOVA Sample: First Experiment	110
V-8.	Overall Results of the Full Linear Model ANCOVA: First Experiment	111
V-9.	ANCOVA with Quadratic Terms: First Experiment	112

TABLE		PAGE
V-10.	Least-Squares Means, Full ANCOVA Model: First Experiment	113
V-11.	Test Results for Pre-Planned Comparisons Using the Full ANCOVA Model: First Experiment	114
V-12.	Estimates of the Differences between Current Units for Treatments I and II at Extremes of the Region of Interest, Full ANCOVA Model: First Experiment	115
V-13.	Estimates of the Differences between Current Units for Treatments I and III at the Extremes of the Region of Interest, Full ANCOVA Model: First Experiment	116
V-14.	Estimates of the Differences between Current Units for Treatments II and III at Critical Points of the Region of Interest, Full ANCOVA Model: First Experiment	117
V-15.	Overall Results of the Analysis of Covariance Using Only Arithmetic and Algebra Covariates: First Experiment	119
V-16.	Least-Squares Means for the Analysis of Covariance with Arithmetic and Algebra Covariates: First Experiment	120
V-17.	Test Results for Pre-Planned Comparisons Using the ANCOVA Model with Arithmetic and Algebra Covariates: First Experiment	122
V-18.	Estimates of the Differences between Current Units for Treatments I and II of the Extremes of the Region of Interest, Using the ANCOVA Model with Arithmetic and Algebra Covariates: First Experiment	122
V-19.	Estimates of the Differences between Current Units for Treatments I and III at the Extremes of the Region of Interest, ANCOVA Model with Arithmetic and Algebra Covariates: First Experiment	123
V-20.	Estimates of the Differences between Current Units for Treatments II and III at Selected Critical Values of the Covariates, ANCOVA Model with Arithmetic and Algebra Covariates: First Experiment	123

TABLE PAGE

V-21. Overall Results of the ANCOVA Using the Pooled Treatment II/III: First Experiment 125

V-22. Estimates of the Differences between Current Units for Treatments I and II/III at the Extremes of the Region of Interest: First Experiment 127

V-23. Success Rates: Percent and Number of Students in Units 14-16, First Experiment 129

V-24. Success Rates for Reduced ANCOVA Sample: Percent and Number of Students in Units 14-16, First Experiment . . . 129

V-25. Success Rates for the Full Sample Using the Pooled Treatment II/III: Percent and Number of Students in Units 14-16, First Experiment 130

V-26. Partial Completion Rates: Percent and Number of Students in Units 10-16, First Experiment 131

V-27. Partial Completion Rates for Reduced ANCOVA Sample: Percent and Number of Students in Units 10-16, First Experiment 132

V-28. Partial Completion Rates Using the Pooled Treatment II/III: Percent and Number of Students in Units 10-16, First Experiment 133

V-29. Attrition Rates: Percent and Number of Students Taking No Tests After the Withdrawal Deadline, First Experiment 135

V-30. Attrition Rates for the Reduced ANCOVA Sample: Percent and Number of Students Taking No Tests After the Withdrawal Deadline, First Experiment 135

V-31. Attrition Rates Using the Pooled Treatment II/III: Percent and Number of Students Taking No Tests After the Withdrawal Deadline, First Experiment 137

V-32. Means and Standard Errors for the Variable of Current Unit: Second Experiment 139

V-33. Number of Students in Each Instructor x Treatment Combination by Unit: Second Experiment 140

TABLE PAGE

V-34. Percent of Students in Each Instructor x Treatment Combination, by Selected Unit Intervals: Second Experiment 141

V-35. Analysis of Variance: Second Experiment 141

V-36. Correlation Coefficients for the Covariates and the Current Unit: Second Experiment 143

V-37. Means, Standard Errors, Minimums and Maximums for the Three Covariates by Treatment and Instructor x Treatment Combinations: Second Experiment 144

V-38. ANCOVA Results, Full Linear Model: Second Experiment . . 145

V-39. Results of the ANCOVA with Quadratic Terms: Second Experiment 146

V-40. Least-Squares Means for the Analysis of Covariance: Second Experiment 147

V-41. Results of Tests for Pre-Planned Comparisons Using the Full Analysis of Covariance Model: Second Experiment 147

V-42. Estimates of the Differences between Current Units for Treatments I and II at Values within the Region of Interest: Full ANCOVA, Second Experiment 149

V-43. Estimates of the Differences between Current Units for Treatments I and III at the Extremes of the Region of Interest: Full ANCOVA, Second Experiment 150

V-44. Estimates of the Differences between Current Units for Treatments II and III at Critical Values for the Region of Interest: Second Experiment 152

V-45. Success Rates: Percent and Number of Students in Units 14-16, Second Experiment 153

V-46. Partial Completion Rates: Number and Percent of Students in Units 10-16, Second Experiment 154

V-47. Attrition Rates: Percent and Number of Students Taking No Tests After Withdrawal Deadline, Second Experiment . . 155

A-1. Topics in Mathematics 107, by Unit and Week of Presentation 183

CHAPTER I

INTRODUCTION

By an interesting coincidence, two major developments in college level education have unfolded almost simultaneously over the past two decades. The first was the significant decline in the general level of academic preparedness of entering college students. The second was the extensive implementation and study of several new teaching methods, all classified under the rubric of individualized instruction. The continuing difficulties encountered in attempts to establish effective remedial programs in mathematics, combined with evidence to support the possible superiority of certain types of individualized programs, suggest that further research needs to be conducted to determine the key features of a successful individualized program in remedial mathematics.

1. STATEMENT OF THE PROBLEM

The decline in the academic preparedness of entering college students has resulted in a dramatic increase in remedial courses in mathematics. A recent survey of the Conference Board of the Mathematical Sciences (Fey, Albers, & Fleming, 1982) reported that 25% of the mathematics students in four-year colleges were enrolled in remedial mathematics courses; in two-year colleges this figure jumped to 42%. While the propriety of such programs still provokes debate, the foremost issue among those involved in remediation concerns the question of effectiveness. Namely, how many of the students enrolled in remedial programs are being successfully prepared for college level courses? While it is difficult to gather accurate data on the subject, there is evidence to indicate that many remedial mathematics programs suffer from 50% to 80% dropout rates, while others with apparently lower dropout rates are discovered to produce graduates most of whom drop out of their next math courses. How can remedial programs be made more effective? The focus of this study is upon this question.

Individualized instruction, the second major development mentioned above, is also central to this study. Usually, innovative teaching methods do not deserve to be referred to as major developments. One of the characteristics of the history of education in this century, and particularly of the last several decades, has been the frequent appearance of innovations heralded as capable of revolutionizing the nature of teaching and/or learning. Unfortunately another historical characteristic has been the eventual--or sometimes abrupt--disappearance of most of these innovations, trailed by strings of studies reporting no significant differences. Dubin and Taveggia summed up the situation in 1968 after reanalyzing 50 years of data from 191 comparative studies. They concluded that the data "demonstrated clearly and unequivocably that there is no measurable difference among truly distinctive methods of college instruction when evaluated by student performance on final examinations" (Dubin & Taveggia, 1968, p. 35).

Oddly enough, 1968 was also the year that Keller published his seminal article on the Personalized System of Instruction (PSI) (Keller, 1968), which was to become the most widely used of the systems of individualized instruction. The impact of individualized instruction was illustrated by the following statement, typical of many reports made in the mid to late seventies:

> In the last ten years, educational research has established that Keller's Personalized System of Instruction is effective in promoting student achievement. Its educational record stands in stark contrast to that of earlier alternatives to the lecture method of teaching, for its use has consistently improved student performance on final examinations in college courses (Kulik & Kulik, 1979, p. 84).

Taveggia himself, who had given the pessimistic report on innovations in 1968, also joined the chorus praising personalized instruction (Taveggia, 1976).

Not only had research in individualized instruction generated evidence of its superiority to traditional instruction in certain situations, it had also gone beyond simple comparisons to focus on component analysis. The goal of the latter was to isolate key components of an effective individualized program. Significant advances have been made in identifying certain features and modifications which contribute most to the success of individualization in a number of academic settings.

Considering the problems encountered in remedial mathematics programs and the glowing reports of the success of individualized instruction, it was reasonable to attempt to apply the latter to the former. Unfortunately, this may be a classic example of the irresistible force meeting the immovable object. The results of experiments with individualized instruction in developmental mathematics are far more mixed than the results reviewed by Kulik and Kulik, Taveggia and others. It seems very likely that the difficulties encountered in developmental mathematics are in large part attributable to the special problems of the subject matter and the remedial students. That special problems exist is attested to by the aforementioned astronomically high dropout rates suffered by many remedial mathematics programs, rates much higher than most of those reported in other academic areas. Such difficulties certainly stem in part from a student population handicapped by such problems as severe academic deficiencies, motivational difficulties, poor study habits and "math anxiety." It is hardly surprising that such characteristics would complicate the problem of establishing a successful individualized remedial mathematics program. It has become quite clear that this problem is not solved by simply transferring wholesale an individualized program that has been successful in another discipline. Programs in remedial mathematics seem to require special features that are of minor or no importance to other programs. Commenting on this very matter, Kulik, Jaksa and Kulik (1978) recommended that research be done to "investigate the possibility that certain PSI features (e.g., tutoring or self-pacing) may be important to certain kinds of students, subject matter, and levels of instruction, but not for others" (p. 12).

Unfortunately, very little research has been conducted to isolate the key characteristics of successful programs in remedial mathematics. Most of the research on individualized mathematics have been comparisons of individualized with more traditional programs. The results of such research are very mixed and it is difficult to discern from the information supplied by the researchers why some individualized programs were successful and others were not. There is some indication that, in common with individualized programs in other fields, individualized mathematics programs benefit from frequent testing, the use of study objectives (commonly in the form of practice tests), and mastery learning used in conjunction with pacing contingencies. However, it is also clear from research and from anecdotal reports that these characteristics are <u>not</u> in and of themselves sufficient for a successful program in many settings. In fact the difficulties entailed in attempting to define the characteristics of a successful individualized remedial program suggest that more subtle factors are involved. The problem is to identify these factors.

One possibility that has received very little direct attention in the literature concerns the organizational structure of the program. Some programs for instance are organized around large lecture presentations, others rely solely on optional math lab attendance, and yet others have regularly scheduled small class meetings conducted by instructors. There is some evidence in the literature to suggest that the more successful programs are more likely to have the lattermost structure. However, there has been no research conducted to investigate a cause and effect relationship between these two characteristics. Likewise, there is some evidence in the literature to suggest that the use of tutors in remedial individualized programs may be beneficial, but that their usefulness seems to depend strongly on the manner in which they are incorporated into the program. Again, little if any research has been conducted to investigate the effect of tutoring behaviors on student performance. In short, very little research has been conducted to isolate the key characteristics of a successful individualized remedial mathematics program, and in particular not enough research has been conducted to affirm or deny the importance of organizational structure or of tutoring behaviors in such a program.

2. PURPOSE OF THE STUDY

The purpose of this study was to determine whether certain organizational structures and certain tutoring behaviors had an impact on student performance in an individualized remedial mathematics program.

Specifically, this study compared the following two types of structures:

1. a lecture/workshop structure in which students alternately attended large lectures and tutor-supervised workshops, with the option of attending the workshop all four days of the week;
2. a more traditional class structure in which 25 to 30 students attended an individualized class supervised by an instructor and a tutor assistant.

In addition, this study compared two types of tutoring situations:

1. a more intensive situation in which students were assigned to specific tutors who closely monitored their attendance records and their test-taking activities; this monitoring was reinforced by a point system of grading based on class attendance and adherence to a test schedule;
2. a less intensive situation in which tutors' duties included little beyond answering students' questions.

There was a common theme addressed in both of these investigations, namely the value of increased personal contact between members of the instructional staff and their students.

The main purpose of this study was to compare the relative effectiveness of these different approaches on student performance in a remedial course in elementary algebra. By manipulating characteristics within the individualized program, this study supplies direct evidence concerning their usefulness in improving the program.

3. IMPORTANCE OF THE STUDY

This study contributes to the body of research aimed at determining the important components of individualized instruction by investigating variables about which little is known. It is particularly informative because it directly manipulates these variables within the context of an individualized mathematics program. These variables of organizational structure and tutoring behaviors are not only of theoretical interest but also of practical interest, as they have a direct bearing on how a program can more effectively use its resources. This is particularly important for remedial mathematics programs, which are often simultaneously plagued by problems of effectiveness and scarce resources, particularly faculty and monetary resources.

4. DESIGN OF THE EXPERIMENT

In order to study the variables of organizational structure and tutoring behaviors described above, three treatments were implemented:

Treatment I: classes organized around a lecture-workshop structure with workshop tutors engaging in low intensity relationships with their students.

Treatment II: classes organized around a lecture-workshop structure with workshop tutors engaging in higher intensity relationships with their students and aided by a point system of grading.

Treatment III: classes of approximately 30 students meeting regularly with an instructor, with the instructor and her tutor assistant engaging in the same higher intensity student relationships as the tutors of Treatment II and using the same point system for grading.

Two instructors were involved in the experiment, each giving lectures to students in Treatments I and II on two days of the week and conducting a Treatment III class on the other three days of the week. Thus the experiment had a 2 x 3 factorial design. Approximately 200 students were randomly assigned to the cells of this design.

In addition, tutors were randomly assigned to the three treatments. The organization of the classes was such that each instructor shared the same tutors for Treatments I and II. The ratio of students to tutors in these two treatments was 15 to 1. The same ratio of students to instructional staff was also present in Treatment III, where 30 students were assigned to one instructor and one tutor, both of whom were present for three of the four class meetings. Only the tutor was present for the fourth class meeting, during which students were encouraged to take tests.

This experiment was replicated at the noon period.

5. ANALYSES

The main measure of performance used in this study was the student's current unit at the end of the Fall Semester, hereafter referred to as the "current unit." To determine if there were significant differences among the treatments for this dependent variable, analyses of variance and analyses of covariance were conducted. For the analyses of covariance (ANCOVAs), the covariates included:

1. pre-semester arithmetic achievement;
2. pre-semester algebra achievement;
3. pre-semester attitude toward mathematics.

For both the analyses of variance and the analyses of covariance, tests based on the variable of current unit were conducted for the five following pre-planned comparisons:

1. Is Treatment I different from Treatment II?
2. Is Treatment II different from Treatment III?
3. Is Treatment I different from Treatment III?
4. Is Treatment I different from the average of Treatments II and III?
5. Is the average of Treatments I and II different from Treatment III?

A number of other measures of student performance were also analyzed, including:

1. success rates, defined in terms of the number of students who reached Unit 14 (the last unit) by the end of the semester;

2. partial completion rates, defined in terms of the number of students who reached Unit 10 by the end of the semester:

   and

3. attrition rates defined in terms of the number of students who stopped taking tests after the official withdrawal deadline (in the ninth week of the semester).

Unless otherwise noted, the denominator for these rates is the number of students originally included in the experiment.

6. SOURCES OF DATA

Variables in this study were measured as follows:

1. Arithmetic achievement was measured by the Arithmetic Skills Test of the Descriptive Tests of Mathematical Skills of the College Board, published by the Educational Testing Service. This test was the first part of a mandatory placement test which most students took prior to the first day of classes. The remaining students took the exam within a week of the beginning of classes.
2. Algebra achievement was measured by Mathematics Test Form BA/1B, a basic algebra test published by the Mathematics Association of America. This was the second part of the mandatory placement test.

3. Attitude toward mathematics was measured by the Aiken and Greger Revised Mathematics Attitude Scale. This instrument was administered to most students on the first day of class. Students who did not attend the first day of class were administered the instrument when they appeared in the math lab to take their first lab tests. A high score indicated a positive attitude toward mathematics.
4. The main measure of performance, the current unit at the the end of the semester, was determined by the number of tests passed. The unit tests were modifications of tests provided by the authors of the textbook and were specifically geared to test all of the unit objectives. There was a minimum of four forms of each unit test and the final examination. The final examination was comprehensive. The mastery criterion for all tests was 85%.

Data collection was facilitated by a computerized record-keeping system used in the remedial program.

7. SCOPE AND LIMITATIONS

It is expected that the results of this experiment may be applicable to similar remedial algebra programs with similar populations. This experiment was conducted at The University of Tennessee at Chattanooga, one of the campuses of The University of Tennessee System. Its enrollment is approximately 7,600 students, approximately 90% of whom are commuters. The minimal admittance requirements are a 2.00 high school grade point average or an ACT composite of 18 or above. These requirements are not far removed from an open door policy, as reflected by the fact that approximately half of all students tested in 1983 placed at the remedial level in mathematics.

Mathematics 107, the remedial elementary algebra course which is the setting of this experiment, is the only remedial mathematics course offered at The University of Tennessee at Chattanooga. Thus the range of backgrounds of its students is quite broad, varying from those with no algebra at all to a few with four years of college preparatory mathematics. Because of the diversity of this student population, the course was structured so that students had the option of finishing the course in two semesters. At institutions with more than one remedial course, course populations may be more homogeneous and thus more manageable. Even so, it seems probable that the results of this experiment will have some bearing on remedial algebra programs at institutions similar to The University of Tennessee at Chattanooga.

8. ASSUMPTIONS

For the statistical models it was assumed that: (a) the random assignment of students to Instructor by Treatment combinations guaranteed the independence of the error terms, (b) the relatively large cell sizes compensated for any departures from normality, and (c) the error variances were homogeneous. It was also assumed that missing values due to unavailable arithmetic and algebra covariate scores did not introduce serious bias into the experiment.

9. DEFINITIONS

The success rate is the ratio of the number of students who reached Unit 14 by the end of the Fall Semester 1983 to the initial number of students included in the experiment at the beginning of the semester.

The partial completion rate is the ratio of the number of students who reached Unit 10 by the end of the semester to the initial number of students included in the experiment at the beginning of the semester.

"W" refers to the withdrawal grade.

The attrition rate is the ratio of the number of students who failed to take any tests in Mathematics 107 after the withdrawal deadline (in the ninth week of the course) to the initial number of students included in the experiment at the beginning of the semester. Note that this includes students who may or may not have officially withdrawn from the course, as well as students included in the experiment who dropped within the first few weeks of the course and hence did not receive even a grade of W.

The FW rate is the ratio of the number of students receiving grades of F or W to the total number of students receiving a grade in the course. A similar definition applies to the DFW rate.

The current unit refers to the course unit that a student reached by the end of Fall Semester 1983.

PSI, the Personalized System of Instruction, is a particular type of individualized instruction which incorporates mastery learning, small units of instruction and frequent testing, heavy reliance on the written word and de-emphasis of lecturing, use of study guides and study objectives, and the use of tutors or proctors.

A tutor in this study is an undergraduate student paid by the hour to work in the remedial mathematics program. Tutors are hired for their apparent competency in mathematics and their ability to empathize with their students.

Mathematics 107 is the remedial algebra course at The University of Tennessee at Chattanooga. Its content is essentially equivalent to first year high school algebra.

A unit is a subdivision of the Mathematics 107 textbook corresponding to approximately one week's work in a one semester course.

10. ORGANIZATION OF THE STUDY

The remaining chapters of this study are organized in the following way:

Chapter II contains a review of the literature relevant to the present study.

Chapter III provides a history of the remedial program at The University of Tennessee at Chattanooga in the last decade.

Chapter IV describes the experimental design and the procedures of the study.

Chapter V contains a presentation and analysis of the data collected in the study.

Chapter VI summarizes the experiment and presents conclusions and recommendations for further study.

CHAPTER II

REVIEW OF THE LITERATURE

1. PREFACE

Before the literature is reviewed on individualized instruction in general and the Personalized System of Instruction (PSI) in particular, it is important to make a few observations regarding interpretation. First, individualized instruction is a very broad term which has been used to describe many different methods of teaching, including programmed learning, audio-tutorial methods, computer-assisted instruction, learning contracts, mastery learning and PSI. The common features seem to be an emphasis on individual rather than group-based instruction and the variability in the rates at which students move through the material. However, the differences among these methods are probably greater than their similarities. Because of this situation research on so-called "individualized instruction" must be interpreted with caution.

An advantage of PSI for research as well as practical purposes is that it is better defined than many other teaching methods, although experience has shown there is room for diversity even within this carefully defined program. Its basic features, based on the behaviorist principles of contingency management and reinforcement theory, are:

1. mastery testing on small units of study;

2. self-pacing, whereby a student is permitted to move through the course at a speed commensurate with his ability and other demands on his time;
3. primary communication through the written word, with lectures used as sources of motivation, not critical information;
4. the use of proctors to permit repeated testing, immediate scoring and personal tutoring.

An analysis of the components of this system and its modifications will be discussed in the following review of the literature.

2. DEFINING AND MEASURING EFFECTIVENESS

To properly interpret the results of studies reported in the literature, it is critical to consider the question of how the effectiveness of an individualized program is to be measured and how programs are to be compared. Final course grades have long been rejected for this purpose for many reasons, among them the problem of subjectivity and diversity of grading systems. Instead, the standard source of comparison has been the mean scores of treatment groups on a common final examination. However, this procedure is usually based on a critical assumption, namely that the treatment groups at the end of the experiment are equivalent. This assumption can be completely unwarranted. This situation will arise if the methods of instruction under study have significantly different effects on dropout rates, even if students were originally assigned at random to the methods. Although there is conflicting evidence over the overall effect of PSI on dropout rates, there is more than enough evidence to indicate that there can be a large difference between PSI and conventional groups on this statistic (Akst, 1976; Born & Whelan, 1973; Hinton, 1978; Robin, 1976; Wood, 1975). Hence it is imperative that interpretation of differences in mean final examination scores be accompanied by a comparison of dropout rates. If such a comparison is not made, then one can not reject the hypothesis that group differences are caused by the systematic deletion of weaker students under one of the methods. Wood (1975) has stated emphatically that the W grades cannot be ignored when comparing PSI to traditional methods; unfortunately many researchers have failed to heed this advice.

This point must also be made about component analyses, in which variations of PSI are compared to one another. However, it may be added that a highly useful measure of effectiveness which takes withdrawals into account is the average number of units completed by all students originally enrolled in the course. Unfortunately, this measure has rarely been reported.

Another general question about the effectiveness of PSI must also be addressed by researchers if credibility is to be more firmly established. This concerns the performance of PSI students in succeeding non-PSI courses. Skeptics have suggested that students who have been "pampered" through a prerequisite PSI-taught course have not been prepared for their next course. The implication is that the superiority of PSI students on final examinations is illusory, based either on "teaching for the test" or on the greater practice PSI students have had with test-taking. To answer these assertions, researchers need to perform follow-up studies, especially in the more hierarchical disciplines, to obtain a second measure of effectiveness, namely how PSI students compare to their counterparts in subsequent courses. The infrequent report of this measure will be highlighted in the following review.

In developmental mathematics there is another, absolutely central reason for conducting follow-up studies. Simply, the primary purpose of a developmental mathematics program is to prepare students to successfully complete a course in college-level mathematics, i.e. their next course in mathematics. Even if a developmental course has a good completion rate and good final examination results, it can hardly be called effective if most of its students proceed to do poorly in college mathematics. A case in point is a developmental mathematics program in Ohio (Romoser, 1978) which passed most of its students in the first quarter and the remainder in the next quarter. A follow-up study revealed that only 23% of these students eventually made a C or above in their next math course (and this percent includes an unspecified number who repeated the course).

Related to this same issue is another measure of effectiveness particularly pertinent to developmental programs. This is the long-term retention rate, where retention can refer to both re-enrollment in the institution and re-enrollment in another math course. The propensity of "open-door" students to become "revolvingdoor" students should make the retention rate of great interest to researchers in developmental areas. For similar reasons, changes in attitude toward mathematics are of interest.

In light of the complications arising from definitions and measures of effectiveness, as well as from ill-defined use of the term "individualized instruction," interpretation of the literature must be made cautiously. Fortunately, some researchers have taken these points into consideration.

3. GENERAL REVIEWS OF INDIVIDUALIZED INSTRUCTION

As indicated, an apparently strong case for the superiority of PSI has been made in the literature. However, reviews of research on PSI have included few studies in mathematics and have not consistently addressed the problem of differential attrition rates. Hursh (1976) examined 23 comparative studies dated after 1963, all of which support the claim that "PSI produces higher student ratings and exam scores, and larger proportions of A and B grades, compared to conventional methods of instruction" (p. 92). However, of these 23 only 12 used a common final exam as the measure of achievement, and only five can be ascertained to have been conducted in non-psychology courses, including at least two studies in statistics and three in physics and engineering. Taveggia (1976) found 28 independent comparisons, again all favoring PSI. However, neither he nor Hursh reported that any of these studies examined dropout or attrition rates, although Hursh referred to the problem and remarked that three early PSI studies reported an average 27% incomplete rate. Since students who receive I grades presumably are often excluded from the studies, they present the same problems as those who drop the course.

In a more comprehensive study, Robin (1976) addressed the question of withdrawal rate when he examined 39 comparisons of PSI. Again, most of the studies were in the social sciences (29), with only one in mathematics and six in the sciences and engineering. A total of 30 favored PSI but only 14 reported withdrawal rates. The average PSI rate among these 14 was 14%, compared to 10% withdrawal for the lecture-discussion conditions. Robin reports that seven studies reported ratios of at least 1.5 to 1.0 in dropout rates, but that only two tested for the equivalence of dropouts and completers and then statistically controlled for any obtained discrepancies. Robin was prompted to remark that there was a "need for careful attention to basic principles of experimental control in future studies" (p. 324).

A number of meta-analyses of comparative studies have been performed recently. This statistical procedure utilizes Effect Size (ES) as the basic index of achievement; ES is defined to be the difference between the means of two groups divided by the standard deviation of the control group. Since an ES is in standard deviation units, cross-study comparisons are made possible. Kulik, Kulik and Cohen (1980) used this procedure to analyze 312 studies of educational technologies in college teaching. The average effect size for student achievement (as measured by final examination scores) was .55 in PSI studies. This is a medium-sized ES and it was strikingly higher than the ES for the other technologies, which was a low .21. However, when restricted to studies in the "hard" sciences which controlled for instructor effects, the ES dropped to approximately .35 for PSI studies. (The corresponding ES for the "soft" sciences was approximately .55). Kulik et al. (1980) also examined course completion rates. Contrary to Robin's report and other results reported in the literature, they found essentially no difference between traditional and experimental methods on this statistic. For both, completion rates were around 80%, although they did find that effect sizes varied considerably on this variable. No pattern emerged even though general type of subject matter and other characteristics were taken into account. Only 66 studies contained data on course completion and 19 reported a significant difference between methods. Nothing is said about adjustments for these differences; however, they did split evenly between the methods. As to completion rates in PSI, the authors remarked that lower completion rates in PSI classes were reported only in early studies in the literature. Reasons for this turnabout will be discussed later.

Willett and Yamashita (1983) also conducted a meta-analysis, centering their attention on comparative studies of 12 innovative instructional systems used in science teaching at the pre-college levels. While noting that the results of the studies may often have been confounded with the value of the instructional materials used, they found that the most successful systems were PSI (ES=.60) and mastery learning (ES=.64). The only other system with even a moderate effect size was the contract learning system. "Individualized instruction," a catch-all category of studies incorporating self-pacing and individual learning packets, had only a small ES of .17. Apparently none of the studies were in mathematics.

Despite the methodological problems present in many of the studies reviewed above, there is still enough evidence to indicate that PSI is a superior method of instruction <u>in certain circumstances</u>. For psychology and other "soft" sciences for instance, the evidence in favor of PSI at the college level is overwhelming. There is also support for PSI in the "hard" sciences, although not as overwhelming. For mathematics, however, and for courses with high attrition rates in PSI, there is very little evidence to be found in these general reviews. For such evidence it is necessary to turn to a handful of specialized reviews and a series of individual studies.

4. REVIEWS OF STUDIES ON INDIVIDUALIZED PROGRAMS IN MATHEMATICS

Only a few reviews have been specifically devoted to individualized instruction in mathematics. Schoen (1976) did an extensive literature search for studies of "self-paced" mathematics instruction at the secondary and post-secondary levels. He defined individualized instruction (apparently synonymous with "self-paced" instruction) as that which specified behavioral objectives, used small units of content and individual learning packets, and did not rely heavily on textbooks. It is possible that the latter specification eliminated some PSI studies. In any case Schoen found very few comparative studies of individualized instruction (as defined above) and traditional instruction (defined as teacher-centered and teacher-paced), and of those he found there was little evidence of the superiority of one method over another. Of the five studies at the post-secondary level, one favored individualized instruction, one favored traditional instruction and three found no significant differences. Results were even less supportive of individualized instruction at the secondary and elementary levels.

An examination of Schoen's bibliography revealed a hodge-podge of teaching methods. His definition of individualized instruction may have netted the same category of studies that Willett and Yamashita lumped under individualized instruction in their meta-analysis for the sciences. If these categories are in fact similar, then the two studies reached the same conclusion: when individualized instruction is defined in this manner, there is little evidence to indicate its superiority. Unfortunately, Schoen did not consider PSI studies in a separate category as the other reviewers did.

Miller (1976) also conducted a review of the literature on individualized instruction in mathematics. He included all studies which identified themselves as individualized as long as they incorporated a self-pacing feature. Of the 88 comparative studies found at all grade levels, 42 produced no significant differences, 32 favored individualized instruction and 14 favored the control group. However, only eight of these studies were at the college level; six of these reported no significant differences and two favored individualized instruction. Miller, unlike Schoen, specifically addressed the question of attrition and found only five studies which reported results on this issue. Of these five, four reported no significant differences and one favored the experimental group. However, none of these involved college level developmental mathematics.

A meta-analysis of four varieties of individualized instruction in elementary and secondary mathematics was reported in a doctoral dissertation (Hartley, 1977). Included were computer-assisted instruction, cross-age and peer tutoring, individual learning packets, and programmed instruction. Tutoring proved to be the superior technique for increasing mathematics achievement, with cross-age tutoring slightly better than peer and adult aide tutoring. Individual learning packets and programmed learning were, on the other hand, frequently inferior to traditional teaching, although on the average they were comparable to traditional instruction. Computer-assisted instruction fell between these two sets of methods. These results throw light on a very important point: examining methods of individualized instruction separately is much more informative than conducting blanket reviews, since different characteristics among the methods apparently have drastically different effects on student achievement. Because of this fact, the main conclusion to be drawn from the first two reviews is that there are many ineffective programs of individualized instruction. The case for or against PSI at the college level must be made elsewhere.

5. OVERVIEW OF REPORTS ON INDIVIDUALIZED PROGRAMS IN MATHEMATICS

Since the mid-1970's there has been a substantial increase in the number of comparative studies using college mathematics courses ranging from developmental arithmetic to first term calculus. Forty-seven such studies have been located by this researcher; half were unpublished doctoral dissertations and many suffered from methodological problems. If the latter difficulty is ignored for the moment and a simple box score is tallied, 24 report significant differences in achievement favoring individualized instruction, 22 report no significant differences and one reports a significant difference in favor of the traditional method. Such results suggest that a closer inspection of these studies may reveal some characteristics distinguishing the superior programs. Unfortunately, many of the reports, particularly the dissertation abstracts, fail to describe either the experimental or the control teaching methods in detail. Thus only a few more box scores can be derived. If the studies which refer to the use of mastery or competency learning are culled, 18 are found to favor the experimental method and 11 report no significant differences. Thirteen studies specifically describe their experimental treatment as PSI. Of these, eight favor PSI and five report no significant differences. While these results are somewhat promising, they certainly do not reflect the overwhelming support for PSI reported in the reviews of primarily non-mathematical studies cited above. It is interesting to note, however, that the results on non-mastery, non-PSI studies do tend to agree with the latter reviews: only six favor individualized instruction, while 11 report nonsignificant results and one favors traditional instruction.

It is curious that practically none of the research reports on PSI reviewed so far favor traditional instruction. There are at least several possible explanations for this phenomenon: either PSI is such a superior method that at worst it is as good as traditional instruction; or PSI, being an experimental method, benefits from the Hawthorne Effect, whereby the newness of the treatment generates the superior results, not the treatment itself; or the failures in PSI are not reported. The second explanation, while initially plausible, loses credibility in the face of evidence that other innovative programs have failed to show the same success as PSI. The third explanation is more serious. It is supported, for instance, by the fact that published reports of PSI show more positive results than unpublished dissertations. Is this due to better research techniques in the former or to the lack of selection inherent in the latter? The question is open. The fundamental question concerns the relationship between research reports in general and the experiences of practitioners "in the field." The latter periodically publish anecdotal articles, with publication again presumably favoring success stories. Such articles usually provide only limited information because of such problems as confounding of variables, lack of comparability of groups treated differently, lack of pertinent data, etc. However, it is worthy to note that, contrary to the impression gained from research reports, anecdotal articles not infrequently refer to failures of the PSI method, resulting either in retreats to more traditional methods of instruction or in major modifications to the individualized approach (Wykoff, 1980; Steele, Legg, & Miles, 1980; Archer, 1978).

Because the research results on PSI are not strongly positive, and because of evidence that practitioners may be having trouble implementing PSI in mathematics, the question mathematics educators need to consider is not whether PSI is superior but whether a particular form of PSI can be found which is consistently superior in mathematics settings. Reports which cast light on this question include both component analyses in other fields and a number of detailed and thoughtful studies in mathematics, several of which use very interesting control groups. These will be the subject of the next sections, as the different components of PSI are investigated. The goal of this investigation is to identify why some PSI programs in mathematics are successful and others are not, in the process identifying the critical features of an effective individualized mathematics program.

6. PACING CONTINGENCIES

Keller's original PSI did not provide any external mechanisms for controlling students' progress through the course. However, many early implementers of PSI found themselves confronted by the "procrastination problem," that is, they were disturbed by the significant number of students who progressed through the course extremely slowly (Keller & Sherman, 1974). This situation produced a number of undesirable results, including an unusually high number of incompletes and an overburdening of a previously underutilized staff toward the end of the course. Researchers were thus inspired to consider ways to control students' behavior, and the result has been numerous studies (exclusively in non-mathematical areas) experimenting with different pacing contingencies. From this research a consensus has been formed that restrictions on pacing need not adversely affect student performance and can control student progress (e.g., Wesp & Ford, 1982; Semb, Conyers, Spencer, & Sanchez-Sosa, 1975; Sutterer & Holloway, 1975; Riedel, Harney, LaFief, & Finch, 1976; Morris, Surber, & Bijou, 1978; Davies, Born, & Semb, 1980; Bijou, Morris, & Parsons, 1976). For example, Glick and Semb (1978) reported a 22% increase in completion rate in an introductory child development course with five as opposed to no deadlines. Riedel et al. (1976) reported that completion rates jumped from 51% to 82% when bonus points were established for making steady progress. In a review of the literature, Reiser (1976) found 46 PSI studies all of which agreed that pacing contingencies were necessary, although positive incentives were preferred over negative incentives. Other reviewers (Kulik et al., 1978; Robin, 1976) also found that positive incentives for progress reduced procrastination, lowered withdrawal rates and did not affect achievement on final examinations.

Authors of studies in mathematics are clearly in accord with this conclusion. Pacing contingencies are not only mentioned as one of the features of most individualized mathematics courses, they are also frequently referred to as being major improvements over previously "self-paced" programs (Wykoff, 1980; Johnson & Steffensen, 1977; Chatterly, 1977; Steele, Legg, & Miles, 1980; Thompson & McCoy, 1979; Taylor, 1978; Overholser, 1979). As for general practices, a subcommittee of the American Mathematics Association of Two Year Colleges (Baldwin, 1976) found that of 104 developmental mathematics programs surveyed, 51% imposed constraints on individualized courses while 23% did not. The subcommittee itself recommended that constraints be used.

The most interesting report on pacing procedures in mathematics was that of Greenwood (1977). He sorted 31 comparative studies of individualized and conventional mathematics instruction in community colleges into two categories. The "self-paced" category included four studies which were open-ended, with no deadlines at all. The second category included 27 studies with various imposed pacing contingencies, ranging from end-of-the-course deadlines (the most common) to a number of deadlines during the term. The results are revealing. None of the four "self-paced" studies favored the individualized method, three reported no significant differences and one favored the conventional method. However, of the studies using deadlines, 13 favored individualized instruction, 12 reported no significant differences and two favored conventional instruction. While tallies of this sort are fallible because of the possibility of alternate explanations from confounded variables, there is enough evidence accumulated to safely conclude that pacing contingencies in an individualized mathematics program are highly desirable.

7. MASTERY LEARNING

Introduction

The mastery learning concept is not unique to PSI. In fact, it is the basis for the philosophy of learning espoused by Benjamin Bloom and his followers. His basic tenet is that almost all students can master the material presented to them if they are given enough time and if the material is presented in a pedagogically sound fashion. In fact, Bloom (1976) also claimed that initial differences in so-called student aptitudes will fade away under a mastery learning program. Specifically, mastery learning strategies should be able to raise the achievement levels of approximately 80% of the students to levels achieved by the upper 20% under non-mastery conditions (Bloom, 1976, p. 5). These claims are often either implicit or explicit in the writings of the original advocates of PSI (e.g., Keller & Sherman, 1974, p. 36), although many users of PSI make no reference to them. Keller's original program required 100% mastery of all unit quizzes. Since then, lower requirements such as 85% have become more common, especially in mathematics. A systematic experiment varying percentage mastery requirements was conducted by Block in 1970. Eighth graders learning elementary matrix theory were randomly assigned to five groups, distinguished only by the different levels of mastery required: no mastery, 65%, 75%, 85% and 95%. Block found that 95% mastery requirements had the effect of dampening student interest, and concluded that 85% was a good compromise.

Whatever the actual mastery rate, mastery learning is usually considered the cornerstone of PSI. After an extensive review of the PSI literature, Hursh (1976) concluded that not only was there strong support for the importance of mastery learning in PSI, but also that mastery learning was "the most powerful of PSI components" (p. 97). As usual, these conclusions were based on non-mathematics studies. There is one study in mathematics which systematically experimented with the mastery learning variable while holding other conditions constant. This experiment was conducted by Akst (1976) in an arithmetic course at a community college. The "re-testing until-mastery" group outscored the "single-testing" group on the common final examination. Unfortunately, completion rates in these two groups were not equivalent, with fewer in the mastery learning group taking the final. Since no indication is given that the researcher adjusted for this discrepancy, the results of the study are inconclusive. However, support for the worth of mastery learning can be found indirectly in the meta-analysis by Kulik et al., previously discussed above. The fact that methods classified under "mastery learning" were the only ones with the same relatively high effect size as PSI suggests that their common feature may be at the root of both of their successes.

As for the underlying assumption of mastery learning, namely that most students can achieve mastery of any material given enough time and proper attention, little research has been conducted in PSI settings. One college did report that it changed its mastery requirement from 90% to 70% on some units because some students could not reach the higher mastery level (Hess, 1977). However, such claims are very difficult to check because of numerous confounding variables such as the overall quality of the program, the time allowed for learning and the level of student preparedness. Even so, some interesting insights into the implications of mastery learning have been gained from some well-done controlled studies. Arlin and Webster (1983), for instance, examined the time costs of mastery learning in a four-day laboratory type of experiment with seventh graders, on the subject of sailing (chosen because of its unfamiliarity to the students and the hierarchical nature of the learning material). Mastery students had to attain 80% on quizzes and received remediation if necessary, while non-mastery students did not receive any feedback at all on their quizzes. This latter condition created a non-realistic situation which must be taken into account when comparisons of the groups are made. The authors reported that the mastery students achieved more than double the scores of the non-mastery students, but they also had to spend twice the time on the materials. The most interesting results came from comparisons among members of the mastery group. First, 15% of the students originally assigned to this group were dropped from the study because they could not achieve mastery under the conditions provided, even though intensive one-to-one tutoring and ample time were available. Of the remaining 85%, the ratio of the learning time of the slowest five to the fastest five learners in the sample of 44 students was found to be a stable 2.5 to 1.0. Since the eliminated 15% were the slowest learners, this ratio would have been even higher if all students had been included. The authors also reported that results from longer-term studies which were in press were in agreement with these short-term results. This study provides evidence to refute the theory that mastery learning techniques can eliminate aptitude differences among students, and it is a warning that mastery learning can exert a heavy time cost on slower learners.

Completion Rates

That mastery learning, or any teaching technique, has not yet solved the dilemma of teaching slow learners is particularly evident in reports on developmental mathematics courses. Many reports fail to report figures, but of those that do, the following are typical. Archer (1978) reported a 51% failure/withdrawal (FW) rate in an arithmetic/introductory algebra course taught by traditional methods, with this rate rising to 68% among black males. Thompson (1977) found an FW rate of 46% in both a traditional and a flexible-credit individualized course in intermediate algebra. Steele et al. (1980) compared two versions of a modified PSI program in basic algebra; an earlier version resulted in a 40% drop rate and a 48% incomplete rate. Since she also reported that only half of the incompletes were ever finished, this is equivalent to a 64% FW rate. A modified program which included test deadlines and strong incentives to attend class led to a 25% drop rate and a 15% incomplete rate, for an equivalent FW rate of 32%. Phillips (1981) also compared two groups in a remedial algebra course at a community college: a modified PSI group had a 52% FW rate, and the traditional group had a 56% FW rate.

Two very interesting studies, both in intermediate algebra, examined completion rates (the other side of the coin) relative to method of instruction and ability level. Gindler, Marosz, and Romano (1977a) examined DF rates among high, medium and low ability students in two groups. In a modified PSI course these DF rates ranged from 11% (high ability) to 41%, averaging 27% for the entire group. For the traditional group the range was 21% to 68%, averaging 39%. Note that these figures do not include W's; it is highly likely that the DFW rate is significantly higher. In the second study, Mendez (1978) physically separated his students into high, middle and low sections on the basis of an algebra placement test. The students had a maximum of three quarters to finish the course and 61% of the high and middle section students eventually did so. However, only 17% to 34% of the low section students finished, with only 11% to 18% scoring C or above. (The higher percents occurred under PSI.) Finally, the authors developed a highly personalized course for these low students, assigning them to small groups in a small class, using a low student-faculty ratio, qualified tutors in a ratio of nine to one, diagnostic tests every other day and various multi-media material. In this setting, 27% of the students made a C or above, 38% had incompletes and 25% dropped. Using this method in a controlled experiment with other low sections, they found success rates (C or above) of 18% in the control groups and 38% in the experimental group. (Many incompletes were reported in both groups, but no indication was given of how they eventually fared.)

One exception to this trend of extremely high FW rates in pre-college algebra courses was reported by Haver (1978). In both mastery based and conventional classes in an intermediate algebra course, he found the FW rate to be about 20%.

Even in calculus the situation is only slightly better. Taylor (1977) reported a 24% DFW rate in a teacher-paced mastery learning course for continuing education students, compared to 46% DFW in a conventional course. In a similar comparative study, Struik and Flexer (1977) reported a 20% W rate and a 15% incomplete rate in a PSI course in which no one received an F, while in the traditional course there was a 22% FW rate and a 14% incomplete rate. Unfortunately, no follow-up figures were reported on the I grades. Finally, in a third study, Klopfenstein (1977) reported in some dismay that a pilot PSI course had resulted in a 51% FW rate, as compared to an apparent departmental average of 17%. More will be said about this study later.

A second word of caution must be made about the W grade. Withdrawals are often overlooked by authors, but there is strong evidence that W's are more highly concentrated among poorer students (Akst, 1976; Hursh, 1976; Harris & Liguori, 1974). Hence, any analysis of completion rates and student performance ought to include all those officially enrolled in the course, not just those receiving A-F grades. A classic case of how a consideration of W's can reverse conclusions occurred in a doctoral dissertation (Phillips, 1981). The author reported a significant difference in achievement between an experimental and a control group in remedial algebra, stating in his abstract that "Eighty-six percent of the experimental group passed the final examination, compared with 68 percent of the control group" (p. 587). However, a re-analysis of the data reveals that these percents are based only on those students who took the final examination. If the original enrollment is used as the denominator of the ratio of success, the picture changes dramatically: only 48% of the total experimental group and 44% of the total control group passed the final examination. Even in non-comparative situations, W grades need to be examined and reported by authors to avoid distorting their reports.

In summary, courses in both developmental mathematics and calculus seem to suffer from much more severe dropout and FW rates than courses in many other disciplines. Among those reporting FW rates and incomplete rates, Wesp and Ford (1982) is typical for psychology: they reported that 80% of the students initially enrolled received a passing grade. Kulik, Kulik and Cohen (1980) reported an average completion rate of around 80% in a total of 66 studies in various unspecified fields. The fact that the completion rate in developmental mathematics seems to average somewhere around 50% suggests that educators in this field need to examine the premises of master learning more carefully.

Mastery Learning in Mathematics

Another facet of Bloom's theory of mastery learning which may shed some light on the situation in developmental mathematics and in other mathematics courses as well is the stress he places on present and past quality of instruction. Bloom (1971) emphasized that instruction must be tailored to fit the needs of the individual student. In addition, he noted that perseverance (the time the learner is willing to spend in learning) can "be increased by increasing the frequency of reward and evidence of learning success. Furthermore, the need for perseverance can be decreased by high quality instruction" (p. 54). Finally, Bloom expected that for courses late in a long sequence of courses, a single term under mastery methods of learning would not bring to mastery those students with a long history of learning difficulties. This is a particularly relevant point for educators in a strongly hierarchical discipline like mathematics, especially for those in developmental mathematics. The danger of implementing PSI without considering this point was discovered by Klopfenstein (1977), who had such a dismal W rate of 38% in calculus. Only 32% of his students finished the course by the end of the first term, and eventually only a total of 47% ever finished. Although there may have been flaws in the program, the reason given most weight by the author for the failure of the program was the lack of prerequisite algebraic, geometric and analytic skills exhibited by many of the students. It may be hypothesized that the gap between a C and an A in standard calculus courses is so wide that an attempt to bridge it with mastery techniques is likely to founder with typical students. The same sort of gap may exist between failure and success in remedial mathematics courses. It may be this discovery that sparked the current keen interest in placement systems among directors of developmental programs. Ironically, Bloom considered algebra a subject with only a few prerequisites (1971, p. 55), and hence a good area in which to initiate mastery learning. He obviously had not visited any community colleges, where the number of pre-algebra courses has proliferated recently (Baldwin, 1976).

It is interesting to note that some mathematical users of PSI refer to their programs as "competency-based" (e.g., Haver, 1978). The change of name illustrates a different attitude toward mastery, namely that it is not just a worthwhile goal but a necessity for the learning of mathematics. Especially in developmental programs, mastery learning has been justified as being essential in order for students to succeed in their next mathematics courses. No systematic study of this assumption has been made, although a few follow-up studies provide evidence in its favor. Thompson (1977), for instance, found that 72% of the students from an individualized algebra course who enrolled in another mathematics course (either calculus or finite math) were successful, compared to 50% of the students taught in the traditional manner. However, the passing criterion in Thompson's "mastery" course was only 67%; it is not clear if, in fact, the two groups had achieved different mastery levels. Greater support for advocates of "competency-based" mathematics is to be found in Eisenberg's report (1981) on the grades of remedial algebra students in their next math course. Of the A students, 85% proceeded to make an A, B or C in their next course; for the B students, this percent fell to 54%; for the C students, 32%; and finally, for the D students, 15% of them made C or above in their next math course. These data are based on the records of 1,600 students gathered over a four year period at a single institution. While this study needs to be duplicated at other institutions before a final conclusion is reached, it seems reasonable in the meantime to prefer at least some sort of mastery learning requirement in developmental mathematics, and even college mathematics in general.

The Value of Repeating Tests

One last objection to mastery learning to be addressed here is that raised by those who believe that students who repeatedly retake tests are not learning much. Whitehurst (1975) has provided direct evidence to refute at least part of this claim. He examined the performance of 300 students in introductory psychology, statistics and child development courses taught with PSI, and found no significant negative correlation between final examination results and the total number of tests repeated by students. In addition, studies in mathematics which have traced the performance of PSI students in follow-up courses provide more evidence against this claim because PSI students have, on the average, done at least as well as non-PSI students (Chatterly, 1977; Maltbie et al., 1974; Mendez, 1978; Carman, 1975). However, no investigation has been made of the subsequent performance of students who take longer than one semester to finish a PSI course. While repeated test-taking may be beneficial for future performance, it is not instantly clear that students who are permitted to prolong the duration of a course are actually being adequately prepared for courses which do not include this option. More research is certainly needed in this area.

8. SMALL UNITS OF STUDY AND FREQUENT TESTING

Another characteristic of PSI is the presentation of course material in short units, followed by quizzing. Major reviewers of PSI studies, including Robin (1976), Hursh (1976) and Kulik et al. (1978) agree that the use of short units with frequent testing is, on the whole, supported by the research, although results are somewhat mixed. Typically, the number of units in a PSI course corresponds roughly to the number of weeks in the term. One study supporting the use of weekly quizzes (Williams & Lawrence, 1975) found a significant difference of 4% in the average scores on a final examination in physiology, favoring students who took weekly quizzes in addition to the five hourly exams taken by the control group. Other factors were well-controlled in this study, so that there were probably no confounding effects. However, in another study, Born (1975) found no difference in achievement among psychology students who were given either 6, 9 or 18 quizzes during the term. Born's study appears to be in the minority.

No studies have been found which address this question directly in mathematics. However, interesting information is to be found by examining comparative studies which specifically refer to the frequency of testing. Seventeen such studies were found. Nine of these reported no significant differences between the individualized and conventional methods. In eight of these nine, either the control group had an equivalent number of tests as the experimental or the latter did not have frequent testing (three or fewer tests were reported in these studies). In the ninth study (Herring, 1975), the PSI group took 14 unit tests compared to four tests taken by the control group. More will be said about this study later. In contrast, of the eight comparative studies in mathematics reporting significant differences, all in favor of individualized instruction, seven of the eight reported giving quizzes at least weekly. Unfortunately, only two of the studies referred to the frequency of testing in the control groups; both reported that it was significantly less frequent than in the experimental groups. Since the control groups in all of these studies were referred to as being traditional, it may be inferred that weekly quizzing was probably not characteristic.

While these tallies are very suggestive, they leave open the possibility that unexamined confounding variables may be the actual cause of the results. For instance, it seems quite plausible that the studies using frequent testing with their experimental groups also conformed more closely to other characteristics of the PSI model, one of which (e.g., mastery learning) may be accounting for the tally results. Luckily, there are two studies which shed some light on this matter. They are both comparative studies in mathematics, using

PSI with their experimental groups and interesting variations of the traditional method with their control groups. In a study by Thompson (1980), the control group was traditionally organized in that the period was used to give lectures and solve problems. However, it had a unique feature: the control group also had weekly quizzes, along with four hourlies. As measured by final examination results, there were no differences between the groups. This study was exceptionally well-done, with random assignment to treatments and other strong methodological procedures. In addition, it did not have any shortcomings related to W rates for the simple reason that students were not permitted to withdraw from the course--the study was conducted at the U.S. Air Force Academy. However, it should be pointed out that the student body was above average in ability before generalizations are too quickly made.

In the second study, Harris and Liguori (1974) reported a very similar situation in an introductory business mathematics course. The experimental group was taught via PSI, with 16 unit tests and an 80% mastery criterion. The control group, taught by the same instructor, included two small lecture/discussion classes who not only took weekly quizzes but also had homework graded daily. There were no significant differences in either achievement or dropout rates between these two groups.

Both of these studies provide support for the conjecture that frequent testing (with or without mastery) can make a critical difference in students' achievement. However, a firm conclusion must wait until more data from controlled experiments are available.

9. STUDY GUIDES AND OBJECTIVES

Part of the standard PSI course is a study guide whose functions, as described by Sherman (1974), include an introduction to the course, a statement of objectives and study questions for every unit, and a description of procedures in the course. Robin (1976) reviewed research that confirmed intuition--students perform better on test questions which mirrored study questions. Hursh (1976) drew the same conclusions in his review. While no studies in mathematics have been found which focused exclusively on this feature, many authors frequently mention study guides in passing. Practice tests, a form of study questions, have become so established that they are included in many new textbooks in developmental and college mathematics. One objection which can be made about such tests is that students may focus their attention solely on the objectives included in the test, to the exclusion of other objectives in the unit. The solution to this problem is to make the practice tests comprehensive. As such a solution is not always tenable, there are grounds for claiming that practice tests may, by narrowing students' focus, be detrimental. No studies have been done to examine this possibility.

One subtle aspect of the role of study guides and objectives in PSI concerns the quality of the material. Very little attention has been paid by researchers to this matter, and rarely has the overall quality of instructional materials been evaluated. One reviewer (Robin, 1976) strongly suggested that more scrutiny be given to this issue, for it could certainly be having a significant effect on students' success or failure in the course. In fact, poor program materials may be the primary cause of program failures; since materials are usually overlooked, the failures are likely to be incorrectly attributed to some other cause, thereby blurring an already hazy picture of the root causes of the effectiveness of PSI. In addition, since so little emphasis is placed on the importance of material quality, unwary practitioners may learn this lesson only through hindsight.

10. IMMEDIATE FEEDBACK

Typically, quizzes taken in a PSI course are immediately graded in the student's presence, and tutoring usually follows. Kulik and Kulik (1979) reviewed several studies all of which supported the hypothesis that immediate feedback leads to higher achievement. These reviewers also noted the possibility that confounding variables such as proctor influence might be present; however, they found several studies which separated the effect of timing of feedback from that of the form of feedback. From these studies they concluded that timing, by itself, was the key factor. The role of the form of feedback will be discussed later.

It should be noted that the above conclusions are primarily based on non-mathematical studies. No studies in mathematics have been found either to corroborate or contradict the importance of immediate feedback of test results. As with study objectives, immediate feedback is common in many PSI mathematics courses, but its effect has not been isolated. Since it is at times inconvenient or even impossible to provide this feedback, research in this area would be useful.

11. TUTORS

Introduction

One of the primary characteristics of PSI is the use of tutors or proctors. This is a feature which has been the subject of much discussion, although not much actual research. Opinions of users of PSI on this subject run the gamut. Many pin the success of their programs on their tutors while a few denigrate their usefulness. Contradictory results from limited experimental research merely add to the confusion. One reason for this situation may be the lack of precision in the definition of this feature. As Robin (1976) pointed out, tutors can be assigned at least three functions: the grading and reviewing of students' tests, explaining course material and answering questions, and providing social interaction. Exactly what tasks a tutor is assigned to do and how well he or she carries them out undoubtedly varies from program to program. It can also be conjectured that the importance of the tutor will vary depending on the level of course content; in a more difficult course, tutoring may become more vital. Likewise, the value of social interaction between tutor and student may be a function of the level of motivation and maturity of the students. Unfortunately, these subtleties are not often considered in the literature on this subject.

This oversight has led to unwarranted generalizations which the following review will attempt to point out. However, it should be mentioned in advance that the lack of detail in typical reports and the almost inevitable confounding of variables makes the investigation of the role of tutoring difficult.

Tutors' Role in Discussing Tests

In some programs, the tutors' sole duly is to grade and discuss tests with students. Farmer, Lachter, Blaustein, and Cole (1972), in an oft-quoted study, found that students who received this aid immediately after taking their tests performed significantly better on the final examination and retook fewer quizzes than those who did not. For students in the latter group, quizzes were returned the next day with correct answers written on their tests. In this study students also varied according to the percent of quizzes for which they received tutoring; these were 25%, 50%, 75% and 100%. The authors found no differences among these groups, and thus concluded that intermittent tutoring was an effective and efficient procedure. At this point it becomes exceedingly pertinent to note that this was an introductory psychology course in which 94% of the students apparently finished the course by the end of the semester. In addition, the tutored students averaged about 1.6 tests per unit, with the mastery criterion set at 100%. These statistics suggest that either the level of difficulty of the course or the average ability of the student population in this study was not comparable to that found in developmental mathematics programs. Thus generalizations about the value of intermittent tutoring cannot be made without further, broader-based research.

Others are also inclined to draw different conclusions from this study. Kulik and Kulik (1979) argued that the critical factor is not how but when test feedback is provided. Since the timing of feedback was confounded with the tutoring/no tutoring condition in the above study, the argument in favor of tutoring per se is undermined. The same confounding occurred in a study by Johnson and Sulzer-Azaroff (1975), which has often been quoted to support the value of tutors in PSI. In light of the results already mentioned concerning the value of immediate feedback, Kulik and Kulik concluded that the proctors' role is not critical as long as immediate feedback is provided to the students.

In another study which questioned the value of tutors, Carsrud (1979) found that introductory psychology students who met in tutorial groups for one to two hours once a week did no better than students who only met for half an hour a week "to pick up tests" (p. 47). Apparently neither group received immediate feedback, so that this study seems to support Kulik and Kulik's position that proctoring is valuable only if it is used to provide immediate feedback. However, this study does not describe in any detail the actual role of the tutors, so that it is hard to draw conclusions about their lack of effectiveness. In addition, this is one of the studies in which the individualized group, which deviated significantly from typical PSI, did not outperform the traditional control group. An alternate hypothesis is that tutoring in and of itself could not compensate for a poorly designed individualized program.

The lack of detail provided about tutoring in this last study is indicative of the general state of affairs. In reviewing the literature on the value of tutors explaining material to students, Robin (1976) also had trouble with this problem; he could reach no conclusion because of the lack of "detailed information concerning the nature of their tutoring" (p. 335). However, there are a few exceptions. For instance, Davis (1976) explains in some detail how the instructor supplemented proctor feedback by giving advice and comments to students who had taken their midterm examinations. This proved to be effective in improving students' later performance.

Methods of Assigning Tutors

Most other studies concerning the role of proctoring in PSI have focused on two issues. The first is the effect of the method of assignment of tutors to students, particularly comparing the effect of assigning students to specific tutors ("constant" tutors) as opposed to simply providing students access to a pool of tutors ("variable" tutors). The second main research issue has concerned the source of tutors, e.g., making use of students within the course as "internal" tutors, or hiring "external" tutors. As usual, reports on both of these subjects are mixed. They are also confined to non-mathematical studies. Johnson and Sulzer-Azaroff, as well as experimenting with the proctoring/no proctoring conditions referred to above, also compared constant to variable tutors and internal to external tutors. None of these conditions resulted in significant differences in performance of the common final examination, so the authors concluded that the type of proctoring did not matter, at least in their introductory psychology course. These results are in direct contradiction to those of Carlson and Minke (1974), in which groups assigned to constant proctors performed significantly better than groups assigned to variable tutors on a number of different measures. In fact, more than 50% of the students in the former group completed all units in the course (the criterion for an A grade), compared to only about 25% of the latter group. In addition, the group assigned to constant tutors progressed more steadily through the course and took significantly fewer retests. Since this study was also in introductory psychology, explanations for this discrepancy in results are not obvious. Carlson and Minke surmised that the constant tutor became a potent source of social reinforcement for the student, and it is possible that the details of their program provided greater potential for the development of such a relationship between tutor and student. However, another factor, that of organizational structure, may have had a strong influence on the results of this program. This factor will be discussed in the next section.

Tutors in Mathematics Courses

Few controlled studies of different forms of tutoring have been found in mathematics. One study did compare two groups which differed only on the factor of internal vs. external tutoring (Harris & Liguori, 1974), and found no difference between the methods. They also did not find any gain in performance among the internal proctors (as compared to students of equivalent background in the other group), and thus questioned the propriety of using this method of tutoring. This study was conducted in an introductory business mathematics course. Another experimental study in an intermediate algebra course compared two groups, both of which were placed in an individualized, mastery learning setting. According to the author, the only difference between the two groups concerned the availability of tutors for one of the groups. A unique finding was made: the tutored group did significantly worse than the non-tutored group. The author surmised that this occurred because an unhealthy dependency relationship developed between the students and the tutors. However, in light of the fact that these results are in sharp contrast to all other studies on the effects of tutoring, it cannot be given much weight. Perhaps unique unexamined circumstances affected the outcome of this study, such as the presence of a hostile tutor.

The most interesting results concerning the effects of tutoring in mathematics were reported by Carman (1975). He randomly assigned developmental mathematics students within an individualized course to three experimental groups, receiving either no tutoring, one hour of tutoring per week, or approximately 3.5 hours of tutoring per week. The academic performance of these students was followed for the succeeding four semesters. The results were impressive. Significantly fewer tutored students withdrew from the developmental mathematics course or from the college during the first semester, and in each of the subsequent three semesters significantly more tutored students re-enrolled in and persevered in college. Such significant long-term effects are extraordinary. Carman concluded that:

> The primary effect of tutoring with low ability students in a developmental mathematics course involves positive changes in attitudes and self-concepts that are reflected in increased persistence in the course, in the college, and in other courses during the semester in which tutoring takes place. . . . The long-term effects of tutoring involve a marked increase in the persistence of tutored students in courses and in the college during the three semesters after tutoring takes place (p. 624).

This study in and of itself provides strong evidence for the value of tutoring in developmental mathematics.

Additional evidence in favor of tutor effectiveness can be found in numerous anecdotal accounts of the use of tutors in individualized programs, although the usual limitation of such accounts must be borne in mind. Eisenberg and Browne (1973), for instance, attributed the improvement in their remedial algebra course to the inclusion of recitation sections staffed by undergraduate tutors. Typically, the tutor effect is confounded with another variable, in this case organizational structure of the course (to be discussed in the next section). Even so, the judgments of program directors--especially since they concur with one another--may be of some value. Hecht (1977) believed that a key element in her successful arithmetic/introductory algebra program was the presence of concerned and compassionate tutors who kept in close contact with their assigned students. Likewise, Hassett, Livermore, and Weis (1977) reported that the use of internal tutors met with success in their intermediate algebra course. In a tutorial program in chemistry, Kean and Welsh (1980) emphatically stated that "The structured interaction of students and tutors is the primary factor which can affect the success of an academic assistance program" (p. 43). In a program in which the instructor played the role of tutor, Johnson and Steffensen (1977) echoed these sentiments: "The personalized individual one-to-one instruction which fosters a better student-teacher relationship is by far the most important aspect of our total program" (p. 53). Steele et al. (1980) found that experienced, well-trained tutors were able to increase student attendance in an individual mathematics course, with the result that drops and incompletes were cut in half from one term to the next.

One last source of information is a pair of meta-analyses on tutoring programs in the elementary and secondary schools. Hartley (1977) analyzed relatively small-scale tutoring programs in mathematics, as previously described, and concluded that tutoring was superior to traditional instruction as well as computer-assisted instruction, instruction via learning packets and programmed instruction. The effect size for tutoring was a moderate 0.6. Cohen, Kulik and Kulik (1981) did a broader meta-analysis of 65 comparative studies of tutoring in several subject areas, and found the same effect size in mathematics as Hartley did, although the overall ES was only 0.4. It appears that tutoring may be more effective in mathematics than in other subjects, at least for small-scale programs in elementary and high schools.

At least one relatively strong conclusion can be drawn from this assorted information about tutoring: tutors in individualized mathematics programs are potentially effective, especially in encouraging perseverance. As to whether they are effective because they provide immediate feedback, or because they help students to understand the material better, or because they form a social bond with their students that reinforces success-oriented behavior, these questions remain open. The fact that there is evidence in developmental mathematics for all three suggests that perhaps all three may be important aspects of tutoring, at least for academically weak students in a subject which they find difficult. It is also not clear if constant or variable tutoring produces significantly different results, but there is some evidence that constant tutoring may influence students to work at a steadier pace and attend classes more regularly (if there is a class to attend). Since these may be of importance in courses with high withdrawal and incomplete rates, the value of constant tutors may be higher in such courses. In sum, there seems to be a potential for students who are assigned to a specific tutor, who receive immediate feedback on tests and frequent one-to-one attention, and who develop a social bond with their tutors to either perform better or persevere longer in developmental mathematics. This hypothesis bears investigation.

Student/Tutor Ratios

One last practical question concerning tutors regards the ideal or at least practical ratio between tutor and students. There does not seem to be any experimental research on this matter in mathematics, but many practitioners, having apparently learned from experience, are not loathe to offer opinions. Green (1971), who conducted a PSI program in physics at Massachusetts Institute of Technology, recommended a 10:1 ratio, but certainly no higher than 12:1. Bijou et al. (1976) also used a 10:1 tutor ratio in a child development course. On the other hand, both Haver (1978) and Cameron (1977) apparently found that a 10:1 ratio in intermediate algebra was workable, as did Kahn (1975) in physical science. Toward the other end of the spectrum, Semb et al. (1975) used a 7:1 ratio in introductory child development and Carlson and Minke (1974) used a 6:1 ratio in introductory psychology. It is most certainly true that the ideal tutor/student ratio is a function of the number of duties the tutor is expected to perform, which may account for some of the variation in the ratios reported above. One might surmise that a 10:1 or 12:1 ratio might be "safe," but of course practical matters such as budget considerations and availability of tutors have to play a role in such a decision. It would seem, however, that 20:1 is probably the upper boundary for a workable ratio.

12. ORGANIZATIONAL STRUCTURE

Introduction

The last characteristic of PSI programs to be discussed here is unusual. It is not one that appears in most lists of the pertinent features of PSI, it has not received the attention of experimenters or reviewers, and it is not emphasized in most descriptions of PSI programs. This characteristic, organizational structure, may be best introduced by raising some of the questions related to it. Is the course organized in traditionally sized classes of 20 to 30 students? Does the class have a regular teacher or is it supervised by tutors? Or is there, instead of regular classes, a "lab" (or "learning center" or "testing center") to which students come whenever they want to take a test? Are they required to report to the lab on a regular basis? Do they ever have contact with the instructor? Who, if anyone, monitors their progress? Is attendance expected at a particular time of the day? Many of these questions are related to the factor of class size, but they not only have to do with the size of the group to which the student belongs but also with the type of leadership provided. This leadership may be from a tutor or a regular faculty member; it may be forceful or inconsequential. On one end of the spectrum are programs with low structure, characterized by a large course in a lab setting with no scheduled meetings of any sort. In this setting students simply show up whenever they wish, and are given tests and graded by whoever happens to be staffing the lab at that time. High structure, on the other hand, is characterized by a course based on regular class meetings, with an instructor and possibly tutors present. Between these two ends of the spectrum are other arrangements providing various levels of structure via a variety of methods.

Once the existence of this spectrum is recognized, it is then possible to look for patterns related to success or failure. One major obstacle in the path of this investigation, unfortunately, is the lack of detail usually provided on the structural organization of many programs. Not atypical is Keller's account of the first fledgling trials of PSI in this country (at Arizona State University in 1965). He mentioned that two different organizational structures were used: a large class of 94 students with 10 proctors, and a smaller class with the instructor and a graduate teaching assistant acting as tutors (Keller, 1974). Very few further details were provided. Since then courses in PSI have evidently taken on a variety of shapes and forms, but even in full-length articles the basic structural features are barely discernible. Despite this handicap, the following discussion will focus on what these features have been in mathematics and how they may have affected student performance.

Categories of Organizational Structures

Before investigating the structure of various individualized mathematics programs, it is useful to describe a number of general categories which range from high structure to low structure: (a) a regular class with an instructor, by which is meant a class of 20 to 40 students which meets regularly several times a week, and which may or may not contain in-class tutors; (b) a regular class with tutors, which fits the description for (a) except that only tutors are regularly present; (c) a lab setting in which some structure is provided, e.g. by requiring students to report in on a regular basis, either for testing or one-to-one tutoring or for a small-group meeting; (d) a lab structure with optional structure, such as non-mandatory lectures; (e) an unstructured lab setting, without any meetings or schedules of any sort.

Even within these categories a great deal of variety can exist. In fact, several of the variations in PSI which have already been discussed are actually factors influencing the amount of structure in a course. Test deadlines, for instance, are a means of requiring students to check in regularly. Likewise, tutors who are assigned duties which increase their level of involvement with students, such as carefully monitoring their progress, are also adding structure to the course. The impact of the instructor may also vary a great deal relative to this point. In fact, the elusive qualities of instructor attitude and commitment to the program may have the strongest bearing of all on how well structural supports are implemented. It is thus no straightforward matter to judge the degree of structure present in a program.

Studies of Organizational Structure

The best starting point for examining the different types and relative effectiveness of organizational structures is to review the handful of reports addressing this very question. In one such study, Slate (1975) compared four modes of operation within an individualized community college course in arithmetic. All students within this course had access to audio-tutorial (A-T) equipment and tutorial services in a math lab, but these were supplemented for the four different groups as follows: (a) no supplementation or "self-instruction"; this fits the description of the unstructured lab setting; (b) lab work supplemented by a weekly lecture for 32 students; (c) lab work supplemented by a weekly group discussion with the instructor and 16 students; (d) lab work supplemented by a weekly seminar of eight students led by the instructor. The seminar group proved to be more effective than all the others on both an attitude and an achievement measure. Self-instruction was the least effective on the achievement measure, and both self-instruction and the lecture group actually showed declines on the attitude measure. This well-done study provides evidence that regular meetings are effective, with smaller interactive meetings with the instructor being the most effective. A previously described study by Carman (1975) which examined the effects of different levels of tutoring, also supports this theory. In the latter study the tutoring effect was confounded with the effect of regular interactive weekly meetings, so that the differences found may actually be primarily attributable to the latter factor.

A report by Eisenberg and Browne (1973) adds more evidence to support the advantages of interactive or individualized meetings over formal lectures. They compared an old system based on four large lectures per week (with 125 students per lecture) to a new system in which lectures alternated with small group recitation sections staffed by undergraduate tutors. The new system was said to improve test scores and attitudes and to reduce the attrition rate. A non-rigorous comparison of final grades (excluding W's) found that 62% of the students under the new system made C or above, compared to 53% under the old system. Since the authors attributed the improvement to the use of turors, they replaced the lectures with 10-15 minute videotapes followed by tutor-supervised study. Unfortunately, the lack of statistical analysis and general vagueness of this report limit its value.

The above studies all suggest that a course with low structure, typified either by an unstructured lab setting or the additional provision of large lectures, is not as effective as courses with higher structure. This was the central theme of an article by Greenwood (1977). The latter focused on the amount of time the student is required to spend with the instructor and what role the instructor can play in an individualized classroom. Defining teacher supplementation to mean any level of teacher involvement beyond simple supervision (such as tutoring or giving supplemental lectures), Greenwood categorized 26 comparative studies in individualized mathematics according to the presence or absence of this characteristic. The outcome was amazingly clearcut: of the 16 studies with teacher supplementation, 10 favored individualized instruction and six reported no significant differences; of the 10 studies without teacher supplementation, six reported no significant differences and four favored traditional instruction. As usual, the possibility of confounding exists with box scores. However, at the very minimum Greenwood's survey suggests that the factor of teacher involvement is worth investigating.

Indirect Information Concerning Organizational Structure

The next source of information on this question of structure is a set of individualized mathematics programs culled from journals and dissertation abstracts, excluding those with inadequate details. There is a rather striking pattern to these studies: most of them fall into category (a), i.e., they are based on regular class meetings supervised by faculty members. This is a rather surprising discovery. Among other things, it indicates that support for the effectiveness of individualized instruction in non-classroom settings is based on a significantly smaller data base than might otherwise have been supposed. The other obvious fact that emerges when these studies are categorized according to the scheme listed above is that no distinct pattern of successful programs exists. Further inspection will show that this is due to the high degree of within-category variation. For instance, regular classes vary according to Greenwood's "teacher supplementation" factor, and programs in lab settings differ enormously on the intensity of relationship between staff and students. Thus it is necessary to examine individual programs carefully for characteristics which affect the degree of program structure. Because certain insights are to be gained by conducting these investigations within categories, this procedure will be followed starting with the low structure end of the spectrum.

Programs without Scheduled Class Meetings

Only 10 studies were found which provided enough information to indicate that the programs were not structured around scheduled class meetings. Of these, two were rather unique. Weir (1977) conducted a PSI course in linear algebra for 16 students at the Naval Postgraduate School. Although there were no scheduled class meetings, these students had access to the instructor and two tutors almost constantly, and a great deal of private tutoring took place. Another study (Anderson & Pritchett, 1977) was conducted in a calculus program at a small elite private residential college. Here the student/tutor ratio was seven to one, and the student/instructor ratio was apparently around 15 to one. The instructors spent at least eight hours per week tutoring students individually and monitoring their progress. Both of these programs showed PSI to be superior, but neither can be regarded as lacking structure. In addition, the student populations were better than average.

Two more non-classroom programs utilized low-structure audio-tutorial instruction. In a program involving over 200 remedial algebra students per term, Billstein (1977) indicated that no structure was provided beyond use of a test calendar for the six tests and final examination. The staff included one director, one graduate student and four work-study students. No comparative statistics were provided for this large operation, although the results of a pilot study favored the audio-tutorial students. However, the results of this pilot study carry little weight as no data were collected to indicate that the two nonrandom groups were equivalent, especially on the matter of completion rates. The larger program, which was established after the pilot study was completed, was reported to suffer from a 50% incomplete/failure rate, with fewer than 17% of the incompletes being resolved. However, no data on completion rates were provided for either of the two groups in the pilot study, so that comparisons cannot be made. One may be suspicious that nonequivalent groups in the pilot study took the final examination.

In a better-controlled and more adequately reported comparative study, Morman (1973) used audio-tutorial instruction with 61 students and one instructor. Students reported to a media center for tests, which they had to take according to a set schedule. The instructor was not present in the media center but could be sought out for assistance. Apparently there were no tutors. There was no significant difference in performance between the audio-tutorial group and the traditional group; withdrawal rates were 51% and 44%, respectively.

In sum, these last two reports do not support the use of a low structure audio-tutorial program to gain superior results.

Two large programs were reported which emphasized lab work. One of these (Zwerling, 1977) was a community college remedial mathematics program handling over 550 students per year. Students were scheduled to spend five hours per week in the lab and apparently relied heavily on programmed texts. Three instructors were constantly available and moved frequently among the students providing help. Presumably other staff was available for handling tests, At least one-half of the students finished the three-course program in one year and eventually 87% completed the program. These students also did well in their subsequent math courses, with 80% to 85% completing all of their math requirements. These statistics indicate that the program was relatively successful. It may be surmised that moderate structure, provided by required lab attendance and a dedicated staff, contributed to the success of this program.

The second large program (Chatterly, 1977) consisted of 1,500 pre-calculus students. These students had the option of either working in a nonstructured lab environment or attending regular lecture/discussion classes. Most of the students opted for the latter. Two retests were permitted on each of the eight modules and there were specific module deadlines. In a comparison with regular students, the modular students were found to have superior performance not only in their pre-calculus course but also in subsequent math courses. Unfortunately, no steps were taken to adjust for the higher dropout rate reported in the modularized course, so the results of this large scale study are questionable. Even though these two reports of large scale programs do not contain too much information, it is nonetheless interesting to note that each provided students with structural support, and that most students in the latter study preferred to work within an even more structured traditional setting.

The remaining four lab-oriented individualized programs are all based on PSI, but also vary considerably. Pascarella (1978) did a comparative study using a relatively small calculus class for mathematics and science majors, with around 60 students. These students may have been enrolled in several smaller sections and had access to optional lecture/problem-solving sessions. A lab was open every afternoon for testing and tutoring, but no information was supplied about tutors. In fact, very little information was provided about students' activities in general, so that it is difficult to know how much structure existed in the course. A well-done analysis indicated that the PSI students outperformed the traditional students, even when students receiving D's or F's in the lecture sections were eliminated from the study. Given the lack of information, few conclusions about structure can be drawn from this study. It should be noted, however, that students in this study were potential math and science majors, not average math students.

Hassett et al. (1977) on the other hand, worked with over 500 intermediate algebra students in his PSI program. The course centered on a testing/tutoring room. There were no lectures, but weekly tests and quizzes were required. The staff consisted of two faculty members, two teaching assistants and 10 work-study students. Under this new system the success rate (% ABC) was 65%, compared to 42% under the previous system of conventional classes taught by teaching assistants. Such comparisons across years leave open the possibility that other factors may be the cause of these differences. In this case the textbook and presumably the tests changed from one year to the next, but the increase is so large as to support the hypothesis that the new program was more successful than the old for more substantial reasons. The percents reported above occurred before mastery criteria were implemented, but apparently similar or even higher success rates were achieved under mastery learning. The structure of this apparently successful program is not clear, but it apparently does not include much personal interaction between instructors and students--the ratio is 90 to one. However, extensive tutor-hours are reported, although it is not known if tutors were assigned to particular students. The only feature known to be adding structure to this course was a set of test deadlines. The lack of information about this course is unfortunate, for it may be an example of a successful even though relatively unstructured program. Further details would be needed in order to draw such a conclusion.

Rogers and Young (1977) had a less successful experience with PSI in an introductory statistics course for the behavioral sciences, with an enrollment of about 200 students. This course operated entirely via a Mathematics Learning Center and was totally unstructured. No deadlines or schedules of any sort were imposed. After several years in this format the course was changed back to a traditional mode of operation because of the high dropout rate experienced under this form of PSI. This rate was reported to be 66% one year, in marked contrast to the rate of 28% experienced in the traditional format the following year. Given these abysmal statistics, the fact that final examination performance was superior under PSI becomes irrelevant. This is the only example of a completely unstructured course found in the literature, and it may be wondered how many similar programs have gone unreported in the face of overwhelming failure. This study certainly adds support to the hypothesis that more structure leads to better programs. It is ironic that this program was lauded by its implementers in an anecdotal article in the American Mathematical Monthly in 1974, before they decided to dismantle it.

The last report of a program in a non-classroom setting is that by Herring (1975). This doctoral dissertation reported no significant differences between a PSI and a traditionally taught group in a mathematics course for liberal arts majors. Despite initial appearances, the structural level of this program was relatively low. Even though two regularly scheduled classes with enrollments of 30 and 19 were originally assigned to the PSI method, only two more class meetings were held after the first day; these were at one-third intervals during the semester. Proctors were available during the scheduled class times but students were not required to attend. In fact, attendance at the 8:00 A.M. time period was described as "sparse." Apparently the instructor rarely had direct contact with most of the students, although he did monitor students' progress and assign tutors to contact and encourage procrastinators. There were no test deadlines and students had to finish only eight of 14 units to have a passing grade up to the final exam. Since procrastination was described as a very serious problem, it is not unlikely that many students who took the final examination with some hope of passing the course had finished far fewer than 14 units. This could easily explain why the PSI students did no better than traditional students on this exam. The author does not provide data on this point, so it must remain speculation. However, figures on dropouts are provided: 10 of the 49 PSI students either formally dropped or failed to show up for the final exam; only one of the 16 traditionally taught students fell into this category. The lack of attendance requirements and test deadlines and the apparent lack of contact between students and instructor reveal that the course had a very low level of structural support, which may have been a prime factor affecting its poor performance.

A similar experience occurred at The University of Tennessee at Chattanooga, where an unstructured PSI course in basic algebra was conducted from 1977 to 1981 (Smith, 1982). This program is discussed in detail in Chapter III. The passing rate in the large day sections of the course (containing hundreds of students) never rose above 20% after the first semester of operation. There were no test deadlines and no attendance requirements, although optional lectures were available. Mastery was set at 100% and the textbook and the tests keyed to it placed more than the usual emphasis on word problems. With a ratio of at least 100 to one, the instructors did not have much opportunity to become personally involved with their students, especially since they spent only one hour a week in the math lab. The case for higher structure can almost be made on the basis of this course alone.

All of the above programs operating in non-classroom settings make a single main point. It is simply that successful programs contain at least some minimal structure. In fact, two of the reports favoring individualized instruction indicated that regular classes were available to the students (Pascarella, 1978; Chatterly, 1977). Two other successful programs incorporated intense teacher/student or tutor/student relationships (Weir, 1977; Anderson & Prichett, 1977). Another also apparently included this feature in spite of a high student/teacher ratio (Zwerling, 1977). Another successful program enforced a rigid test schedule, had a large number of tutors and a zealous staff (Hassett et al., 1977). Two of the remaining three reports (Billstein, 1977; Morman, 1973) were of audio-tutorial programs with low structure, neither of which could be shown to be more successful than conventional instruction. The last report (Rogers, 1977) of an unsuccessful unstructured statistics course merely confirms the trend apparent in the others. There is evidence to indicate that the inclusion of structural support can improve the performance of an individualized course, and that the lack of such support can prove to be disastrous in an unstructured lab setting.

Programs with Class Meetings Supervised by Tutors

As for the numerous programs identified as including regular class meetings, only three of almost three dozen were staffed by tutors. None of these three was subject to well-controlled experimentation. However, Haver (1978) did report comparative data favoring PSI on a common final examination in intermediate algebra, but equivalence of the two groups was somewhat suspect. Jackson (1978) wrote a dissertation on the same program and also showed a significant difference in performance. Unfortunately, he was not able to obtain data on attrition rates, so the results must remain inconclusive. The tutor-run classes in this program alternated with lectures given by teaching assistants or faculty members, and attendance was required of all those behind schedule. This latter requirement was the central issue of a report by Steele et al. (1980). In her program, classes supervised by undergraduate tutors met regularly for three hours a week. As previously reported, this program had an extremely high dropout rate until test deadlines were enforced and attendance was strongly encouraged. The third program in this category (Eisenberg & Browne, 1973) originally was like the one described by Haver and Jackson, but was modified to more closely resemble Steele's program, with the addition of 15 minute videotapes.

Comparisons across semesters indicated that use of tutor-supervised classes increased performance.

With so few reports of tutors supervising classes, no strong conclusions can be reached concerning their effectiveness. However, there is enough evidence to suggest that such programs may be effective as long as other structural supports such as test deadlines and attendance requirements are provided. The value of lectures in such programs is not clear. Although there is evidence from other situations against their effectiveness, no study has been found which tests their value in this particular setting.

Programs with Class Meetings Supervised by Faculty Members

Twenty-nine studies were identified as using instructor-led classes. They included several different types of individualized instruction and will be examined according to type.

Eight of these studies seemed to have the major characteristics of PSI. Of these there were five reports of significant differences in performances, all favoring PSI, and three reports of no significant differences. These box scores are actually somewhat misleading, as all three of the latter studies are special cases. They include the reports of Harris and Liguori (1974) and Thompson (1980), which have already been discussed at some length because their control groups incorporated some of the apparently vital features of PSI. The third is a report by Klopfenstein (1977), who tried without success to use PSI in a calculus course for science and engineering majors. For this course, mastery for each of 20 quizzes in a 10-week term was set at 100%, there were no test deadlines (even at the end of the quarter), and the course grade (A, B or C) for students finishing the course depended only on their performance on the final exam. It may be conjectured that the exceedingly demanding nature of this course, combined with the weak pre-calculus backgrounds of many of the students and the lack of schedules, led to the high dropout rate (twice the rate of regular courses).

Of the five courses reporting significant differences in favor of PSI, one was an arithmetic course, three were developmental algebra courses and the fifth was a math course for non-science majors. One study (Gindler et al., 1977a) failed to describe any structural details, barely providing enough information to deduce that teaching assistants supervised PSI classes of about 30 to 35 students. Of the remaining four studies, three (Maltbie, 1974; Akst, 1976; Mendez, 1978) clearly indicated that strong personal contact existed between the instructors and their students. The fourth program (Peluso & Baranchik, 1977) was an exception; the instructor's duties were limited to supervising the tutors, giving occasional lectures and assigning final grades. As for class size, four of the five studies provided enough information to deduce that the range was 20 to 35 students, with one to four tutors per class. Also, in at least four of the five programs, students did not take the final exam until all units were completed (unlike Herring's program). It is also interesting to note that only one program (Mendez, 1978) used lectures in PSI and even here the lectures were not more than 15 minutes long.

As for dropout rates, they were roughly equivalent in two studies, lower for PSI in one and higher in another, and unreported in the fifth study. Unfortunately, Akst did not adjust for the lower completion rate in PSI, so the results of two studies are somewhat suspect. On the plus side, two of the five reports included information showing the equivalent or superior performance of PSI students in subsequent mathematics courses. On the whole, this barrage of information indicates that most, if not all, of these five programs were well-run and well-structured, which may account in good part for their success.

Seven more of the 29 mathematical studies with regular class structure lacked some of the features of PSI but included mastery learning. Only five were comparative studies, and of these only one favored mastery learning while the others reported no significant differences. One of the last four (Williams & Lawrence, 1975) did not require that all units be covered before the final exam; another (Nott, 1971) was based on programmed instruction; a third (Price, 1971) used only three tests, From information presented previously in this paper, it is not unlikely that these flaws contributed to the lack of program success. In the fourth study reporting no significant difference (Schoen, 1974), the experimental and control groups were not randomized and no analyses at all were undertaken to check their equivalence. This major methodological weakness makes any comparison of dubious value. It is nonetheless interesting to note that all but one of 75 students completed 18 study modules, with mastery set at 70%, by the end of the term. Finally, in the one study in this category favoring mastery learning, Urban (1971) failed to provide many details about his program. The number of quizzes, for instance, cannot be determined. However, it may be inferred that teacher/student interactions were high because only 45 students were assigned to a team of three teachers, who gave lectures, conducted small group discussions, and provided individual tutoring. Otherwise the structural details of this program are not known. With so few studies and so little information about non-PSI mastery learning programs, it is dangerous to draw any firm conclusions. One tentative hypothesis suggested by the meager information presented above is that mastery learning in a mathematics classroom should be augmented by other features such as frequent testing, required completion of all units, and strong teacher involvement to be more effective.

Another four studies of individualized programs in a regular class setting permitted optional retesting. Three favored the experimental and one reported no significant difference. All three of the former included frequent quizzing and other features of PSI. Struik et al. (1977), for instance, apparently included every PSI feature except required mastery learning. Only four retests were permitted on each of the 12 quizzes. Three differently paced presentations were available to the 105 students in the program, and the authors reported that positive comments were made on end-of-the-term evaluations about the amount of personal attention and contact they received. The experimental group outperformed the traditional group on a common final even after the D and F students were eliminated from the latter. The only major shortcoming of this experiment was that students self-selected the experimental course. Apparently, as a result, differences in prior mathematical ability favored this group. Although analysis of covariance was used to attempt to correct for this difference, this problem casts some suspicion on the results.

Miller (1976) also reported a significant difference in performance favoring the experimental group. His program included weekly quizzing and the use of study objectives, but instruction was primarily lecture-based and retesting was limited to one per quiz. The third study favoring the use of retesting (Wine & Olan, 1983) also limited it to one per quiz. However, this course in intermediate algebra was otherwise very close to PSI: most of the class period was spent on individual work, with the instructor acting as tutor; homework was assigned and corrected regularly, nine quizzes were given during the semester, and deadlines for both were enforced. On a common final the students in this experimental program outperformed the traditionally taught students, although not by a wide margin. The authors observed that withdrawal rates were equivalent.

The one study in this group which did not report a significant difference in favor of retesting was for a program in remedial algebra (Merritt, 1974). There was a seven-point difference in average unit scores favoring the retesting group, so the problem here may only have been lack of statistical power. However, only five tests were given during the term, compared to at least nine in the other programs. It may be that the availability of a total of only five retests weakened the influence of retesting on overall results. In any case, the results of this study do not supply any evidence to contradict the hypothesis that in a regular classroom setting, even limited retesting may lead to improved performance if enough structural support is provided.

The remaining 10 studies of programs set in regular classrooms either failed to provide adequate structural information or lacked reliable comparative data, so they were not analyzed. For what it is worth, of the seven rather dubious comparisons six favored individualized instruction over traditional instruction. These were the last of the individual reports examined concerning the variability and influence of program structure in individualized mathematics courses.

Summary

The above investigation of 42 individual studies in individualized mathematics at the college level has provided information about the effect of organizational structure on student performance in 10 lab courses, three courses based on tutor-supervised classes, and 29 courses based on instructor-supervised classes. Although analysis is hampered by the lack of detail and various methodological problems found in many of these studies, certain patterns can be discerned, especially when reports comparing different structural organizations are considered (Slate, 1975; Greenwood, 1977). First of all, individualized courses which provide extremely low structure seem to be very unsuccessful, as manifested by their extremely low completion rates. Studies by Rogers and Young

(1977), Herring (1975) and Smith (1982) support this conclusion: there are no studies reporting successful unstructured programs. Next, greater structure, achieved in a variety of ways, seems to increase the probability that the individualized program will be successful. In successful lab courses, increased structure was achieved either by intense teacher-student involvement (Anderson & Pritchett, 1977; Weir, 1977) or by a rigid weekly test schedule (Hassett et al., 1977), or by required lab attendance and a dedicated staff (Zwerling, 1977). The bulk of the programs in the literature provided structure via regular meetings led by regular faculty members. In such settings frequent interaction between instructors and their students usually takes place. If these types of programs also incorporated the features of frequent testing with required or optional retesting, along with most of the other features of PSI including the use of study guides and objectives, tutors, and a de-emphasis of lectures, they proved to be almost invariably better than their traditional counterparts. It may thus be hypothesized that a relatively high degree of structure, achieved in various ways, is a necessary (although not sufficient) condition for a successful individualized program.

13. CONCLUSION

The above review of the literature indicates very clearly that the success of individualized instruction in mathematics is heavily dependent on the specific characteristics of the program.

Unfortunately, it is not yet possible to categorically state which characteristics are necessary and sufficient for a successful program, as no general method has proved to be consistently effective, and only limited research has been done on component analysis in mathematics. However, the preceding close examination of studies in mathematics, combined with information from more general reviews, reveals certain patterns of success which provide evidence to support the value of certain key characteristics of an individualized program. A number of these characteristics are incorporated into PSI, including frequent testing, immediate feedback of test results, required or possibly optional mastery, and the use of study objectives (usually in the guise of practice tests). However, it is also clear that these characteristics in and of themselves are not sufficient for a successful program. In particular, pacing contingencies to control student progress seems to be absolutely critical in most individualized mathematics courses. In addition, a few research reports and numerous anecdotal accounts strongly suggest that the use of tutors may be beneficial, but their usefulness seems to depend strongly on the manner in which they are incorporated into the program. As to specific tutoring behaviors, little research has been done in mathematics to determine the relative effectiveness of such variations as constant or variable tutoring, and more intensive or less intensive involvement of tutors with their students. Finally, there is evidence to indicate that the organizational structure of a program, a characteristic which has received little attention, has a substantial influence on the success of the program. Courses which provide structure, either by regular class meetings, attendance requirements, test schedules and/or high levels of interaction between students and their teachers, or possibly tutors, seem more likely to be successful than programs which fail to provide such structural support.

CHAPTER III

THE INDIVIDUALIZED REMEDIAL MATHEMATICS PROGRAM
AT THE UNIVERSITY OF TENNESSEE AT CHATTANOOGA

The review of the literature presented in Chapter II illustrated the need not only for more reports of individualized programs in mathematics but also for more thorough descriptions of such programs, so as to increase the benefit both to practitioners and researchers. This chapter is included to help meet this need. It is a description and analysis of events taking place in the individualized developmental algebra program at The University of Tennessee at Chattanooga, the site of the experiment to be described in Chapter IV. The difficulties experienced in this program indicate some of the subtleties involved in developing and implementing a successful individualized remedial mathematics program.

An individualized program in remedial algebra was instituted at The University of Tennessee at Chattanooga in the Fall of 1977, with an enrollment of approximately 250 students. This program was very unstructured even though it included the main features of PSI, including mastery learning (with a criterion of 100%), small units of instruction, written study objectives in the form of practice tests, optional lectures and a Mathematics Learning Center (Math Lab) which served as a tutoring as well as testing center. However, there were no class attendance requirements, no homework assignments and no test deadlines. Although no comparative study was ever carried out, this program could by no means be classified as a success--it never achieved a completion rate over 20% after its first semester of operation. One obvious problem was the complete lack of deadlines; students who failed to complete the course in one semester were assigned non-penalty grades of no credit and permitted to continue where they had left off whenever they decided to re-enroll in the course.

In 1981 the course was restructured. An alternating lecture/workshop format was established with attendance expected four days a week. End-of-semester deadlines were instituted to require students to complete at least half of the course in a semester. In addition, in order to continue where they had left off, students were required to immediately re-enroll in the course. Finally, a more readable and better written text was introduced, and the mastery criterion was changed to 85%. These changes did have a significant effect on student performance: almost one-third of the students finished the course in one semester and half finished by the end of two semesters. Even with this improvement, however, the completion rate was not considered satisfactory and the course was once again revised. This time biweekly test deadlines and penalties for poor attendance were introduced. Neither of these techniques improved completion rates significantly.

An evaluation of this program in light of the preceding review of the literature leads to several hypotheses about its possible weaknesses. First of all, the structure of the class did not permit instructors to become familiar with most of their students--the student-instructor ratio was 120 to one. It was hoped that the interaction of students with their tutors would compensate for the lack of contact with the instructor; there was a 20 to one ratio of students to tutors in the workshops. However, various indicators such as discipline and attendance problems suggested that tutors in their current role were not as influential as desired. Their activities included answering students' questions, checking homework and taking roll. They were not usually directly involved with students' test-taking, although they received weekly reports of their students' progress. Exchanges between tutors and students were usually initiated by the students, and tutors were not consistently active in observing and encouraging their students' progress. While tutors eventually became familiar with students who attended their workshops regularly, feelings of anonymity may easily have pervaded the first few weeks of the course.

Another problem in the course concerned communications, particularly in regard to test deadlines and attendance policies. Students receiving F's based on poor attendance records often acted surprised, either expressing ignorance of the penalty for poor attendance or disbelief that they had exceeded the permissible number of absences (eight out of approximately 58 class sessions). Students also missed test deadlines because of ignorance of course procedures, but a more common reason seemed to be that students did not allow themselves time to retake failed tests. As mentioned previously, their test-taking was not strongly supervised.

In addition, there were various problems with the course materials. For instance, there were some discrepancies between the practice tests and the actual tests, and there were a few weaknesses in the textbook. A major problem with the course was that an entire unit was devoted to word problems. The existence of this unit undoubtedly created a major stumbling block for many students. However, it seemed unlikely that these problems in and of themselves were the main reasons for the low completion rate.

The major problems afflicting this remedial mathematics program were hypothesized to be due to the organizational structure of the course and the lack of a strong relationship between the staff and their students. The review of the literature on the subject of tutors and organizational structure provided support for these hypotheses. It seemed likely, for instance, that a regular class structure, in which the students were supervised by a faculty member, could make a significant difference in the effectiveness of the program. In addition, it also seemed likely that a more intensive relationship between the tutors and their students could increase perseverance and thereby improve the completion rate. The experiment to be described in Chapter IV was designed to test these theories.

CHAPTER IV

METHODOLOGY AND DESIGN

1. INTRODUCTION

The purpose of the experiment to be described in this chapter was to determine if student performance was differentially affected by two types of organizational structure, lecture/workshop and regular class, and two types of tutoring behaviors, more intensive and less intensive. The setting within which the experiment was conducted is described first, followed by descriptions of the sample, the three treatments used to investigate the variables, the experimental design, the planned analysis and the sources of data.

2. GENERAL DESCRIPTION OF MATHEMATICS 107

The setting for the experiment was the remedial mathematics course at The University of Tennessee at Chattanooga. This course, Mathematics 107, was approximately equivalent to first year high school algebra, with a special emphasis on solving word problems. A copy of the syllubus is provided in Appendix A. Students who placed at the lowest placement level (as determined by high school background and mathematics placement test scores) were required to enroll in Mathematics 107 before taking any college level mathematics course. Since the latter was included as part of the general education requirements for all majors, those placed in Mathematics 107 had to complete this course in order to fulfill graduation requirements, regardless of major. As previously mentioned, approximately 50% of all students tested at UTC in 1983 placed at the Mathematics 107 level, and of those many were clearly in need of an even lower level remedial mathematics course.

All sections of the Mathematics 107 program shared the following features:

1. each student was required to pass all 13 unit tests and the final comprehensive examination in order to satisfactorily complete the course;
2. a mastery rate of 85% was required for all unit tests and the final examination; unlimited retests were permitted, but students who failed three tests were required to have a conference with their instructor before being retested;
3. a mathematics lab was open to all students from 8:00 A.M. to 4:00 P.M. three days a week and from 8:00 A.M. to 8:00 P.M. the other two days of the week; the staff's time was mainly devoted to test-related activities (checking homework, grading tests, discussing tests with students, recording tests, etc.) but some time was also available for students seeking assistance;
4. homework was assigned after each failed test and it was spot-checked before another test was taken;
5. a study guide was provided for all students; it outlined

homework assignments and included tips and additional practice problems on the more difficult topics in the course;

6. all sections met for four 50-minute classes a week for 15 weeks;
7. attendance was required in all classes;
8. students who passed at least five tests usually had the option of deferring completion of the course until the Spring Semester;
9. students were provided with progress charts in their study guides; these charts contained two sets of test deadlines: (a) the first set served as a guideline for finishing the course by the end of the semester and (b) the second indicated the lowest acceptable level of progress in the course; the penalty for missing a deadline in the second set was a mandatory meeting with the instructor;
10. students were permitted to take tests in the math lab during only one of the four weekly class periods; those who needed to take tests more frequently had to do so during their free time.

3. THE SAMPLE

The sample for this experiment consisted of all students who registered for a day section of Mathematics 107 at The University of Tennessee at Chattanooga in the Fall Semester of 1983, excluding:

1. students who had previously received a non-W grade in the course;
2. students who enrolled after classes began;
3. students who either did not attend class at all or did not attend after the first day;
4. a few students who were obliged to change sections.

At the 8:00 A.M. period the sample size was 186; at the noon period it was 160.

4. THE TREATMENTS

General Description

There were three treatments, summarized as follows:

Treatment I: A lecture/workshop structure with workshop tutors engaging in low intensity relationships with their students.

Treatment II: A lecture/workshop structure with workshop tutors engaging in high intensity relationships with their students, supported by the use of a point system for grading.

Treatment III: A standard class structure of 30 students primarily supervised by an instructor, with the instructor and her tutor assistant engaging in the same high intensity relationships with their students and supported by the same point system as the tutors of Treatment II.

Comparison of Treatments I and II

Tutoring behaviors. The 75 students assigned to Treatments I and II shared a common lecture which met twice a week in a large tiered lecture hall. The lectures covered all 14 units of the course in one semester at a rate of approximately one unit per week. However, during the other two class periods of the week the students attended different types of workshops. Under Treatment II various conditions were created in the workshops with the intention of encouraging a supportive bond between tutors and their students. These conditions included specific tutoring behaviors and a different grading system. The differences in tutoring behaviors in the two treatments are contrasted in Table IV-1.

The grading systems. The grading systems for all treatments were based in part on the students' progress in the course, but they differed in regard to the use of attendance records and the point system. In Treatment I there was no point system; only the attendance records were used in conjunction with course progress to determine the final grade. In Treatments II and III attendance points were combined with points gained by meeting first or second test deadlines in order to determine the final grade. These grading systems are summarized in Tables IV-2 and IV-3, with the point system summarized in Table IV-4.

The purpose of the point system was twofold: to offer incentives for students to meet test deadlines, and to be a vehicle

TABLE IV-1

COMPARISON OF TUTORING BEHAVIORS IN TREATMENT I AND TREATMENT II

Treatment I	Treatment II
1. Thirty students were assigned as one group to two tutors	1. Fifteen students were assigned to each tutor
2. Tutors were not provided with any information about their students prior to the first day of classes	2. Tutors were provided with the names of their students prior to the first day of classes and encouraged to become familiar with them for the purpose of learning their names quickly; tutors also spent some time inspecting their students' mathematics placement scores
3. Tutors spent the first day supplying students with information regarding the effect of attendance on grading; they did not collect information cards	3. Tutors spent the first day of class supplying students with information regarding the point system of grading. They also collected information cards from their students on their math backgrounds, hobbies, physical descriptions, etc., and spoke individually to as many students as possible
4. During the semester the tutors were simply told to keep attendance records and answer their students' questions	4. During the semester the tutors were encouraged to learn their students' names as quickly as possible and to greet as many students as possible individually at each class meeting
5. Tutors never saw their students' tests	5. Most or all of their students' tests were delivered to tutors during class, and tutors gave the students a second chance to look at them and then recorded the tests on progress charts

TABLE IV-1 (continued)

Treatment I	Treatment II
6. Tutors' record-keeping involved only the recording of attendance	6. In addition to keeping attendance records, tutors also kept track of students' test-taking activities via progress charts for each student (identical to the progress charts provided to the students)

TABLE IV-2

GRADING SYSTEM FOR TREATMENT I

Grade	Criteria
S	Pass all 13 unit tests and the final examination at an 85% mastery level
I	Reach Unit 14 (the last unit in the course) and accumulate fewer than eight absences
NC	Reach Unit 10 and accumulate fewer than eight absences
F	Fail to reach Unit 10, or accumulate at least eight absences without completing the course
FF	Fail to reach Unit 6, or accumulate at least 16 absences without completing the course. Added penalty: must start the course over again at Unit 1.

Note: Students with grades of I, NC and F were permitted to continue the course into the Spring Semester.

TABLE IV-3

GRADING SYSTEM FOR TREATMENTS II AND III

Grade	Criteria
S	Pass all 13 unit tests and the final examination at an 85% mastery level
I	Reach Unit 14 and accumulate at least 67 points
NC	Reach Unit 10 and accumulate at least 67 points
F	Fail to reach Unit 10, or accumulate less than 67 points without completing the course
FF	Fail to reach Unit 6, or accumulate fewer than 54 points without completing the course. Extra penalty: must start the course over again at Unit I

Note: Students with grades of I, NC and F were permitted to continue the course into the Spring Semester.

TABLE IV-4

POINT SYSTEM FOR TREATMENTS II AND III

For meeting a first test deadline	5 points
For meeting a second test deadline	2 points
For attending each class	1 point

for promoting a stronger involvement between tutors and their students. The tutors were encouraged to regularly tally their students' points and to encourage more effort from those who were lagging in their point counts. The intention was to give students in Treatments II and III a greater sense that their tutors cared about their progress and wanted to help them meet their goals, and also to increase the tutors' awareness of their students' progress. The point system was directly tied to the use of the progress charts. By using the progress charts to keep track of points, both the tutors and the students received a visual display of students' progress.

Optional lecture attendance. In both Treatments I and II students whose test progress did not conform with the lecture schedule were permitted to attend their workshops all four class periods of the week instead of attending lecture. The schedule for the course was such that students choosing this option worked under the same tutor all four days of the week and mingled with students who were in the same treatment under the other instructor.

Description of Treatment III

Students assigned to Treatment III reported to a typical classroom for all four of their weekly classes, never attending lectures in a large lecture hall as students in the other treatments did. The instructor was present for three of the four class periods while the tutor was present for all four. Each class in Treatment III initially contained approximately 30 students. Both instructors spent part or all of the period working with the class as a group, mainly through interactive lectures. The instructor and her tutor followed the same procedures as the tutors for Treatment II, as listed in Table IV-1. As previously mentioned, Treatment III also used the same grading system as Treatment II (see Table IV-3). Thus the only difference between Treatments II and III was the structural organization of the class.

5. DESIGN OF THE EXPERIMENT

Two instructors participated in the experiment, the experimenter and a highly rated full-time instructor experienced in teaching remedial algebra. Both instructors were crossed with all treatments; thus a 2 x 3 factorial design was used. Students who enrolled at each of the two daytime periods that Mathematics 107 was offered were randomly assigned to the six cells of the design. Approximately 28 to 30 were assigned to each cell of Treatments I and III and approximately 42 to 45 were assigned to each cell in Treatment II.

The replication of the experiment at the noon period used the same two instructors, although most of the tutors were different.

Instructor 1 lectured to her 75 students in Treatments I and II on Mondays and Wednesdays, then met with her students in Treatment III on the other three days of the week. Instructor 2 had a similar schedule except that she lectured on Tuesdays and Thursdays. Because the two instructors' workshops for Treatments I and II met on different days of the week, their workshops shared the same tutors. However, the tutors for the two Treatment III classes were different. A list of sections and their corresponding instructors and tutors is given in Table IV-5.

Due to space limitations, the first two of the three sections of Treatment II listed for each instructor shared a classroom; each section was assigned to a different side of the room. The third section of Treatment II met by itself in a separate smaller classroom.

The tutors were randomly assigned to the treatments. For the most part the tutors for the noon period were different from those at the 8:00 A.M. period, although there were a few overlaps as indicated in Table IV-5.

6. ANALYSIS

The main statistical method used to analyze the results of this 2 x 3 (Instructor by Treatment) unbalanced factorial design was the analysis of covariance. The dependent variable was the unit the student had reached by the end of Fall Semester 1983, and the covariates were the arithmetic and algebra scores of the mathematics placement test and mathematics attitude scores. Analyses of variance

TABLE IV-5

INFORMATION ON TREATMENTS, INSTRUCTORS AND TUTORS BY SECTION

Section	Treatment	Instructor	Tutor	Section	Treatment	Instructor	Tutor
20	I	1	1 & 2	40	I	1	8 & 9
21	II	1	3	41	II	1	10
22	II	1	4	42	II	1	4
23	II	1	5	43	II	1	11
24	III	1	6	44	III	1	5
30	I	2	1 & 2	50	I	2	8 & 9
31	II	2	3	51	II	2	10
32	II	2	4	52	II	2	4
33	II	2	5	53	II	2	11
34	III	2	7	54	III	2	12

were also performed using the same dependent variable. In addition, less powerful chi-square tests were performed on success rates, partial completion rates and attrition rates.

Five pre-planned comparisons were tested:

1. Is there a significant difference between Treatment I and Treatment II?
2. Is there a significant difference between Treatment I and Treatment III?
3. Is there a significant difference between Treatment II and Treatment III?
4. Is there a significant difference between Treatment I and the average of Treatments II and III?
5. Is there a significant difference between Treatment III and the average of Treatments I and II?

The Type I error was set at $\alpha = .10$. This high level was chosen in order to decrease the probability of a Type II error. This decision was made because of practical considerations: in this experiment the consequences of incorrectly rejecting a true null hypothesis were considered less severe than the consequences of incorrectly failing to reject a false null hypothesis. In the latter case the superior treatment would not have been discovered, while in the former case a treatment would have been chosen as better than another even though it was not. Since overlooking a superior treatment was regarded as the more costly mistake, the Type I error was set at a higher than average level.

Using $\alpha = .10$, the power of various statistical tests conducted in this study was calculated based on the definition of effect size and the accompanying tables given by Cohen (1977). For the test of the main effect of treatment in the analyses of covariance, the power for a medium effect size was 85%; for a small effect size the power was only 28%. For the chi-square tests in the first experiment, the power for a medium effect size was 94%, while the power for a small effect size was 24%. In the second experiment, the power of the chi-square test for a medium effect size was 91%, and for a small effect size it was 22%. Thus, with $\alpha = .10$, the power for a medium effect size is quite good for all of the above tests, although the power is extremely weak for small effect sizes.

CHAPTER V

PRESENTATION AND ANALYSIS OF THE DATA

1. INTRODUCTION

In this chapter, data and analyses of the data are presented on the effects of treatments on student performance in Mathematics 107 at The University of Tennessee at Chattanooga. Data from the first experiment at the 8:00 A.M. period will be examined first, followed by the data from the replication at the noon period. For each of these sets of data, progress in the course as determined by the current unit was analyzed via analyses of variance and analyses of covariance. Success rates, partial completion rates and attrition rates were analyzed via chi-square tests. Analyses were performed using SAS (SAS User's Guide: Statistics, 1982) and SPSS (Nie, Hull, Jenkins, Stairbrenner, & Bent, 1975). The raw data are listed in Appendix B.

1. DATA FROM THE FIRST EXPERIMENT

The Sample

The sample at the 8:00 A.M. period consisted of 186 students. Of the 208 students who were originally assigned to the cells of the design, 18 were dropped because they either never attended class or attended only the first day of class (during which only general information was disseminated). Removal of these students from the experiment was expected to reduce the error variance without introducing bias. In addition, four students assigned to Treatment III sections were dropped from the study because they could not attend class on Fridays. While Treatment I and Treatment II sections met at the officially scheduled times from Monday through Thursday, the Treatment III sections met on Fridays (as well as three of the other four days of the week) to permit the instructors to be crossed with the treatments. It was assumed that the loss of such a small number of students due to conflicts with the Friday class did not introduce serious bias into the experiment.

Course Progress

The variable used to measure student progress in the course was each student's current unit at the end of the semester. The last unit in the course was Unit 14. For the sake of brevity, students who passed Unit 14 but did not pass the final examination are referred to as being in Unit 15, while those who passed the final examination are referred to as being in Unit 16. A histogram for the current unit for the entire sample is presented in Figure V-1.

Descriptive statistics. Means and standard errors of the current unit for each instructor, treatment and instructor/treatment combination are listed in Table V-1. The number of students in each unit for each combination is tabulated in Table V-2, and percents are computed over certain unit intervals in Table V-3. Inspection of the treatment means indicates that Treatment I has a lower mean

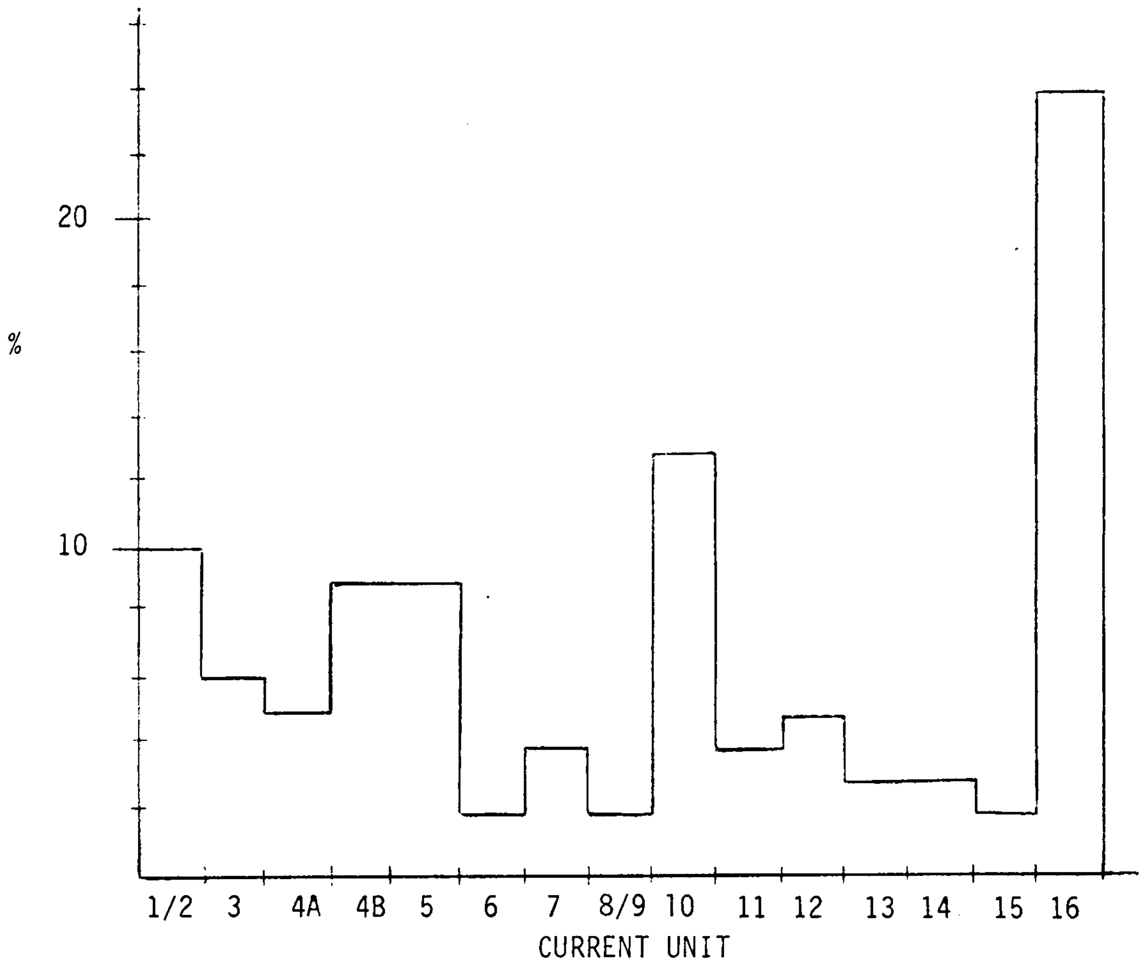

FIGURE V-1. HISTOGRAM FOR CURRENT UNIT: FIRST EXPERIMENT

TABLE V-1

MEANS AND STANDARD ERRORS FOR THE CURRENT UNIT: FIRST EXPERIMENT

Instructor	Treatment	Total No. in Group	Mean	Standard Error
1		99	9.56	.54
2		87	8.10	.59
	I	55	7.78	.72
	II	81	9.36	.60
	III	50	9.30	.79
1	I	29	9.34	1.04
	II	44	9.61	.76
	III	26	9.69	1.13
2	I	26	6.04	.91
	II	37	9.05	.96
	III	24	8.88	1.12
Total		186	8.88	.40

TABLE V-2

NUMBER OF STUDENTS IN EACH UNIT FOR EACH INSTRUCTOR X TREATMENT COMBINATION: FIRST EXPERIMENT

	Instructor 1 Treatment			Instructor 2 Treatment			
Unit	I	II	III	I	II	III	Total
1 and 2	2	4	2	4	4	2	18
3	1	1	0	2	5	3	12
4 (Part A)	1	1	4	4	0	0	10
4 (Part B)	5	3	1	3	3	1	16
5	1	4	3	4	2	3	17
6	1	0	1	0	1	1	4
7	3	1	0	1	1	1	7
8 and 9	0	3	0	0	0	1	4
10	4	9	2	2	6	1	24
11	0	2	0	2	1	2	7
12	1	2	3	1	1	1	9
13	0	2	1	1	0	1	5
14	0	1	0	0	2	2	5
15 (lacking final)	0	1	0	1	1	0	3
16 (finished)	10	10	9	1	10	5	45
Total	29	44	26	26	37	24	186

TABLE V-3

PERCENT OF STUDENTS IN SELECTED UNIT INTERVALS FOR EACH INSTRUCTOR X TREATMENT COMBINATION: FIRST EXPERIMENT

	Instructor 1			Instructor 2			
	Treatment			Treatment			
Interval	I	II	III	I	II	III	Total
1-5	34%(10)	30%(13)	38%(10)	65%(17)	38%(14)	38%(9)	39%
6-9	14%(4)	9%(4)	4%(1)	4%(1)	5%(2)	13%(3)	8%
10-13	17%(5)	34%(15)	23%(6)	23%(6)	22%(8)	21%(5)	24%
14-16	34%(10)	27%(12)	35%(9)	8%(2)	35%(13)	29%(7)	28%
Total	29	44	26	26	37	24	186

than Treatments II and III, while the means of the latter two are very close together. Examination of the combination means indicates that, while this same order is preserved for both instructors, only the mean of Treatment I for Instructor 2 appears to be well below the means for for the other two treatments.

Analysis of variance. The results of the analysis of variance conducted for a 2 x 3 unbalanced factorial design, including the results of the tests for the five pre-planned comparisons, are presented in Table V-4. The overall F was $F(5,180) = 1.78$, $p=.12$. None of the five null hypotheses could be rejected at $\alpha = .05$, but the following conclusions could be made for $\alpha = .10$:

TABLE V-4

ANALYSIS OF VARIANCE, INCLUDING TESTS OF PRE-PLANNED COMPARISONS: FIRST EXPERIMENT

Source	Type III SS	df	F	p
Instructor	108.14	1	3.72	.06
Treatment	101.74	2	1.75	.18
Instructor x Treatment	68.17	2	1.17	.31
Error	5230.00	180		
Total	5488.16			
Treatment I vs. II	87.92	1	3.72	.08
Treatment I vs. III	66.23	1	2.28	.13
Treatment I vs. II & III	99.23	1	3.42	.07
Treatments I & II vs. III	21.45	1	.74	.39

1. the hypothesis that Treatment I was equivalent to Treatment II could be rejected, p=.08;
2. the hypothesis that Treatment I was equivalent to the average of Treatments II and III could be rejected, p=.07.

Analyses of covariance with three covariates. Three covariates, including arithmetic scores, algebra scores and mathematics attitude scores, were added to the model. The expanded cell means model was thus: $y_{ijk} = \mu_{ij} + \beta_{ij}x + \delta_{ij}y + \gamma_{ij}z + \varepsilon_{ijk}$, where

y_{ijk} is the current unit,

μ_{ij} is the mean for the current unit in cell ij,

x is the arithmetic score,

y is the algebra score,

z is the mathematics attitude score,

ε_{ijk} is the error term.

The correlations among the three covariates and the current unit are presented in Table V-5, and descriptive statistics for the covariates are presented in Table V-6. The correlations between the current unit and the three covariates range from .22 to .26, all significant at $\alpha = .01$.

The use of the expanded covariance model reduced the sample size from 186 to 146 because of missing values. Specifically, six students had no arithmetic and algebra scores, five students failed to take the attitude questionnaire and 29 students failed to properly complete the attitude questionnaire. Of the first six, three also failed to take any unit tests. Since students were informed that

TABLE V-5

CORRELATION COEFFICIENTS FOR THE COVARIATES AND THE CURRENT UNIT: FIRST EXPERIMENT

Variables	1	2	3	4
1. Current Unit	--	.26*	.25*	.22*
2. Arithmetic		--	.32*	.08
3. Algebra			--	.05
4. Attitude				--

*Significant at α = .01.

they either had to take the placement test or drop the course, it seems likely that these three chose the latter option. The records of the remaining three either do not exist or are not traceable. Likewise, two of the five students who did not take the attitude questionnaire also took no unit tests and attended class only briefly; the remaining three were apparently overlooked. The greatest source of missing values was the set of 29 students who failed to complete the second page of the two page attitude questionnaire. The possibility that these missing values introduced bias into the experiment will be investigated in the next section. However, it is assumed that the first 11 missing values, including the six missing mathematics test scores and the five completely missing attitude scores, occurred in a random manner and introduced no serious bias.

TABLE V-6

MEANS, STANDARD ERRORS, MINIMUMS AND MAXIMUMS FOR THE THREE COVARIATES, BY TREATMENT AND BY INSTRUCTOR/TREATMENT COMBINATION: FIRST EXPERIMENT

		Arithmetic				Algebra				Attitude			
Instr	Trtmnt	Mean	S.E.	Min	Max	Mean	S.E.	Min	Max	Mean	S.E.	Min	Max
	I	74.6	1.85	40	100	31.2	1.59	8	68	35.1	2.21	0	71
	II	71.2	1.65	22	94	31.4	1.08	0	52	38.0	1.66	12	62
	III	74.5	1.86	31	100	30.4	1.30	12	52	34.7	2.13	12	52
1	I	77.5	2.61	45	100	33.6	2.54	12	68	34.9	2.22	21	58
	II	71.5	2.26	40	94	30.0	1.58	0	52	37.9	2.37	12	62
	III	78.0	1.68	62	94	32.2	1.74	12	52	36.3	3.09	10	62
2	I	71.3	2.51	40	88	28.6	1.74	8	44	35.4	3.94	0	71
	II	70.8	2.44	22	88	33.0	1.44	12	52	38.1	2.31	20	59
	III	70.8	3.25	31	100	28.5	1.89	12	40	33.1	2.95	15	64
Total		73.0	1.04	22	100	31.1	.75	0	68	36.2	1.13	0	71

The distribution of missing values is presented in Table V-7, along with the new means based on the reduced sample.

To determine if the full linear model could be reduced to a simpler model, the hypothesis that there were no differences among the three sets of covariate coefficients was tested. Specifically, the null hypothesis was: $\beta_{11} = \beta_{12} = \beta_{13} = \beta_{21} = \beta_{22} = \beta_{23}$ and $\delta_{11} = \delta_{12} = \delta_{13} = \delta_{21} = \delta_{22} = \delta_{23}$ and $\gamma_{11} = \gamma_{12} = \gamma_{13} = \gamma_{21} = \gamma_{22} = \gamma_{23}$. In other words, it was hypothesized that there were no interactions between the covariates and the class variables. This hypothesis was rejected, $F(15,122) = 1.81$, $p<.04$. Thus the full linear model was retained. Overall results are presented in Table V-8.

The possibility that the data would be better fitted by a quadratic model was also investigated. The results of an analysis of covariance using a model with quadratic covariate terms are presented in Table V-9. Since none of the tests for the quadratic terms were significant, the linear model was retained.

In the full linear model each treatment was represented by a hyperplane in four-dimensional space. The geometrical interpretation of the rejection of the first hypothesis tested above is that these hyperplanes were not parallel. Given this situation it was necessary to compare the treatment hyperplanes at a number of different values of the covariates to obtain an understanding of their relative positions. The behavior of the three treatment hyperplanes was first examined by obtaining the least-square means, which were the expected values of the cell means "for a balanced design involving the class

TABLE V-7

SUBSAMPLE SIZES, MEANS AND THE DISTRIBUTION OF MISSING VALUES FOR THE REDUCED ANCOVA SAMPLE: FIRST EXPERIMENT

Instructor	Treatment	Missing Values due to Incomplete Attitude Questionnaires	Missing Values from Other Sources	Reduced Sample Size	Unit Means Based on Reduced Sample
1	I	7	1	21	9.38
	II	8	2	34	10.21
	III	2	2	22	10.45
2	I	3	3	20	6.35
	II	7	2	28	9.36
	III	2	1	21	8.67
Total		29	11	146	9.21

TABLE V-8

OVERALL RESULTS OF THE FULL LINEAR MODEL ANCOVA: FIRST EXPERIMENT

Source	df	Type III SS	MS	F	p	R^2
Model	23	1279.6	55.63	2.51	.0007	.32
Error	122	2700.8	22.13			
Corrected Total	145	3980.4				
Instructor (I)	1	27.9		1.26	.26	
Treatment (T)	2	130.4		2.94	.06	
I x T	2	43.3		.98	.38	
Arithmetic (AR)	1	.0		.00	.98	
Algebra (AL)	1	187.1		8.45	<.01	
Attitude (ATT)	1	109.7		4.95	.03	
AR x I	1	24.4		1.10	.30	
AL x I	1	17.8		.80	.37	
ATT x I	1	85.8		3.88	.05	
AR x T	2	143.9		3.25	.04	
AL x T	2	31.4		.71	.49	
ATT x T	2	172.4		3.89	.02	
AR x I x T	2	100.3		2.26	.11	
AL x I x T	2	17.6		.40	.67	
ATT x I x T	2	74.7		1.69	.19	

TABLE V-9

ANCOVA WITH QUADRATIC TERMS: FIRST EXPERIMENT

Source	df	Type III SS	F	p
Cell Means (CM)	6	337.99	23.47	<.01
Arithmetic (AR) x CM	6	326.07	2.27	.04
Algebra (AL) x CM	6	116.14	.81	.57
Attitude (ATT) x CM	6	353.32	2.46	.03
AR^2 x CM	6	99.34	.69	.66
AL^2 x CM	6	39.31	.27	.95
ATT^2 x CM	6	88.77	.62	.72

variables with all covariates at their mean values" (SAS User's Guide: Statistics, 1982, p. 177). These least-squares means are listed in Table V-10. There is one apparently significant difference between these means and the raw means presented in Table V-1: the mean of Treatment I for Instructor 1 is 2.1 to 2.7 units below the means of Treatments II and III for the same instructor, a gap which is much wider than the gap of approximately .3 which existed between the comparable raw means.

To determine if in fact the differences among the treatment hyperplanes were significant at the mean values of the covariates, the tests of the five pre-planned comparisons were conducted using these mean covariate values. The results of these tests are presented in Table V-11. These results, combined with an inspection of the least-squares means in Table V-10, indicate that at the covariate

TABLE V-10

LEAST-SQUARES MEANS, FULL ANCOVA MODEL: FIRST EXPERIMENT

Instructor	Treatment	Least-Squares Mean
1		9.79
2		8.39
	I	7.62
	II	9.77
	III	9.89
1	I	8.06
	II	10.14
	III	10.77
2	I	6.81
	II	9.61
	III	9.34

TABLE V-11

TEST RESULTS FOR PRE-PLANNED COMPARISONS USING THE FULL ANCOVA MODEL: FIRST EXPERIMENT

Contrast	df	SS	F	p
Treatment I vs. II	1	176.49	7.97	<.01
Treatment I vs. III	1	147.07	6.64	.01
Treatment II vs. III	1	.00	.00	.99
Treatment I vs. II & III	1	201.78	9.11	<.01
Treatments I & II vs. III	1	40.86	1.85	.18

means the current unit for Treatment I is significantly less than both the current unit for Treatment II ($p<.01$) and the current unit for Treatment III ($p<.01$). In contrast, the current units for Treatments II and III at the covariate means are extremely close to each other ($p=.99$).

To gain more information about the relative positions of the three treatment hyperplanes in the region of pertinent values of the covariates (hereafter called "the region of interest"), estimates were made of the differences between the current units for Treatment II and Treatment I at extreme values of the covariates. These estimates are presented in Table V-12. They indicate that the Treatment I hyperplane lies below the Treatment II hyperplane over almost the entire region of interest, with the exception of the "high corner" where the two stronger covariates, arithmetic and attitude, are at their maximums. It may be concluded that Treatment II is significantly

TABLE V-12

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS I AND II AT EXTREMES OF THE REGION OF INTEREST, FULL ANCOVA MODEL: FIRST EXPERIMENT

Covariate Values			Current Unit Difference Between Treatments II and I[a]
Arithmetic	Algebra	Attitude	
25	0	0	26.0
60	24	8	13.9
25	8	76	12.2
80	52	0	7.5
72	31	33	6.1
60	24	71	3.9
100	64	0	1.5
60	48	71	0.5
80	16	71	-0.5
80	52	71	-4.5
100	8	76	-5.1
100	64	76	-11.4

[a]Positive values indicate that the current unit for Treatment II is greater than the current unit for Treatment I, and vice versa.

better than Treatment I except for students with high scores on both the arithmetic test and the attitude questionnaire.

Comparison of the Treatments I and III hyperplanes yielded somewhat different results. Estimates of the differences between current units for these two treatments at critical values of the covariates are presented in Table V-13. These estimates indicate that there were more interactions between the covariates and these two treatments than for Treatments I and II. It may nonetheless be deduced from the estimates that Treatment III was superior to Treatment I over most of the region of interest, except for the region of high arithmetic scores paired with lower attitude scores.

Finally, estimates were made of the differences between the Treatments II and III hyperplanes. These estimates are presented in Table V-14. They indicate that in the region around the mean values of the covariates, the two treatments are equivalent. However, they also indicate that a strong interaction existed between the two treatments and the attitude covariate: students with lower attitude scores tended to perform better in Treatment II and students with higher attitude scores tended to perform better in Treatment III.

TABLE V-13

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS I AND III AT THE EXTREMES OF THE REGION OF INTEREST, FULL ANCOVA MODEL: FIRST EXPERIMENT

Covariate Values			Difference between Current Units for Treatments III and I[a]
Arithmetic	Algebra	Attitude	
25	8	76	37.2
60	24	71	19.8
25	8	0	19.7
80	24	71	9.0
100	64	76	5.8
72	31	33	5.7
60	0	8	1.3
80	52	0	-2.9
100	8	76	-3.3
80	36	0	-5.4
100	64	0	-11.7
100	0	0	-22.1

[a]Positive values indicate that the current unit for Treatment III is greater than the current unit for Treatment I, and vice versa.

TABLE V-14

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS II AND III AT CRITICAL POINTS OF THE REGION OF INTEREST, FULL ANCOVA MODEL: FIRST EXPERIMENT

Covariate Values			Current Unit Differences between Treatments III and II[a]
Arithmetic	Algebra	Attitude	
40	16	71	20.6
80	44	71	15.9
80	16	71	8.2
35	16	36	8.1
50	44	0	2.6
60	0	8	1.7
30	32	0	1.1
50	48	18	.4
50	0	18	- .2
72	31	33	- .4
40	0	24	-1.6
35	48	36	-2.9
50	0	18	-8.1
80	52	0	-10.4
60	16	8	-10.8
80	44	0	-12.5
80	16	0	-20.2

[a]Positive values indicate that the current unit for Treatment III is greater than the current unit for Treatment II, and vice versa.

This interaction between attitude and Treatments II and III is further illustrated by the positions of these two treatments relative to Treatment I: Treatment II was essentially equivalent to Treatment I for high arithmetic scores and high attitude scores, while Treatment III was essentially equivalent to Treatment I for high arithmetic scores and low attitude scores.

Analysis of covariance with two covariates. The possibility that the above analysis of covariance with three covariates was based on a biased sample because of the loss of 29 partially completed attitude questionnaires was examined by conducting a second analysis of covariance. This second analysis used only the arithmetic and algebra covariates, for which there were only six missing values. The sample size was thus reduced from 186 to 180. There was little reason to believe that this reduced sample was seriously biased since the number of missing values was so low and the missing values themselves were apparently randomly distributed across the treatments.

The test for equal slopes in this model yielded $F(10,162) = 1.48$, $p=.15$. Since this indicated that there was an 85% probability that at least one of the coefficients for one of the covariates was different from the others, the full model was retained as the more accurate model. Therefore, the same procedures were followed for this analysis as for the preceding analysis. Overall results of the analysis are presented in Table V-15; the overall $F(17,162) = 2.68$ with $p<.001$. Least-squares means are presented in Table V-16 and test results for the pre-planned comparisons are

TABLE V-15

OVERALL RESULTS OF THE ANALYSIS OF COVARIANCE USING ONLY ARITHMETIC AND ALGEBRA COVARIATES: FIRST EXPERIMENT

Source	df	Type III SS	MS	F	p	R^2
Model	17	1146.6	67.4	2.68	<.01	.22
Error	162	4082.6	25.2			
Corrected Total	179	5229.2				
Instructor (I)	1	.5		.02	.89	
Treatment (T)	2	107.8		2.14	.12	
I x T	2	85.4		1.69	.19	
Arithmetic (AR)	1	33.2		1.32	.25	
Algebra (AL)	1	246.0		9.76	<.01	
AR x I	1	15.9		.63	.43	
AL x I	1	12.0		.47	.49	
AR x T	2	119.5		2.37	.10	
AL x T	2	34.3		.68	.51	
AR x I x T	2	171.6		3.41	.04	
AL x I x T	2	36.3		.72	.49	

TABLE V-16

LEAST-SQUARES MEANS FOR THE ANALYSIS OF COVARIANCE WITH ARITHMETIC AND ALGEBRA COVARIATES: FIRST EXPERIMENT

Instructor	Treatment	Least-Squares Mean
1		9.58
2		8.18
	I	7.35
	II	9.57
	III	9.72
1	I	8.37
	II	10.01
	III	10.36
2	I	6.34
	II	9.12
	III	9.08

listed in Table V-17. These results are all in concurrence with the results of the ANCOVA with three covariates.

In addition, a clearer picture is presented by the estimates of the differences between the planes for the treatments. Estimates for Treatments I and II, presented in Table V-18, indicate that only for the relatively rare cases of high arithmetic and high algebra scores are the two treatments essentially equivalent. Over the rest of the region of interest Treatment II is superior. The estimates for Treatments I and III, listed in Table V-19, indicate that Treatment I may be superior to Treatment III for high arithmetic scores coupled with low algebra scores, and the treatments may be equivalent when both arithmetic and algebra scores are high. The estimates for Treatments II and III, presented in Table V-20, indicate that there were interactions with both arithmetic and algebra covariates: Treatment II was superior for higher arithmetic scores paired with lower algebra scores, while in general Treatment III was superior for lower arithmetic scores. The situation is illustrated in Figure V-2.

Analysis of covariance using the pooled Treatment II/III. Because of the number of similarities found between Treatments II and III in the above analyses, another model was considered in which Treatment II and Treatment III were pooled into one treatment, hereafter referred to as Treatment II/III. The test for equal slopes in this model yielded $F(9,130) = .11$, so the full model was retained. Results are presented in Table V-21. In this model the difference

TABLE V-17

TEST RESULTS FOR PRE-PLANNED COMPARISONS USING THE ANCOVA MODEL WITH ARITHMETIC AND ALGEBRA COVARIATES: FIRST EXPERIMENT

Contrast	df	SS	F	p
Treatment I vs. II	1	148.77	5.90	.02
Treatment I vs. III	1	133.70	5.31	.02
Treatment II vs. III	1	2.21	.09	.77
Treatment I vs. II & III	1	184.64	7.33	<.01
Treatments I & II vs. III	1	56.26	2.23	.14

TABLE V-18

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS I AND II OF THE EXTREMES OF THE REGION OF INTEREST, USING THE ANCOVA MODEL WITH ARITHMETIC AND ALGEBRA COVARIATES: FIRST EXPERIMENT

Covariate Values		Difference between Current Units for Treatments II and I[a]
Arithmetic	Algebra	
25	0	11.22
25	72	6.15
100	0	4.04
100	52	.38

[a]Positive values indicate that the current unit for Treatment II is greater than the current unit for Treatment I, and vice versa.

TABLE V-19

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS I AND III AT THE EXTREMES OF THE REGION OF INTEREST, ANCOVA MODEL WITH ARITHMETIC AND ALGEBRA COVARIATES: FIRST EXPERIMENT

Covariate Values		Differences between Current Units for Treatments III and I[a]
Arithmetic	Algebra	
25	52	28.41
25	8	19.89
100	52	- 2.23
100	8	-10.75

[a]Positive values indicate that the current unit for Treatment III is greater than the current unit for Treatment I and vice versa.

TABLE V-20

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS II AND III AT SELECTED CRITICAL VALUES OF THE COVARIATES, ANCOVA MODEL WITH ARITHMETIC AND ALGEBRA COVARIATES: FIRST EXPERIMENT

Covariate Values		Differences between Current Unit for Treatments III and II[a]
Arithmetic	Algebra	
40	24	8.8
25	0	7.1
60	36	5.7
60	24	2.5
80	44	1.5
72	31	.6
100	64	.6
60	16	.4
80	36	- .6
100	48	-3.7
80	16	-5.9
100	16	-12.1

[a]Positive values indicate that the current unit for Treatment III is greater than the current unit for Treatment II, and vice versa.

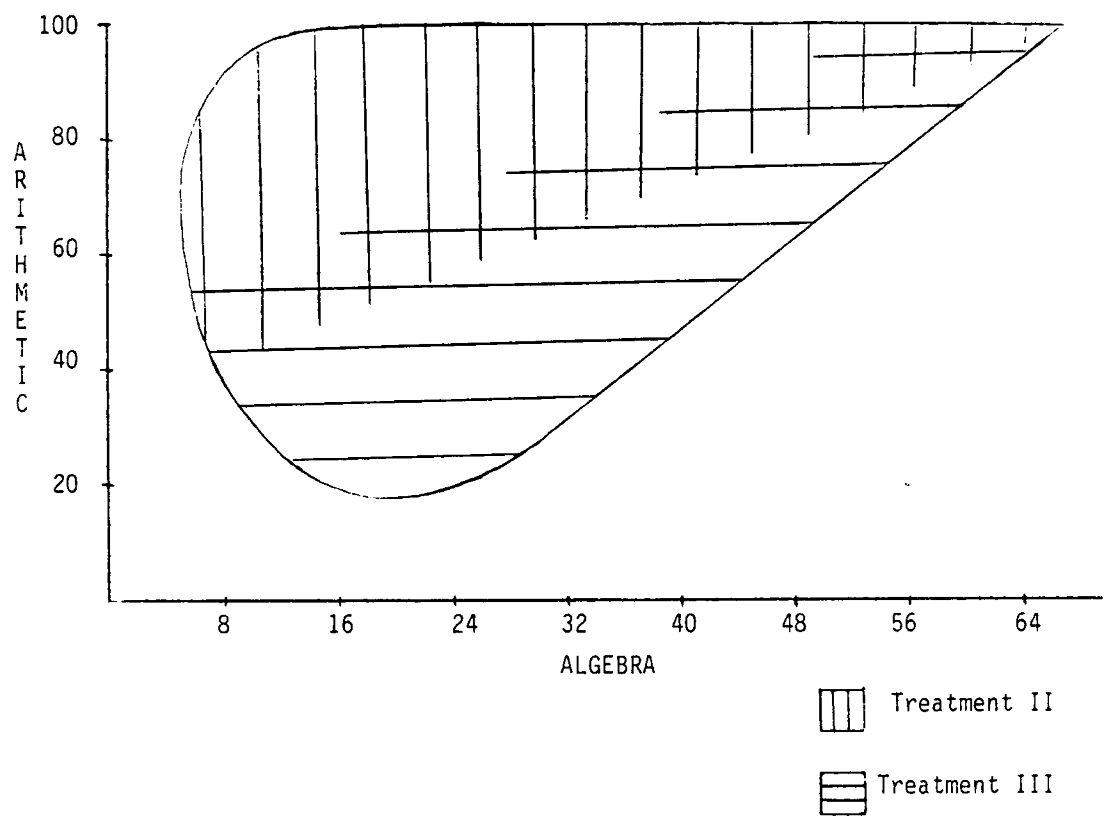

FIGURE V-2. INTERACTIONS BETWEEN TREATMENT II AND TREATMENT III AND THE ARITHMETIC AND ALGEBRA COVARIATES, ILLUSTRATED IN THE ARITHMETIC/ALGEBRA PLANE: FIRST EXPERIMENT

TABLE V-21

OVERALL RESULTS OF THE ANCOVA USING THE POOLED TREATMENT II/III: FIRST EXPERIMENT

Source	df	Type III SS	MS	F	p	R^2
Model	15	1014.1	67.61	2.96	<.01	.25
Error	130	2966.3	22.82			
Corrected Total	145	3980.4				
Instructor (I)	1	44.5		1.95	.17	
Treatments (T)	1	90.9		3.99	.05	
I x M	1	42.6		1.87	.17	
Arithmetic (AR)	1	37.7		1.65	.20	
Algebra (AL)	1	157.5		6.90	.01	
Attitude (ATT)	1	72.9		3.20	.08	
AR x I	1	66.3		2.91	.09	
AL x I	1	37.6		1.65	.20	
ATT x I	1	48.5		2.13	.15	
AR x M	1	72.5		3.18	.08	
AL x M	1	.1		0.01	.94	
ATT x M	1	.2		0.01	.93	
AR x I x M	1	101.2		4.43	.04	
AL x I x M	1	10.9		.48	.49	
ATT x I x M	1	16.5		.72	.40	

between the current units for Treatment I and Treatment II/III at the covariate means was found to be significant, $F(1.130) = 7.12$, $p<.01$. Estimates of the differences between the two planes at extremes of the region of interest are presented in Table V-22. These estimates indicate very strongly that Treatment II/III is superior to Treatment I for students with low arithmetic scores, while the treatments are approximately equivalent for students with high arithmetic scores. Since the two treatments are significantly different from each other at the covariate means, it may be deduced that Treatment II/III is significantly better than Treatment I over most of the region of interest except for the area of high arithmetic scores.

Success Rates

Another measure of student performance which was examined in this experiment was the percent of students completing or almost completing the course by the end of Fall Semester. This percent, referred to as the success rate, was defined as the percent of students either reaching or surpassing Unit 14, the last unit in the course. This definition included not only students who had completed all requirements for the course but also students who had made enough progress to be eligible for the grade of Incomplete. The decision to include the latter students in the success rate was based on the observation that students in Treatments II and III were more likely to qualify for the Incomplete grade than were students in Treatment I, and thus had less incentive to complete the course. This situation arose because of the different grading systems for

TABLE V-22

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS I AND II/III AT THE EXTREMES OF THE REGION OF INTEREST: FIRST EXPERIMENT

Covariate Values			Difference between Current Units
Arithmetic	Algebra	Attitude	for Treatments II/III and I[a]
25	52	76	20.17
25	8	76	19.47
25	52	2	19.20
25	8	2	18.51
100	8	2	2.76
100	52	76	-2.43
100	8	76	-3.13
100	52	2	-3.39

[a]Positive values indicate that the current unit for Treatment II/III is greater than the current unit for Treatment I, and vice versa.

the treatments (listed in Tables IV-2 and IV-3 on pages 92 and 93 respectively). Students in Treatment I were more likely to have had too many absences to qualify for the Incomplete grade than were students in Treatments II and III to have had too few points. The fact that only one Treatment I student was found in Units 14 and 15, in contrast to seven Treatment II and III students (see Table V-2, page 103) attests to this situation. It was thus the researcher's opinion that the best measure of success was the percent of students in Unit 14 and above.

Full sample. The success rates for the full sample are presented in Table V-23. The chi-square test for this contingency table was not significant, but it is interesting to note that there was a 10% difference between the success rate for Treatment I and the average success rate for Treatments II and III. Note, however, that this is entirely due to the poor performance of one cell, the Instructor 2/Treatment I combination, in which the success rate of 7.7% was almost 20% below the next higher success rate.

Reduced ANCOVA sample. Success rates were also computed for the reduced sample of 146 students used for the full analysis of covariance for the purpose of comparing these rates with rates for the full sample. These success rates are presented in Table V-24. Again, the chi-square test was not significant. In addition, a comparison of these rates with the rates for the full sample listed in Table V-23 reveal that no major differences existed between these two samples on this measure.

Pooled treatments. Another analysis of success rates was conducted after pooling Treatments II and III. Use of the pooled Treatment II/III, which was justified on the grounds that so few differences were detected in the analyses of covariance, resulted in increasing the power of the chi-square test. The success rates for Treatments I and II/III are presented in Table V-25. The chi-square test for this reduced contingency table was significant with $p=.08$. Inspection of the cell rates indicates that this is completely

TABLE V-23

SUCCESS RATES: PERCENT AND NUMBER OF STUDENTS IN UNITS 14-16, FIRST EXPERIMENT

Instructor	Treatment I	Treatment II	Treatment III	Total
1	34.5% (10)	27.3% (12)	34.6% (9)	31.3% (31)
2	7.7% (2)	35.2% (13)	29.2% (7)	25.3% (22)
Totals	21.8% (12)	30.9% (25)	32.0% (16)	28.5% (53)

$X^2 = 7.35$, $p=.20$.

TABLE V-24

SUCCESS RATES FOR REDUCED ANCOVA SAMPLE: PERCENT AND NUMBER OF STUDENTS IN UNITS 14-16 , FIRST EXPERIMENT

Instructor	Treatment I	Treatment II	Treatment III	Total
1	33.3% (7)	26.5% (9)	40.9% (9)	32.5% (25)
2	10.0% (2)	32.1% (9)	23.8% (5)	23.2% (16)
Totals	22.0% (9)	29.0% (18)	32.6% (14)	28.1% (41)

$X^2 = 5.78$, $p=.33$.

TABLE V-25

SUCCESS RATES FOR THE FULL SAMPLE USING THE POOLED TREATMENT II/III: PERCENT AND NUMBER OF STUDENTS IN UNITS 14-16, FIRST EXPERIMENT

Instructor	Treatment I	Treatment II/III	Total
1	34.5% (10)	30.0% (21)	31.3% (31)
2	7.7% (2)	32.8% (20)	25.3% (72)
Totals	21.8% (12)	31.3% (41)	28.5% (53)

$X^2 = 6.66$, $p=.08$.

due to the low rate in the Instructor 2/Treatment I combination.

Partial Completion Rates

Students in Mathematics 107 were not under heavy duress to finish the course in one semester since, except in extreme cases, they were permitted to continue the course in the next semester. It was thus of interest to examine not only the success rates but also the partial completion rates, defined to be the percent of students who reached at least Unit 10 by the end of Fall Semester. Unit 10 was chosen for this definition because it was part of the criteria for receiving the non-penalty grade of No Credit (NC).

Full sample. Partial completion rates for the full sample are presented in Table V-26. The chi-square test was not significant.

TABLE V-26

PARTIAL COMPLETION RATES: PERCENT AND NUMBER OF STUDENTS IN UNITS 10-16, FIRST EXPERIMENT

Instructor	Treatment I	II	III	Total
1	51.7% (15)	61.4% (27)	57.7% (15)	57.6% (57)
2	30.8% (8)	56.8% (21)	50.0% (12)	47.1% (41)
Totals	41.8% (23)	59.3% (48)	54.0% (27)	52.7% (98)

$X^2 = 6.93$, p=.23.

It is interesting to note, however, that the average partial completion rate for Treatment I (41.8%) is more than 15% below the average rate for Treatments II and III (57.3%). Examination of the cell rates indicates that this is mainly due to the extremely low partial completion rate for the Instructor 2/Treatment I combination, although the rate for the Instructor 1/Treatment I combination is also below the other rates by 6% to 10%.

Reduced ANCOVA sample. Partial completion rates for the reduced ANCOVA sample of 146 students were also computed and are presented in Table V-27. While the chi-square test is still not significant, its p-value of .12 is relatively low. Inspection of the rates indicates that the average rate for Treatment I was 14% below the average rate for Treatment III and 23% below the average rate for

TABLE V-27

PARTIAL COMPLETION RATES FOR REDUCED ANCOVA SAMPLE:
PERCENT AND NUMBER OF STUDENTS IN UNITS 10-16,
FIRST EXPERIMENT

Instructor	Treatment I	Treatment II	Treatment III	Total
1	52.4% (11)	67.7% (23)	63.6% (14)	62.3% (48)
2	30.0% (6)	60.7% (17)	47.6% (10)	47.8% (33)
Totals	41.5% (17)	64.5% (40)	55.8% (24)	55.5% (81)

$X^2 = 8.81$, $p=.12$.

Treatment II. Further inspection of the individual cell rates shows that the Treatment I rates for both instructors were from 11% to 37% below the Treatment II and Treatment III rates for both instructors.

Comparison of Tables V-26 and V-27 indicates that increases in partial completion rates occurred in several of the cells of the reduced sample, increases which especially favored Treatment II and to a lesser extent Treatment III. This suggests that the reduced sample may have been somewhat biased.

Pooled treatments. The partial completion rates for Treatments I and II/III are presented in Table V-28. The Chi-square test for this smaller contingency table is significant, $p=.09$. Again, inspection of the cell rates indicates that there is one abnormal rate, namely that for the Instructor 2/Treatment I combination.

TABLE V-28

PARTIAL COMPLETION RATES USING THE POOLED TREATMENT II/III: PERCENT AND NUMBER OF STUDENTS IN UNITS 10-16, FIRST EXPERIMENT

Instructor	Treatment I	Treatment II/III	Total
1	51.7% (15)	60.0% (42)	57.6% (57)
2	30.8% (8)	54.1% (33)	47.1% (41)
Totals	41.8% (23)	57.3% (75)	52.7% (98)

$X^2 = 6.57$, $p=.09$.

Attrition Rates

A fourth measure of student performance examined in this study was the rate of attrition. Attrition rate is often defined to be the number of students who officially drop a course divided by the total number of students receiving a grade in the course. However, a different definition was considered more accurate for this study. Attrition rate was defined to be the number of students included in the experiment who did not take any tests in the course after the official withdrawal deadline, divided by the total number of students in the experiment. (The withdrawal deadline occurred during the ninth week of classes.) This definition of attrition rate was chosen so as to include students who stopped taking tests by this time even though they did not officially withdraw from the course, and also to include students who withdrew so quickly that they did not receive any grade in the course. Note, however, that students who either never attended the course or attended only the first day of class are not included in this definition of attrition since they were excluded from the experiment regardless of the grades they may have received in the course.

Full sample. The attrition rates for the full sample are presented in Table V-29. The chi-square test was not significant. It is nonetheless interesting to observe that a pattern occurs among the cell rates: the two Treatment II cell rates were the two lowest rates in the table. In fact the Instructor 1/Treatment II rate of 22.7% was 11% to 31% below the rates in the four cells for the other two treatments. Conversely, the Instructor 2/Treatment I cell had an attrition rate of 53.9% which was much higher, practically speaking, than the other rates, by a margin of 19% to 31%.

Reduced ANCOVA sample. The attrition rates for the reduced sample used for the analysis of covariance are presented in Table V-30. The chi-square test is not significant. These rates are approximately the same as the rates for the full sample listed in Table V-29, except that the difference between the lowest rate in the Instructor I/Treatment II cell and the non-Treatment II cells becomes more marked.

TABLE V-29

ATTRITION RATES: PERCENT AND NUMBER OF STUDENTS TAKING NO TESTS AFTER THE WITHDRAWAL DEADLINE, FIRST EXPERIMENT

Instructor	Treatment I	Treatment II	Treatment III	Total
1	34.5% (10)	22.7% (10)	34.6% (9)	29.3% (29)
2	53.9% (14)	29.7% (11)	33.3% (8)	37.9% (33)
Totals	43.6% (24)	25.9% (21)	34.0% (17)	33.3% (62)

$X^2 = 7.40$, $p=.19$.

TABLE V-30

ATTRITION RATES FOR THE REDUCED ANCOVA SAMPLE: PERCENT AND NUMBER OF STUDENTS TAKING NO TESTS AFTER THE WITHDRAWAL DEADLINE, FIRST EXPERIMENT

Instructor	Treatment I	Treatment II	Treatment III	Total
1	33.3% (7)	14.7% (5)	31.8% (7)	24.7% (19)
2	50.0% (10)	28.6% (8)	33.3% (7)	36.2% (25)
Totals	41.5% (17)	21.0% (13)	32.6% (14)	30.1% (44)

$X^2 = 7.86$, $p=.16$.

Pooled treatments. The attrition rates for Treatment I and the pooled Treatment II/III are presented in Table V-31. The chi-square test for this smaller table is significant, p=.10. The overall Treatment II/III rate of 29.0% is almost 15% below the Treatment I attrition rate. Examination of the individual cell rates indicates that this is primarily due to the abnormally high attrition rate in the Instructor 2/Treatment I cell, but even so, the Instructor 1/Treatment I cell rate is also below the corresponding Treatment II/III rate.

3. DATA FROM THE SECOND EXPERIMENT

The Sample

A replication of the experiment was performed at the noon period with a sample of 160 students. This sample was obtained in the same manner as the sample for the first experiment. Of the 187 students originally assigned to the six cells of the 2 x 3 (instructor by treatment) factorial design, one student was dropped from the study because of a conflict with the Treatment III Friday class, and 26 were dropped because they either never attended class or attended only the first day of class.

Course Progress

Descriptive statistics. A histogram for the current unit for the noon sample is presented in Figure V-3. Means and standard errors of the current unit for each instructor, treatment and

TABLE V-31

ATTRITION RATES USING THE POOLED TREATMENT II/III:
PERCENT AND NUMBER OF STUDENTS TAKING NO TESTS
AFTER THE WITHDRAWAL DEADLINE,
FIRST EXPERIMENT

Instructor	Treatment I	Treatment II/III	Total
1	34.5% (10)	27.1% (19)	29.3% (29)
2	53.9% (14)	31.2% (19)	37.9% (33)
Total	43.6% (24)	29.0% (38)	33.3% (62)

$X^2 = 6.28$, $p = .10$.

instructor/treatment combination are listed in Table V-32. The number of students in each unit for each combination is tabulated in Table V-33 and percents are provided for certain unit intervals in Table V-34.

Analysis of variance. The results of the analysis of variance for an unbalanced design are presented in Table V-35. The overall F was not significant, $F(5,154) = .38$, $p=.87$, so no tests of preplanned comparisons were conducted.

Analysis of covariance with three covariates. For the analysis of covariance the three covariates of arithmetic, algebra and mathematics attitude were added to the model. The correlations among

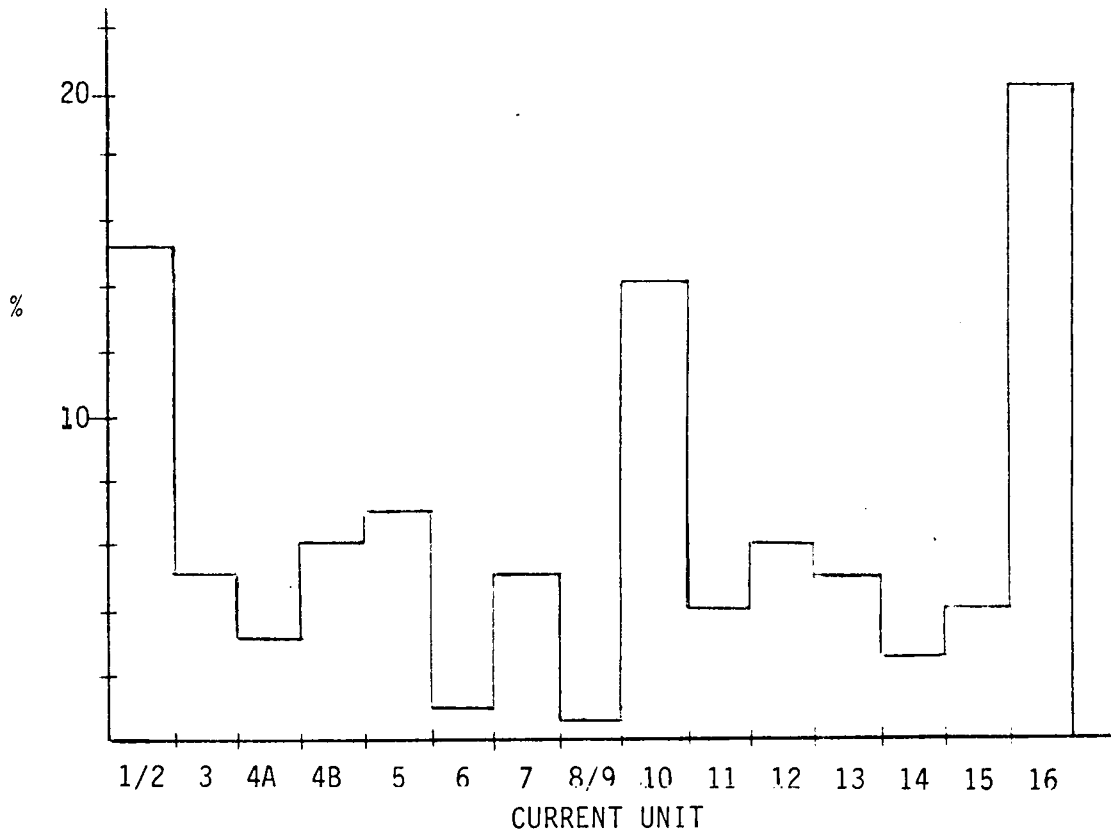

FIGURE V-3. HISTOGRAM FOR CURRENT UNIT: SECOND EXPERIMENT

TABLE V-32

MEANS AND STANDARD ERRORS FOR THE VARIABLE OF CURRENT UNIT: SECOND EXPERIMENT

Instructor	Treatment	Number	Mean	Standard Error
1		78	8.81	.63
2		82	8.98	.60
	I	50	8.44	.75
	II	61	9.25	.68
	III	49	8.92	.84
1	I	24	8.96	1.12
	II	30	8.57	1.04
	III	24	8.96	1.17
2	I	26	7.96	1.03
	II	31	9.90	.89
	III	25	8.88	1.23
Total		160	8.89	.43

TABLE V-33

NUMBER OF STUDENTS IN EACH INSTRUCTOR X TREATMENT COMBINATION BY UNIT: SECOND EXPERIMENT

	Instructor 1			Instructor 2			
	Treatment			Treatment			
Unit	I	II	III	I	II	III	Total
1 and 2	2	5	5	4	3	5	24
3	1	2	1	2	1	1	8
4A	2	3	0	0	0	0	5
4B	2	1	0	3	2	2	10
5	3	0	1	2	2	3	11
6	0	1	0	0	2	0	3
7	0	1	4	2	0	1	8
8 and 9	1	0	1	0	0	0	2
10	4	4	2	5	4	3	22
11	1	2	0	1	2	0	6
12	0	3	2	1	4	0	10
13	1	0	1	2	4	0	8
14	1	1	1	0	0	1	4
15 (lacking final)	0	2	0	0	2	3	7
16 (completed course)	6	5	6	4	5	6	32
Totals	24	30	24	26	31	25	160

TABLE V-34

PERCENT OF STUDENTS IN EACH INSTRUCTOR X TREATMENT COMBINATION, BY SELECTED UNIT INTERVALS: SECOND EXPERIMENT

	Instructor 1			Instructor 2			
Unit	Treatment			Treatment			
Interval	I	II	III	I	II	III	Total
1-5	42%(10)	37%(11)	29%(7)	42%(11)	26%(8)	44%(11)	37%
6-9	4%(1)	7%(2)	21%(5)	8%(2)	6%(2)	4%(1)	8%
10-13	25%(6)	30%(9)	21%(5)	35%(9)	45%(14)	12%(3)	29%
14-16	29%(7)	27%(8)	29%(7)	15%(4)	23%(7)	40%(10)	27%
Totals	24	30	24	26	31	25	160

TABLE V-35

ANALYSIS OF VARIANCE: SECOND EXPERIMENT

Source	Type III SS	df	F	p
Instructor	.30	1	.01	.92
Treatment	16.52	2	.27	.76
Instructor x Treatment	38.49	2	.63	.54
Error	4719.59	154		
Total	4777.19			

these variables and the dependent variable of current unit are presented in Table V-36. Unlike the correlations for these variables in the first experiment, both arithmetic and algebra are moderately correlated with current unit (.58 and .48, respectively) but attitude is not (.17). Other descriptive statistics for the covariates are presented in Table V-37, including means, standard errors, minimums and maximums.

The use of the expanded covariance model reduced the sample size from 160 to 147. The sources of the missing values were seven students who did not correctly complete their attitude questionnaires and six students who had no arithmetic and algebra scores. Since the number of missing values was relatively small, it was assumed that the reduced sample was not seriously biased by their loss.

The test of the parallelism hypothesis resulted in $F(15,123) = 1.51$, $p=.11$. Since this indicated an 89% probability that at least one of the coefficients for one of the covariates was different from the others, the full model was retained for the remaining analyses. Overall results are presented in Table V-38.

The possibility that the data would be better fitted by a quadratic model was also investigated. The results of an analysis of covariance using a model with quadratic covariate terms are presented in Table V-39. Since none of the tests of the null hypotheses for the quadratic coefficients were significant, the linear model was retained.

The behavior of the three treatment hyperplanes was first examined by obtaining the least squares means, which are presented

TABLE V-36

CORRELATION COEFFICIENTS FOR THE COVARIATES
AND THE CURRENT UNIT:
SECOND EXPERIMENT

Variables	1	2	3	4
1. Unit	--	.58*	.48*	.17**
2. Arithmetic		--	.39*	.18**
3. Algebra			--	.25*
4. Attitude				--

*Significant at α = .01

**Significant at α = .05

TABLE V-37

MEANS, STANDARD ERRORS, MINIMUMS AND MAXIMUMS FOR THE THREE COVARIATES BY TREATMENT AND INSTRUCTOR X TREATMENT COMBINATIONS: SECOND EXPERIMENT

		Arithmetic				Algebra				Attitude			
Instr	Treatment	Mean	S.E.	Min	Max	Mean	S.E.	Min	Max	Mean	S.E.	Min	Max
	I	73.0	2.32	31	97	32.7	1.64	8	52	27.0	2.20	3	59
	II	72.7	2.09	25	100	30.7	1.41	8	52	34.8	2.01	4	76
	III	70.0	2.65	31	100	30.8	1.78	12	64	35.8	2.21	2	67
1	I	72.7	3.58	42	97	34.6	2.08	20	52	25.0	3.27	3	54
	II	72.6	3.41	25	97	27.0	1.93	8	44	33.4	2.75	4	60
	III	68.0	4.09	31	100	30.6	2.44	12	64	33.5	3.16	2	63
2	I	73.4	3.03	31	88	30.8	2.51	8	52	28.6	3.00	4	59
	II	72.9	2.54	34	100	34.1	1.88	16	52	36.2	2.95	4	76
	III	72.0	3.43	34	97	30.8	2.63	12	52	37.9	3.09	6	67
Total		72.0	1.35	25	100	31.0	.92	8	64	32.8	1.26	2	76

TABLE V-38

ANCOVA RESULTS, FULL LINEAR MODEL: SECOND EXPERIMENT

Source	df	Type III SS	MS	F	p	R^2
Model	23	2203.1	95.79	5.40	<.01	.50
Error	123	2179.9	17.72			
Corrected Total	146	4382.9				
Instructor (I)	1	43.1		2.43	.12	
Treatments (T)	2	0.9		.02	.98	
I x T	2	61.7		1.74	.18	
Arithmetic (AR)	1	552.2		31.16	<.01	
Algebra (AL)	1	356.6		20.12	<.01	
Attitude (ATT)	1	0.8		.04	.84	
AR x I	1	130.8		7.38	<.01	
AL x I	1	127.3		7.19	<.01	
ATT x I	1	33.3		1.88	.17	
AR x T	2	45.0		1.27	.28	
AL x T	2	55.3		1.56	.21	
ATT x T	2	20.8		.59	.56	
AR x I x T	2	46.0		1.30	.28	
AL x I x T	2	168.0		4.74	.01	
ATT x I x T	2	23.1		.65	.52	

TABLE V-39

RESULTS OF THE ANCOVA WITH QUADRATIC TERMS: SECOND EXPERIMENT

Source	df	Type III SS	F	p
Cell means (CM)	6	3460.6	31.46	<.01
AR x CM	6	654.2	5.95	<.01
AL x CM	6	460.9	4.19	<.01
ATT x CM	6	42.8	.39	.88
AR^2 x CM	6	137.7	1.25	.29
AL^2 x CM	6	67.4	.61	.72
ATT^2 x CM	6	46.6	.42	.86

in Table V-40. There is a clear pattern in the least-squares means for the six instructor/treatment combinations: the two means for Treatment I are both below the means for the other two treatments by a margin of 1.0 to 2.18 units, while the means for Treatments II and III are relatively close together. This pattern did not occur among the raw means.

Next the tests of the five pre-planned comparisons were conducted using the mean covariate values. The results of these tests are presented in Table V-41. At α = .10 the following conclusions can be made about the treatments at the means of the covariates:

1. The current unit for Treatment II is significantly higher than the current unit for Treatment I (p=.08).
2. The average current unit for Treatments II and III is significantly higher than the current unit for Treatment

TABLE V-40

LEAST-SQUARES MEANS FOR THE ANALYSIS OF COVARIANCE: SECOND EXPERIMENT

Instructor	Treatment	Least Squares Means
1		8.83
2		9.05
	I	7.85
	II	9.55
	III	9.44
1	I	7.46
	II	9.40
	III	9.64
2	I	8.24
	II	9.69
	III	9.24

TABLE V-41

RESULTS OF TESTS FOR PRE-PLANNED COMPARISONS USING THE FULL ANALYSIS OF COVARIANCE MODEL: SECOND EXPERIMENT

Source	df	SS	F	p
Treatment I vs. II	1	51.93	2.93	.09
Treatment II vs. III	1	.33	.02	.89
Treatment I vs. III	1	42.31	2.39	.12
Treatment I vs. II & III	1	58.03	3.27	.07
Treatments I & II vs. III	1	14.17	.80	.37

I (p=.07). It is also noteworthy that while the hypothesis of the equivalence of Treatments I and III could not be rejected, the p-value of this test was p=.12.

As in the case of the first experiment, the retention of the full linear model called for further investigation of the relative positions of the three treatment hyperplanes. Inspection revealed a more complicated situation than that discovered in the first experiment because all of the hyperplanes intersected each other within the region of interest. It was necessary to find large numbers of estimates in order to gain insight into the many interactions taking place. Some of the estimates found for the Treatments I and II hyperplanes are presented in Table V-42. These estimates indicate that over most of the region Treatment II was superior to or at least equivalent to Treatment I; one major exception was the region of low attitude scores paired with medium to high algebra scores. Investigation of the differences between the Treatments I and III hyperplanes indicated interactions as well: estimates presented in Table V-43 demonstrate that over much of the region of interest, Treatment III was superior to Treatment I, with one major exception being the area of high attitude scores paired with medium to high algebra scores.

Estimates of the differences between the Treatments II and III hyperplanes demonstrated the strong interaction between these two treatments and the attitude covariate that is implied by the

TABLE V-42

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS I AND II AT VALUES WITHIN THE REGION OF INTEREST: FULL ANCOVA, SECOND EXPERIMENT

Covariate Values			Differences between Current Units
Arithmetic	Algebra	Attitude	for Treatments II and I[a]
100	0	76	20.8
100	0	0	15.1
25	0	76	10.8
72	31	33	3.4
25	8	0	2.4
100	64	76	- .9
40	40	71	-1.1
80	44	0	-2.5
30	32	0	-5.0
80	52	0	-5.2
100	64	0	-6.6

[a]Positive values indicate that the current unit for Treatment II is greater than the current unit for Treatment I, and vice versa.

TABLE V-43

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS I AND III AT THE EXTREMES OF THE REGION OF INTEREST: FULL ANCOVA, SECOND EXPERIMENT

Covariate Values			Differences between Current Units
Arithmetic	Algebra	Attitude	for Treatments I and III[a]
100	8	0	17.9
100	8	76	14.0
72	31	33	3.1
25	8	0	2.9
100	64	0	-0.4
25	8	76	-1.0
60	36	71	-2.8
40	24	71	-2.9
100	64	76	-4.3

[a]Positive values indicate that the current unit for Treatment III is greater than the current unit for Treatment I and vice versa.

above relationships with the Treatment I hyperplane. For lower values of the attitude covariate, Treatment III was superior to or at least equivalent to Treatment II, but for higher values of the attitude covariate Treatment II was superior to Treatment III. Estimates for these two treatments, presented in Table V-44, are paired to illustrate this pattern: if arithmetic and algebra scores are held constant, the sign of the difference between the two planes is reversed from high to low attitude scores.

Success Rates

Success rates, defined as the percent of students who either reached or surpassed Unit 14, are presented in Table V-45, along with the associated chi-square test.

Inspection of the six success rates in the instructor/treatment cells shows that the success rates for all three treatments under Instructor 1 were very close to one another. This is not the case for the cells of Instructor 2: the rate for Treatment III under this instructor is 40%, which is approximately double the rates of 15% and 23% for the other two treatments. However, the overall chi-square test is not significant.

Partial Completion Rates

Partial completion rates, defined to be the percent of students who either reached or surpassed Unit 10, are presented in Table V-46, along with the associated chi-square test.

TABLE V-44

ESTIMATES OF THE DIFFERENCES BETWEEN CURRENT UNITS FOR TREATMENTS II AND III AT CRITICAL VALUES FOR THE REGION OF INTEREST: SECOND EXPERIMENT

Covariate Values			Differences between Current Unit for Treatments III and II[a]
Arithmetic	Algebra	Attitude	
30	32	0	1.1
30	32	71	-7.8
60	48	0	3.3
60	48	71	-5.6
80	52	0	4.7
80	52	71	-4.3
80	24	71	-4.6
72	31	33	- .4

[a]Positive values indicate that the current unit for Treatment III is greater than the current unit for Treatment II, and vice versa.

TABLE V-45

SUCCESS RATES: PERCENT AND NUMBER OF STUDENTS IN UNITS 14-16[a], SECOND EXPERIMENT

Instructor	Treatment I	Treatment II	Treatment III	Totals
1	29.2% (7)	26.7% (8)	29.2% (7)	28.2% (22/78)
2	15.4% (4)	22.3% (7)	40.0% (10)	25.6% (21/82)
Totals	22.0% (11)	24.6% (15)	34.7% (17)	26.9% (43)

$X^2 = 4.36$, $p=.50$

[a]Students in "Unit 15" are those who still need to pass the final exam; students in "Unit 16" have passed the final exam and thus completed the course.

TABLE V-46

PARTIAL COMPLETION RATES: NUMBER AND PERCENT OF STUDENTS IN UNITS 10-16, SECOND EXPERIMENT

Instructor	Treatment I	Treatment II	Treatment III	Total
1	54.2% (13)	56.7% (17)	50.0% (12)	53.8% (42)
2	50.0% (13)	67.7% (21)	52.0% (13)	57.3% (47)
Totals	52.0% (26)	62.3% (38)	51.0% (25)	55.6% (89)

$X^2 = 2.65$, $p=.75$

Inspection of the six partial completion rates in the instructor/treatment cells indicates that these rates for all three treatments under Instructor 1 are within 7% of each other, reflecting the same similarity found among the success rates for these cells. However, a different pattern emerged for the cells of Instructor 2: the partial completion rate for Treatment II was 16% to 18% above the partial completion rates for Treatments I and III, respectively. This is in contrast to the pattern of the success rates for these same cells, where the success rate for Treatment III was much higher than the success rates for Treatments I and II and the latter two rates were fairly close to each other. However, the differences in partial completion rates were not large enough to produce a significant chi-square test.

<u>Attrition Rates</u>

Attrition rates, defined as the percent of students who did not take any tests after the official withdrawal deadline, are presented in Table V-47, along with the associated chi-square test. These rates seem to be very close to one another, as indicated by a very non-significant p-value for the chi-square test.

TABLE V-47

ATTRITION RATES: PERCENT AND NUMBER OF STUDENTS TAKING NO TESTS AFTER WITHDRAWAL DEADLINE, SECOND EXPERIMENT

	Treatment			
Instructor	I	II	III	Total
1	33.3% (8)	36.7% (11)	29.2% (7)	33.3% (26)
2	34.6% (9)	22.6% (7)	28.0% (7)	28.0% (23)
Totals	34.0% (17)	29.5% (18)	28.6% (14)	30.6% (49)

$X^2 = 1.84$, $p=.87$

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

1. INTRODUCTION

The purpose of this study was to investigate two variables within the context of an individualized remedial algebra course. These variables were organizational structure and tutoring behaviors. Two types of organizational structure were investigated. The first was a lecture/workshop structure in which students alternately attended large lectures and smaller tutor-supervised workshops four periods per week; by the middle of the semester the majority of students attended workshop during all four periods because they had failed to keep pace with the lecture schedule. The second organizational structure was that of a class with the traditional size of 25 to 30 students, supervised primarily by an instructor.

The second variable under investigation was that of tutoring behaviors. Specifically, two levels of involvement between tutors and their students were studied. The first was less intensive, with tutors playing a more passive role, primarily that of answering questions. These tutors never saw their students' tests and did not regularly keep track of their students' progress. In this situation two tutors were assigned to 30 students. The second set of tutoring behaviors was more intensive. Tutors deliberately became more acquainted with each of their students, kept track of their test-taking activities and reviewed their tests almost daily, and interacted regularly with their students regarding their progress in the course. In this situation each tutor was specifically assigned to 15 students.

The experiments used three treatments to investigate these variables. Specifically, the experiments had the following purposes:

1. to compare the performance of students within a lecture/workshop structure (Treatments I and II) to the performance of students within a small class structure (Treatment III);
2. to compare the performance of students under more intensive tutoring conditions (Treatments II and III) to the performance of students under less intensive tutoring conditions (Treatment I).

Performance was measured by the number of units completed by the end of Fall Semester. To accomplish this purpose five pre-planned comparisons were tested:

1. Is Treatment I equivalent to Treatment II?
2. Is Treatment I equivalent to Treatment III?
3. Is Treatment II equivalent to Treatment III?
4. Is Treatment I equivalent to the average of Treatments II and III?
5. Is Treatment III equivalent to the average of Treatments I and II?

2. SUMMARY OF FINDINGS

Because the major findings were somewhat different for each of the two replications in this study, they are presented separately.

Findings from the First Experiment

After the data were adjusted for the three covariates of arithmetic, algebra and attitude toward mathematics, the major findings were as follows:

1. Treatment II was superior to Treatment I at the covariate means and over most of the possible values of the covariates, with the exception of the region of high arithmetic scores paired with high attitude scores.
2. Treatment III was superior to Treatment I at the covariate means and over many of the possible values of the covariates, with the exception of the region of high arithmetic scores paired with low attitude scores.
3. The hypothesis that Treatment II was equivalent to Treatment III at the covariate means could not be rejected. In fact these two treatments had almost equivalent values at the covariate means.
4. The pooled Treatment II/III was superior to Treatment I except over the region of high arithmetic scores and either high algebra scores or high attitude scores. (Treatments II and III were pooled in order to increase statistical power; pooling was justified by the probable equivalence of these treatments at the covariate means.)

5. The hypothesis that the mean of Treatment III was equivalent to the average of the means of Treatments I and II at the covariate means could not be rejected.

Because adjustment for the covariates made a significant difference in the relationship of the treatments to each other, as demonstrated by a comparison of raw cell means to adjusted cell means, the analyses of the other measures of performance (all based on the unadjusted data) were not in full accord with the above results. Nonetheless, for all three measures of success rates, partial completion rates and attrition rates, Treatment I was found to be significantly inferior to the pooled Treatment II/III. Inspection of the rates for individual cells of the design revealed that this was primarily due to the poor performance of students in only one of the two Treatment I cells, namely the cell for the second instructor. For the first instructor the rates for all three treatments on all three measures were relatively close to each other, a situation which is not surprising in light of the fact that the three unadjusted current unit means for the first instructor were also close to one another. However, the results of the analysis of covariance indicate that after adjustment was made for the covariates, the Treatment I cell for the first instructor became comparable to the Treatment I cell for the second instructor; that is, both cells ranked third in performance after the Treatments II and III cells. Inspection of covariate means for the cells reveals that the Treatment I cell for the first instructor had the highest mean algebra score and the second highest mean arithmetic score; the conclusion to be drawn from the analysis of covariance is that the students in Treatment I under the first instructor performed as well as students in the other two treatments only because of their better mathematical preparation.

In light of these results, it may be concluded that for the population of the first experiment:

1. the differences in organizational structure (lecture/workshop versus small class) did not have a significant effect on student performance;
2. the differences in tutoring behaviors and grading systems did have a significant effect on student performance.

Findings from the Second Experiment

The findings of the second experiment were not completely consistent with the findings for the first experiment, but there were several major similarities. After the data were adjusted for the covariates, the results were as follows:

1. Treatment II was significantly better than Treatment I at the covariate means and also apparently over most of the region of interest, except for the region of high arithmetic and high algebra scores.
2. The hypothesis that Treatment III was equivalent to Treatment I at the covariate means could not be rejected, although the relatively low p-value of .12 and comparisons of the treatments over other covariate values cast doubt on the similarity of the treatments.

3. The hypothesis that Treatment II was equivalent to Treatment III at the covariate means could not be rejected. In fact these two treatments had very similar values at the covariate means.

4. The average of Treatments II and III was superior to Treatment I at the covariate means.

5. The hypothesis that Treatment III was equivalent to the average of Treatments I and II at the covariate means could not be rejected.

The only major differences between the results of the first experiment and the results of the second experiment were that no significant difference was found between Treatment I and Treatment III, and more interactions occurred between the treatments and the covariates within the region of interest. It may be inferred that these differences arose from the fact that the experiments drew on different populations--it would seem that students who chose to take Mathematics 107 at the 8:00 A.M. period were somewhat different from students who chose to take the same course at the noon period.

The success rates, partial completion rates and attrition rates found in the second experiment did not reflect the results of the analysis of covariance. In fact the rates for all three treatments were very similar for Instructor 1. For Instructor 2 there were wider fluctuations among the three treatments, although not wide anough to be statistically significant. However, as in the case of the first experiment, even in the face of this apparently contradictory evidence, the results of the analysis of covariance prevail since they were based on adjusted data.

It may thus be concluded that for the population of the second experiment:

1. differences in organizational structure did not have a significant effect on student performance;
2. differences in tutoring behaviors and grading systems within the context of a lecture/workshop structure had a significant effect on student performance.

2. DISCUSSION

Preface

Before interpreting the findings of this study it is first obligatory to make several observations about the treatments tested in these experiments. Objectively speaking, none of the treatments was successful. The overall success rate was 28% and the highest success rate for a single cell was 40%. In comparison to completion rates reported in the literature for successful programs, these rates are low. While the differences found between treatments were of practical as well as theoretical significance, they were not large enough to be considered satisfactory. It must therefore be concluded that there were characteristics common to all the treatments that inhibited their success. Since such characteristics could easily have had an effect on the outcome of these experiments, it is relevant to speculate on what they may have been. There are several strong possibilities:

1. Many students who could have completed the course in one semester chose instead to extend the course into the following semester simply because this was the least demanding option. This conjecture is supported by the shape of the histograms for the current unit: there is a peak at Unit 10, the unit students had to reach to avoid the F grade, Further evidence is found in the distributions for previous semesters: when the minimal unit was Unit 8 and Unit 9 rather than Unit 10, the peak occurred at Unit 8 and Unit 9. (Units 8 and 9 were combined into one unit because they were both short.)
2. Many students lacked the necessary background to handle what was in essence a review of first year high school algebra. The histograms also support this conjecture: there is a large peak over the first few units, which review integer arithmetic, simplifying basic algebraic expressions, operations with fractions with integral denominators and solving basic linear equations.
3. The course may be more demanding than many other remedial algebra courses because it places relatively heavy emphasis on word problems. That students were impeded by the units containing the word problems is illustrated by the peak in the histogram at Unit 4B and Unit 5, the units which are heavily devoted to word problems.

In sum, this course is characterized by:

1. the heterogeneous student population, many of whom apparently need a lower level course;
2. the lack of an end-of-the-semester deadline for completing the course; and
3. the inclusion of relatively difficult material which hinders students' progress in the course.

How strongly these characteristics affect the extent to which the results of this experiment apply to other remedial mathematics programs is unknown, but certainly bears investigation.

Implications about Organizational Structure

In both experiments the failure to find a significant difference between Treatments II and III strongly indicates that when other conditions are held constant, a small class supervised by a faculty member is no more or less effective than a class organized around large lectures and tutor-supervised workshops. The only role the faculty member played in Treatment II was that of lecturer. Since many students started attending workshop instead of lecture about halfway through the semester, the faculty member had even less impact on these students. On the other hand, in Treatment III the faculty members' contact with their students was maximized: they spent part or all of the class period working with students individually, they monitored the progress of each of their students and discussed their progress charts with them regularly, and they came to know each of their students in Treatment III far better than the students to whom they lectured in Treatments I and II. Thus, from the faculty member's viewpoint the treatments were very different. However, from the standpoint especially of the students who attended workshop four days a week in Treatment II, the differences between Treatments II and III were apparently not of major consequence, at least in terms of their performance. In effect, the tutor in the workshop performed the same services equally well as the faculty member in the small classroom.

Thus it must be concluded that under the conditions in which this experiment was conducted, undergraduate tutors are just as effective (or just as ineffective) as faculty members in the individualized classroom. This is a very interesting result in light of the differences in cost of personnel! However, it must be cautioned that this conclusion was based on the equivalence of two treatments neither of which could be considered particularly effective. It is theoretically possible that under conditions which were more conducive to effective teaching methods, this conclusion would not apply.

As indicated in the review of literature, there is a dearth of studies comparing the relative effectiveness of different types of organizational structure in individualized mathematics programs. The study presented here was intended to fill a gap in the research in this field. It had been surmised by the experimenter that because most successful individualized mathematics programs were conducted within the structure of small classes conducted by faculty members, the latter structure might be causally related to the success of the program. Such does not appear to be the case, at least within the limitations of this study. Instead, the results indicate that programs such as those described by Jackson (1979), Steele et al. (1980), and Eisenberg and Brown (1973), which were all organized around tutor-supervised classes, are just as effective as programs in which the faculty member plays a more central role. They also indicate that the problems associated with less successful programs such as the one at The University of Tennessee at Chattanooga are not simply due to a large faculty to student ratio.

Implications about the Role of Tutors and Grading Systems

Both experiments also provide strong evidence to indicate that the different conditions in Treatment II made it significantly more effective than Treatment I. The two major differences between these treatments are confounded and therefore no conclusion can be made regarding how much either individually affected the outcome of the experiments. It is thus necessary to speculate as to why either may have improved student performance.

The Possible Effect of Grading Systems on Performance

Treatments I and II operated under different grading systems. In Treatment II the point system provided incentives for students to meet test deadlines; no such incentives were provided in Treatment I. There is considerable support in the literature for the claim that incentives such as these can increase student performance. For instance, Riedel et al. (1976) reported that completion rates jumped from 51% to 82% when bonus points were established for meeting test deadlines. In addition, many reviewers, including Reiser (1976), Kulik et al. (1978) and Robin (1976), have concluded that research supports the hypothesis that positive incentives for progress reduce procrastination in individualized programs.

It may be further speculated that the point system in Mathematics was not as effective as it might have been because it was tied to an S/NC/F/FF grading system rather than the traditional A through F system. The effect of accumulating points was therefore not as gratifying.

The Possible Effect of High Intensity Tutor Behaviors on Performance

The second major difference between Treatment I and Treatment II was the level of involvement between tutors and their students. In Treatment II the tutors became acquainted with their students more quickly and made more direct inquiries about their students' lives outside of the classroom. In addition, through the vehicle of the progress charts the tutors were more likely to be conscious of the regularity with which their students were taking tests, and thus more likely to offer specific encouragement to students who were procrastinating. The students thus may have perceived their tutors as being interested both in their mathematical progess and in themselves as individuals. There is some support in the literature for the claim that students who are assigned to specific tutors perform better than students who are simply assigned to a pool of tutors (Carlson & Minke, 1974), and it is hypothesized that such specifically assigned tutors were a potent source of social reinforcement for their students. Carman (1975) found that tutoring increased perseverance, and it may be conjectured here that more intensive tutoring would increase perseverance even more. This same conjecture has been made by a number of other researchers and practitioners (Hecht, 1977; Kean & Welsh, 1980; Johnson & Steffensen, 1977). The theory, supported by personal experience but not yet by the results of controlled experimental research, is that a strong bond between the tutors and their students increases students' perseverance and hence improves their performance.

One other difference between Treatment II and Treatment I was that students in Treatment II usually had the opportunity to discuss the mistakes they had made on their tests with their tutors, as these tests were returned to the workshops on a daily basis. However, all students were required to discuss failed tests with the lab staff either immediately after their tests were graded or, if necessary, at a later time. The tests were returned to the workshops mainly for the purpose of allowing tutors to see how their students had fared, to sympathize with those who had failed and to encourage their students to forge ahead. It seems unlikely that students would have performed better in the course simply because they saw their mistakes a second time. In fact, many times the students were not interested in seeing the test again.

It thus seems appropriate to conclude that either the difference in grading systems or the difference in tutoring behaviors, or the two differences taken together caused a significant increase in student progress in this individualized mathematics program.

3. SUGGESTIONS FOR TEACHING

The results of this study suggest that it may be worthwhile to incorporate the following features into an individualized program in remedial mathematics:

1. students should be assigned to specific tutors and these tutors should be encouraged to establish close relationships with their students and to monitor their progress regularly;
2. incentives should be established for meeting test deadlines;
3. progress charts should be provided to both students and tutors both as a device for students to keep track of their progress and as a vehicle for promoting tutor-student interactions.

In addition, the failure of any of the treatments to be reasonably effective suggests that certain characteristics of the course in general were impeding success. Based on speculation as to what these characteristics may have been, it is further suggested that the following features be tested in an individualized program in remedial mathematics:

4. students should be separated into homogeneous groups and placed into courses appropriate for their level of preparation;
5. all courses should have end-of-the-term deadlines for course completion, in conjunction with a reasonable but strict policy on grades of Incomplete.

4. RECOMMENDATION FOR FURTHER RESEARCH

In light of the results of this study, the following recommendations are made for further research on individualized instruction, especially individualized instruction in the area of college-level remedial mathematics:

1. To separate the effects of grading systems and tutor behaviors, research should be conducted with treatments in which one or the other, but not both, are present.
2. To determine if organizational structure has a significant effect on student performance in settings in which successful programs are organized around small classes, an experiment could be performed in which students are assigned to tutors instead of to faculty members.
3. There were a number of interesting interactions among the treatments and the covariates which could be investigated to determine which students benefit most from which type of teaching method.
4. This experiment should be replicated in other settings in order to determine the scope of its conclusions. In particular, it should be replicated in situations in which students are under greater pressure to finish the program in a single term, and in situations in which a more homogeneous student population is present.

LIST OF REFERENCES

LIST OF REFERENCES

Akst, G. R. A study of the effect on learning of pacing and testing procedures in a two year college remedial mathematics course (Doctoral dissertation, Columbia University Teachers College, 1976). Dissertation Abstracts International, 1976, 37, 2035A.

Anderson, J. T., and Pritchett, G. D. A report on self-paced calculus at Hamilton-Kirkland Colleges. In T. L. Sherman, M. J. Hassett and R. B. Thompson (Eds.), Innovative teaching methods in introductory college mathematics. Tempe, Arizona: Rocky Mountain Mathematics Consortium, 1977.

Archer, J. A. Developmental mathematics: an ex post facto study. Community/Junior College Research Quarterly, 1978, 3, 75-85.

Arlin, M., and Webster, J. Time costs of mastery learning. Journal of Educational Psychology, 1983, 75(2), 187-195.

Baldwin, J. Survey of developmental mathematics courses at colleges in the United States. 1976. (ERIC Document Reproduction Service No. 125688.)

Bijou, S. W., Morris, E. K., and Parsons, J. A. A PSI course in child development with a procedure for reducing student procrastination. Journal of Personalized Instruction, 1976, 1, 36-40.

Billstein, R. Can you be innovative on a low budget? In T. L. Sherman, M. J. Hassett and R. B. Thompson (Eds.), Innovative teaching methods in introductory college mathematics. Tempe, Arizona: Rocky Mountain Mathematics Consortium, 1977.

Bloom, B. Mastery learning. In J. H. Block (Ed.), Mastery learning: theory and practice. New York: Holt, Rinehart & Winston, Inc., 1971.

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

Born, D. G. Examination performance and study behavior as a function of study unit size. In J. M. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Ill.: Charles C. Thomas, 1975.

Born, D. G., and Whelan, P. Some descriptive characteristics of student performance in PSI and lecture courses. The Psychological Record, 1973, 23, 145-152.

Cameron, L. Intermediate algebra (short report). In T. L. Sherman, M. J. Hassett, and R. B. Thompson, Innovative teaching methods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977.

Carlson, J. G., and Minke, K. A. The effects of student tutors on learning by unit mastery instructional methods. Psychological Record, 1974, 24, 533-543.

Carman, R. The effects of tutoring in developmental mathematics on the academic performance, attrition, and attitudes of community college students (Doctoral dissertation, University of California, 1975). Dissertation Abstracts International, 1975, 36, 623A.

Carsrud, A. L. Undergraduate tutors: are they useful? Teaching of Psychology, 1979, 6, 46-48.

Chatterly, L. J. BYU's modularized pre-calculus program: flexibility in two directions--course offering and credit obtainable. In T. L. Sherman, M. J. Hassett, and R. B. Thompson, Innovative teaching methods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977.

Cohen, J. C. Statistical power analyses for the behavioral sciences. New York: Academic Press, 1977.

Cohen, P. A., Kulik, J. A., and Kulik , C-L. C. Educational outcomes of tutoring: a meta-analysis of findings. American Educational Research Journal, 1982, 19, 237-248.

Davies, C. S., Born, D. G., and Semb, G. Effects of performance-based pacing schedules on student performance in the Personalized System of Instruction. Journal of Personalized Instruction, 1980, 4, 3-11.

Davis, M. L. An analysis of examination feedback verbalized by the instructor. In L. E. Fraley and E. A. Vargas (Eds.), Behavior research and technology in higher education. Gainesville: University of Florida, Society for Behavioral Technology and Engineering, Psychology Department, 1976.

Dubin, R., and Taveggia, T. C. The teaching-learning paradox. Eugene, OR: Center for the Advanced Study of Educational Administration, University of Oregon, 1968.

Eisenberg, T. Remedial mathematics and open admissions. School Science and Mathematics, 1981, 81, 341-346.

Eisenberg, T., and Browne, J. G. Using student-tutors in pre-calculus instruction. American Mathematical Monthly, 1973, 80, 685-688.

Farmer, J., Lachter, G. D., Blaustein, J. J., and Cole, B. K. The role of proctoring in personalized instruction. Journal of Applied Behavior Analysis, 1972, 5, 401-404.

Fey, J. T., Albers, D. J., and Fleming, W. H. Survey reveals dramatic changes in collegiate mathematics. Focus, 1982, 2(1), 1-2; 4-6.

Gindler, H. A., Marosz, W. A., and Romano, A. Who benefits most from individualized instruction in mathematics? Journal of Personalized Instruction, 1977a, 2, 204-208.

Gindler, H. A., Marosz, W. A., and Romano, A. Mastery learning in intermediate algebra. In T. L. Sherman, M. J. Hassett, and R. B. Thompson (Eds.), Innovative teaching meth ods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977b.

Glick, D. M., and Semb, G. Instructor-set pacing contingencies versus the absence of such contingencies in a personalized university course. Journal of Personalized Instruction, 1978, 3, 131-138.

Green, B. A. Physics teaching by the Keller plan at MIT. American Journal of Physics, 1971, 39.

Greenwood, M. Two factors involved in successful individualized mathematics programs. Two Year College Mathematics Journal, 1977, 8, 219-222.

Harris, M. B., and Liguori, R. A. Some effects of a personalized system of instruction in teaching college mathematics. Journal of Educational Research, 1974, 68, 62-66.

Hartley, S. S. Meta-analysis of the effects of individually paced instruction in mathematics (Doctoral dissertation, University of Colorado, 1977). Dissertation Abstracts International, 1977, 38, 4003A.

Hassett, M., Livermore, P., and Weis, D. An instructor-paced, individualized algebra system. In T. L. Sherman, M. J. Hassett, and R. B. Thompson (Eds.), Innovative teaching methods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977.

Haver, W. Developing skills in college algebra--a mastery approach. Two Year College Mathematics Journal, 1978, 9, 282-287.

Hecht, M. Personalized arithmetic remediation at Hunter College. In T. L. Sherman, M. J. Hassett, and R. B. Thompson, Innovative teaching methods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977.

Herring, C. R. An investigation into the effects of mastery learning on three variables in a college general mathematics course (Doctoral dissertation, The University of Tennessee, 1975).

Hess, J. H., Jr. Large-scale individualized instructional systems: a comparison of three working models. Educational Technology, 1977, 17, 37-44.

Hinton, J. R. Individualized learning: a summary of recent research. NSPI Journal, 1978, 17(4), 15-16.

Hursh, D. E. PSI: What do the data indicate? Journal of Personalized Instruction, 1976, 1, 91-105.

Jackson, H. E. A comparison of the effectiveness of three instructional formats in introductory calculus student achievement and attrition (Doctoral dissertation, The University of Tennessee, 1978). Dissertation Abstracts International, 1978, 39, 6606A.

Johnson, K. R., and Sulzer-Azaroff, B. The effects of different proctoring systems upon student examination performance and preference. In J. W. Johnston and G. W. O'Neill (Eds), Research and technology in college and university teaching. Atlanta, GA: Georgia State University Urban Life Center, 1975.

Johnson, M., and Steffensen, A. Individualized instruction at Northern Arizona University. In T. L. Sherman, M. J. Hassett, and R. B. Thompson (Eds.), Innovative teaching methods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977.

Kahn, P. B. Instructional innovations in physics at Stony Brook. American Journal of Physics, 1975, 43, 400-407.

Kean, E. S., and Welsh, C. A. C. Remedial tutorial programs at the university level. Educational Technology, 1980, 20 (July), 41-45.

Keller, F. S. Goodbye teacher . . . . Journal of Applied Behavior Analysis, 1968, 1, 78-79.

Keller, F. S., and Sherman, J. G. The Keller Plan handbook. Reading, MA: W. A. Benjamin, Inc., 1974.

Klopfenstein, K. The personalized system of instruction in introductory calculus. American Mathematical Monthly, 1977, 84, 120-124.

Kulik, C-L. C., Kulik, J. A., and Cohen, P. A. Instructional technology and college teaching. Teaching of Psychology, 1980, 7, 199-205.

Kulik, J. A., and Jaksa, P. PSI and other educational technologies in college teaching. Educational Technology, 1977, 17(9), 12-19.

Kulik, J. A., Jaksa, P., and Kulik, C. C. Research on component features of Keller's Personalized System of Instruction. Journal of Personalized Instruction, 1978, 3, 2-14.

Kulik, J. A., and Kulik, C-L. C. College teaching. In P. L. Peterson and H. J. Walberg (Eds.), Research on teaching: concepts, findings and implications. Berkeley: McCutchan Pub. Co., 1979.

Kulik, J. A., Kulik, C-L. C., and Cohen, P. A. A meta-analysis of outcome studies of Keller's PSI. American Psychologist, 1979, 34, 307-318.

Maltbie, A., Savage, R. G., and Wasik, K. L. The operation and evaluation of a proctorial system of instruction in mathematics. American Mathematical Monthly, 1974, 81, 71-78.

McDermott, C. W. The development of an individualized learning system for students studying intermediate algebra, final report. (ERIC Document Reproduction Service No. 077774).

Mendez, C. G. A mixed system of instruction in basic mathematics. Journal of Personalized Instruction, 1978, 3, 221-224.

Merritt, P. Some research support for a second chance for beginning algebra students. Two Year College Mathematics Journal, 1974, 5(2), 50-54.

Miller, R. L. Individualized instruction in mathematics: a review of research. Mathematics Teacher, 1976, 69, 345-351.

Morman, S. J. An audio-tutorial method of instruction vs. the traditional lecture-discussion method. Two-Year College Mathematics Journal, 1973, 4(3), 56-61.

Morris, E. K., Surber, C. F., and Bijou, S. W. Self-pacing vs. instructor pacing: achievement, evaluation and retention. Journal of Educational Psychology, 1978, 70, 224-230.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D.H. Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill, Inc., 1975.

Nott, M. E., Jr. New results of research comparing programmed and lecture-text instruction. Two Year College Mathematics Journal, 1971, 2(1), 19-22.

Overholser, J. S. Individualized instruction in mathematics (Doctoral dissertation, University of Oregon, 1979). Dissertation Abstracts International, 1979, 40, 4940A.

Pascarella, E. T. Student motivation as a differential predictor of course outcomes in personalized systems of instruction and conventional instruction methods. The Journal of Educational Research, 1977, 71, 21-26.

Pascarella, E. T. Interactive effects of prior mathematics preparation and level of support in college calculus. American Educational Research Journal, 1978, 15, 275-285.

Peluso, A. J., and Baranchik, A. J. Self-paced mathematics instruction: a statistical comparison with traditional teaching. American Mathematical Monthly, 1977, 84, 124-129.

Phillips, J. A self-paced multi-media remedial algebra course at an urban community college (Doctoral dissertation, New York University, 1981). Dissertation Abstracts International, 1981, 42, 587A.

Price, W. E. A mastery learning strategy for college freshman mathematics (Doctoral dissertation, State University of New York at Buffalo, 1971). Dissertation Abstracts International, 1971, 32, 841A.

Reiser, R. Self-pacing and instructor-pacing in PSI courses. 1976. (ERIC Document Reproduction Service No. 125688).

Riedel, R., Harney, B., LaFief, W., and Finch, M. The effect of time as a contingency on student performance in an individualized course. In B. A. Green (Ed.), Proceedings of the Second National Conference on Personalized Instruction in Higher Education. Washington, D.C.: Center for Personalized Instruction, 1976.

Robin, A. L. Behavioral instruction in the college classroom. Review of Educational Research, 1976, 46, 313-354.

Rogers, G. S., and Young, D. L. An evaluation of some aspects of a PSI introductory statistics course. In T. L. Sherman, M. J. Hassett, and R. B. Thompson (Eds.), Innovative teaching methods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977.

Romoser, R. C. Results of the second assessment study of developmental education programs in Ohio. State-wide Advisory Committee on Developmental Education, Ohio, 1978. (ERIC Document Reproduction Service No. 157587.)

SAS Institute Inc. SAS user's guide: Statistics, 1982 edition. Cary, NC: SAS Institute Inc., 1982.

Schoen, H. A plan to combine individualized instruction with the lecture method. Mathematics Teacher, 1974, 67, 7.

Schoen, H. L. Self-paced mathematics instruction: How effective has it been in secondary and post-secondary schools? Mathematics Teacher, 1976, 69, 352-357.

Semb, G., Conyers, D., Spencer, R., and Sanchez-Sosa, J. J. An experimental comparison of four pacing contingencies. In J. M. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Ill: Charles C. Thomas, 1975.

Sherman, J. G. The personalized system of instruction: 41 germinal papers. Menlo Park, CA: W. A. Benjamin, Inc., 1974.

Slate, A. R. The comparative effectiveness of four techniques of instruction of arithmetic in the community college (Doctoral dissertation, George Peabody College, 1975). Dissertation Abstracts International, 1975, 36, 2078A-2079A.

Smith, B. D. Completion rates in Mathematics 107. Internal report, The University of Tennessee at Chattanooga, 1982.

Steele, K., Legg, S. M., and Miles, G. B. Managing the procrastination syndrome in self-paced instruction. Educational Technology, 1980, 20(9), 46-49.

Struik, R. R., and Flexer, R. J. Self-paced calculus: a preliminary evaluation. American Mathematical Monthly, 1977, 84, 129-134.

Sutterer, J. R., and Holloway, R. E. An analysis of student behavior with and without limiting contingencies. In J. M. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Ill: C. C. Thomas, 1975.

Taveggia, T. C. Personalized instruction: a summary of comparative research, 1967-1974. American Journal of Physics, 1976, 44, 1028-1033.

Taylor, V. Individualized calculus for the "life-long" learner: A two-semester comparison of attitude and effectiveness. 1977. (ERIC Document Reproduction Service No. 146417.)

Taylor, V. Individualized mathematics for continuing education students. NSPI Journal, 1978, 17(4), 11-12.

Thompson, R. B. The individual learning system for algebra service courses. In T. L. Sherman, M. J. Hassett, and R. B. Thompson (Eds.), Innovative teaching methods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977.

Thompson, R. B., and McCoy, J. S. Student pacing or instructor pacing? American Mathematical Monthly, 1979, 86, 217-221.

Thompson, S. B. Do individualized mastery and traditional instructional systems yield different course effects in college calculus? American Educational Research Journal, 1980, 17, 361-375.

Urban, J. D., Jr. The design and evaluation of an undergraduate mathematics course incorporating team teaching, individualized instruction, and team supervision (Doctoral dissertation, University of Pittsburgh, 1971). Dissertation Abstracts International, 1971, 33, 6301A.

Weir, M. D. PSI in linear algebra at the Naval Postgraduate School. In T. L. Sherman, M. J. Hassett, and R. B. Thompson (Eds.), Innovative teaching methods in introductory college mathematics. Tempe, AR: Rocky Mountain Mathematics Consortium, 1977.

Wesp, R., and Ford, J. E. Flexible instructor pacing assists student progress in a personalized system of instruction. Teaching of Psychology, 1982, 9(3), 160-162.

Whitehurst, C., and Madigan, J. Slow learners in PSI courses: Do they learn less? Journal of Higher Education, 1975, 46, 55-62.

Willett, J. B., Yamashita, J. J. M., and Anderson, R. D. A meta-analysis of instructional systems applied to science teaching. Journal of Research in Science Teaching, 1983, 20, 405-407.

Williams, R. L., and Lawrence, J. The effects of frequency of quizzing in a lecture course. In J. W. Johnston and G. W. O'Neill (Eds.), Research and technology in college and university teaching. Atlanta, GA: Georgia State University Urban Life Center, 1975.

Wine, J. D., and Olan, M. Individualizing the beginning mathematics course. International Journal of Mathematics Education in Science and Technology, 1983, 14, 1-7.

Wood, W. S., and Wylie, R. G. Individualized systems of instruction are better . . . for whom? In J. M. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Ill: Charles C. Thomas, 1975.

Wykoff, J. A. A modular, mastery-based method of teaching intermediate algebra at the university level. Journal of Personalized Instruction, 1980, 4, 148-151.

Zwerling, L. S. Developmental math with a difference. Change, 1977, 9(1), 28-31.

APPENDIXES

APPENDIX A

SYLLABUS FOR MATHEMATICS 107

TABLE A-1

TOPICS IN MATHEMATICS 107, BY UNIT AND WEEK OF PRESENTATION

Week	Unit	Topics[a]
1	1	Integer arithmetic
2	2	Simplifying polynomial expressions
3	3	Operations with algebraic fractions with integer dnominators
4	4A	Solving linear equations and literal equations
5	4B	Percent problems and linear inequalities
6	5	Word problems
7	6	Graphing linear equations and inequalities
8	7	Systems of linear equations
9	8 & 9	Laws of exponents, including negative exponents, and multiplication and division of polynomials
10	10	Factoring
11	11	Multiplication and division of algebraic fractions
12	12	Addition and subtraction of algebraic fractions
13	13	Radical expressions (square roots only)
14	14	Quadratic equations and absolute value equations

[a]Text: Algebra, a First Course, by Baley and Holstege.

APPENDIX B

STATISTICAL DATA

Statistical data, including current unit and arithmetic, algebra and mathematics attitude covariates for each subject by instructor and treatment, first experiment:

Inst	Trtmt	Unit	Ar	Al	Att	Inst	Trtmt	Unit	Ar	Al	Att
1	I	1	51	20	32	1	II	1	51	16	
1	I	16	85	28	44	1	II	10	94	24	49
1	I	7	80	32	36	1	II	2	51	28	35
1	I	4	62	36		1	II	11	74	40	62
1	I	1			27	1	II	8	62	32	25
1	I	4	54	32	21	1	II	16	88	28	50
1	I	6	82	40	21	1	II	16	85	40	57
1	I	10	65	44		1	II	3	68	20	12
1	I	16	85	60	58	1	II	10	77	20	26
1	I	16	100	12		1	II	5	88	36	59
1	I	10	77	32	38	1	II	10	54	32	20
1	I	4	80	16	28	1	II	10	68	20	47
1	I	16	85	32	28	1	II	15	88	24	30
1	I	10	77	28	22	1	II	16	57	28	60
1	I	16	94	20		1	II	10	88	52	30
1	I	4	45	36		1	II	10	88	40	29
1	I	10	97	44	42	1	II	16	74	48	58
1	I	16	85	20	32	1	II	1	68	32	19
1	I	16	91	44	29	1	II	8	82	32	36
1	I	12	74	40	57	1	II	10	42	32	25
1	I	16	85	36	43	1	II	4	71	28	32
1	I	5	77	44	38	1	II	5	85	28	16
1	I	7	88	24	46	1	II	8	77	44	
1	I	16	94	68	25	1	II	13	65	32	51
1	I	16	77	24		1	II	1	40	24	
1	I	7	71	20		1	II	10	82	36	53
1	I	4	80	56	28	1	II	10	62	20	46
1	I	3	62	32	35	1	II	16	77	28	
1	I	2	68	20	37	1	II	16	88	40	38
1	II	7	62	0	53	1	II	13	85	44	45
1	II	4	51	40		1	II	11	57	24	
1	II	12	42	12	30	1	III	16	85	24	60
1	II	16	91	40	50	1	III	1			26
1	II	4	85	24		1	III	12	62	28	21
1	II	16	71	40		1	III	5	80	40	13
1	II	5	91	24	36	1	III	16	85	28	30
1	II	14	65	44		1	III	12	71	28	29
1	II	5	71	24	36	1	III	16	74	44	31
1	II	12	80	28	30	1	III	6	65	28	35
1	II	16	68	28	34	1	III	16	80	36	62
1	II	1			28	1	III	12	68	12	47
1	II	16	62	16	19	1	III	16	71	28	32

Inst	Trtmt	Unit	Ar	Al	Att
1	III	3	74	24	37
1	III	13			
1	III	16	74	36	51
1	III	3	82	36	
1	III	16	94	52	57
1	III	16	62	44	40
1	III	5	82	24	
1	III	16	85	40	60
1	III	3	80	32	28
1	III	4	82	40	34
1	III	10	88	32	21
1	III	5	82	32	33
1	III	3	82	24	49
1	III	1	82	28	10
1	III	10	82	32	28
2	I	12	62	28	44
2	I	11			42
2	I	2	74	36	
2	I	4	65	28	45
2	I	7	88	32	8
2	I	5	80	36	36
2	I	10	48	32	
2	I	5	85	28	71
2	I	4	71	44	49
2	I	1	74	24	21
2	I	1	74	20	52
2	I	3	74	28	
2	I	1	40	16	
2	I	16	85	40	17
2	I	3	77	40	35
2	I	5	60	20	9
2	I	11	80	20	40
2	I	3	68	28	0
2	I	4	88	36	24
2	I	2	74	20	38
2	I	13	82	40	51
2	I	5	51	28	28
2	I	3	62	8	
2	I	1	82	20	36
2	I	10	77	32	32
2	I	15	62	32	65
2	II	4	80	44	35
2	II	16	60	32	32
2	II	2	22	24	
2	II	10	71	28	53
2	II	14	80	44	27
2	II	6	60	24	35
2	II	10	77	28	29
2	II	12	71	40	26
2	II	16	82	36	
2	II	1	82	36	49

Inst	Trtmt	Unit	Ar	Al	Att
2	II	2	68	28	44
2	II	10	60	40	28
2	II	10	65	40	44
2	II	1	31	24	50
2	II	16	65	32	32
2	II	16	80	44	22
2	II	11	80	24	47
2	II	14	77	16	
2	II	10	65	12	30
2	II	1	77	28	24
2	II	10	77	36	26
2	II	16	77	40	47
2	II	2	82	32	
2	II	4	77	36	
2	II	15	62	28	59
2	II	4	85	20	59
2	II	16	88	40	28
2	II	16	88	36	20
2	II	16	62	32	29
2	II	5	82	28	54
2	II	7	85	52	45
2	II	2	40	32	58
2	II	5	74	36	35
2	II	16	71	36	
2	II	16	77	40	
2	II	1	57	24	
2	II	2	82	48	
2	III	5	91	40	26
2	III	13	71	24	53
2	III	1	31	12	
2	III	14	74	24	
2	III	16	65	32	59
2	III	8	91	36	40
2	III	4	100	32	22
2	III	14	51	32	20
2	III	16			22
2	III	5	88	24	64
2	III	1	51	16	20
2	III	16	74	36	28
2	III	2	68	36	21
2	III	2	68	12	18
2	III	11	80	28	49
2	III	10	60	16	46
2	III	16	57	40	38
2	III	16	88	40	35
2	III	7	60	20	15
2	III	12	71	28	22
2	III	6	68	24	35
2	III	5	68	36	31
2	III	2	71	40	33
2	III	11	82	28	32

Statistical data, including current unit and arithmetic, algebra and mathematics attitude covariates for each subject by instructor and treatment, second experiment:

Inst	Trtmt	Unit	Ar	Al	Att
1	I	16	88	48	
1	I	16	60	44	49
1	I	5	77	40	
1	I	16	71	32	15
1	I	13	68	28	16
1	I	10	65	28	27
1	I	2	85	36	41
1	I	5	54	32	52
1	I	16	77	44	28
1	I	5	48	28	21
1	I	10	57	20	14
1	I	4	42	24	26
1	I	14	97	52	40
1	I	16	97	44	54
1	I	4	77	24	6
1	I	16	97	44	4
1	I	10	94	44	24
1	I	3	94	20	15
1	I	11	71	40	32
1	I	1			21
1	I	10	82	36	3
1	I	1	60	24	16
1	I	8	62	44	
1	I	3	48	20	22
1	II	1	25	8	24
1	II	16			40
1	II	4	54	20	42
1	II	11	77	44	29
1	II	3	80	24	20
1	II	10	68	44	31
1	II	2	71	20	9
1	II	14	80	28	20
1	II	1	31	16	
1	II	12	85	36	40
1	II	15	88	32	41
1	II	16	80	44	38
1	II	1	68	24	38
1	II	10	94	40	60
1	II	2	57	20	33
1	II	3	40	28	4
1	II	15	77	36	13
1	II	16	94	28	59
1	II	11	80	24	31

Inst	Trtmt	Unit	Ar	Al	Att
1	II	6	77	20	6
1	II	16	88	36	41
1	II	1	57	12	27
1	II	10	85	16	54
1	II	10	82	32	41
1	II	7	88	12	46
1	II	12	51	32	18
1	II	1	68	20	53
1	II	3	97	24	36
1	II	16	82	44	34
1	II	12	80	20	41
1	III	12	85	32	43
1	III	1	45	28	30
1	III	5	91	36	52
1	III	1	42	20	45
1	III	8	54	44	
1	III	7	65	48	27
1	III	12	54	16	29
1	III	10	88	32	19
1	III	16	71	32	31
1	III	13	88	16	19
1	III	14	85	28	34
1	III	7	68	28	22
1	III	16	88	44	22
1	III	1	51	16	49
1	III	10	65	24	63
1	III	7	48	32	45
1	III	16	77	36	20
1	III	7	60	32	2
1	III	1	31	12	13
1	III	1	40	16	29
1	III	16	94	28	29
1	III	2	88	36	53
1	III	16	54	36	54
1	III	16	100	64	41
2	I	11	82	48	41
2	I	10	60	48	23
2	I	1	60	28	25
2	I	1	31	8	8
2	I	10	82	28	22
2	I	7	71	44	14
2	I	16	85	24	37
2	I	5	80	12	4

Inst	Trtmt	Unit	Ar	Al	Att
2	I	4	71	28	25
2	I	1	45	28	38
2	I	2	71	28	30
2	I	13	88	32	22
2	I	16	82	24	58
2	I	10	88	40	28
2	I	16	82	36	48
2	I	16	85	52	34
2	I	12	74	44	36
2	I	7	82	12	37
2	I	10	85	20	26
2	I	4	74	36	59
2	I	2			43
2	I	10			13
2	I	13	80	40	
2	I	1	54	24	6
2	I	5	77	24	8
2	I	4			30
2	II	16	82	24	26
2	II	16	77	32	21
2	II	6	62	24	22
2	II	10	68	28	15
2	II	15	77	44	40
2	II	13	77	28	44
2	II	10	68	28	42
2	II	1	34	32	48
2	II	4	57	32	65
2	II	6	80	52	51
2	II	12	97	52	53
2	II	16	57	44	76
2	II	12	80	32	38
2	II	13	85	20	29
2	II	5	80	44	50
2	II	10	77	24	28
2	II	16	100	44	60
2	II	1	68	16	22
2	II	16	97	44	38
2	II	15	85	44	
2	II	10	88	40	43
2	II	12	60	32	42
2	II	11	62	32	10
2	II	12	68	48	43
2	II	2	62	44	4
2	II	13	77	28	21
2	II	4	71	48	29
2	II	13	82	32	27
2	II	11	54	16	26
2	II	5	68	24	35
2	II	1	60	24	38

Inst	Trtmt	Unit	Ar	Al	Att
2	III	15	77	44	61
2	III	5	54	36	59
2	III	15	82	40	53
2	III	5	80	20	41
2	III	16	97	44	35
2	III	1	62	12	35
2	III	16	80	40	35
2	III	16	88	52	23
2	III	15	74	40	28
2	III	5	62	28	30
2	III	10	71	12	13
2	III	16	91	32	40
2	III	1	34	16	24
2	III	10	71	24	60
2	III	16	88	28	28
2	III	10	54	44	46
2	III	4	88	40	47
2	III	14	82	24	45
2	III	2	62	36	67
2	III	4			30
2	III	7	74	12	41
2	III	1	82	16	24
2	III	1	45	12	25
2	III	1	42	40	6
2	III	16	88	48	52

VITA

Betsy Darken Smith was born in Plainfield, New Jersey on October 9, 1950. In 1972, she received a Bachelor of Arts degree in mathematics from Kirkland College in Clinton, New York. From 1972 to 1976 she was a teaching fellow in the Department of Mathematics at the University of Michigan, where she received a Master of Arts degree in mathematics in 1974.

From 1976 to 1979, she taught developmental mathematics, first at Armstrong State College in Savannah, Georgia, and later at Augusta College in Augusta, Georgia. Since 1979 she has been a member of the Mathematics Department at The University of Tennessee at Chattanooga, where she became the Director of Developmental Mathematics in 1981. She received her Doctor of Philosophy degree with a major in Education from the University of Tennessee, Knoxville in August 1984, and will shortly receive a Master of Science degree in Statistics from the same institution.

She has published several articles as well as delivered a number of papers in the areas of mathematics education over the past several years. She is a member of the National Council of Teachers of Mathematics, the Mathematical Association of America, the Tennessee Mathematics Teachers' Association and the National Association for Developmental Education.

She is married to Larry W. Smith.