THE ASSOCIATION BETWEEN CLASS SIZE, ACHIEVEMENT,
AND OPINIONS OF UNIVERSITY STUDENTS
IN FIRST-SEMESTER CALCULUS

DISSERTATION

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Eddie N. Warren, B.S., M.A.
Denton, Texas
May, 1988
The purposes of the study were: to determine the relationship between class size and academic achievement among university students in first-semester calculus classes, and to compare opinions about the instructor, course, and classroom learning environment of university students in small first-semester calculus classes with those in large classes.

The sample consisted of 225 university students distributed among two large and two small sections of first-semester calculus classes taught at the University of Texas at Arlington during the fall of 1987. Each of two tenured faculty members taught a large and small section of approximately 85 and 27 students, respectively. During the first week of the semester, scores from the Calculus Readiness Test (CR) were obtained from the sample and used as the covariate in each analysis of covariance of four periodic tests, a comprehensive final examination, and final grade average. The CR scores were also used in a logistic regression analysis of attrition rates between each pair of
large and small sections of first-semester calculus. Three semantic differentials were used to test the hypotheses relating to student opinion of the instructor, course, and classroom learning environment.

It was found that for both pairs of large and small first-semester calculus classes there was no significant difference in the adjusted means for each of the four periodic tests, the final examination scores, the final grade averages, and the attrition rates. It was also found that the means of the student evaluation of the course by students in small and large classes were not significantly different, and the results of the student evaluations of the instructor and classroom learning environment by students in small and large first-semester calculus classes were mixed.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v.</td>
</tr>
</tbody>
</table>

### Chapter

#### I. INTRODUCTION

- Statement of the Problem
- Purposes of the Study
- Hypotheses
- Significance of the Study
- Theoretical Frame of Reference
- Summary
- Chapter Bibliography

#### II. SYNTHESIS OF THE LITERATURE

- Research of Class Size and Achievement
- Research of Class Size and Student Opinion
- Summary
- Chapter Bibliography

#### III. PROCEDURES FOR COLLECTION OF DATA

- Introduction
- Population and Sample Selection
- Research Design
- The Treatment
- Delimitations
- Description of the Instruments
- Summary
- Chapter Bibliography

#### IV. PRESENTATION AND ANALYSIS OF THE DATA

- Description of the Sample
- Findings
- Hypothesis Two
- Hypothesis Three
- Hypotheses Four
- Summary of Findings
- Chapter Bibliography
V. SUMMARY OF THE STUDY

Summary
Discussion of the Findings
Hypotheses One and Two
Hypotheses Three and Four
Demographics
Conclusions
Recommendations
Chapter Bibliography

APPENDICES

BIBLIOGRAPHY
LIST OF TABLES

Table | Page
--- | ---
1. Description of the Sample | 56
2. Description of the Sample | 57
3. Analysis of Covariance Summary for Periodic Test One | 61
4. Analysis of Covariance Summary for Periodic Test Two | 64
5. Analysis of Covariance Summary for Periodic Test Three | 66
6. Analysis of Covariance Summary for Periodic Test Four | 69
7. Analysis of Covariance Summary for Final Examination | 71
8. Analysis of Covariance Summary for Final Grade Averages | 73
9. Attrition Rates for Each Pair of Small and Large Classes | 75
10. Logistic Regression Analysis of Attrition Rates | 76
11. Attrition Rates for Subgroups Within Two Pairs of Large and Small Classes | 78
CHAPTER I

INTRODUCTION

With increasing frequency large sections of university mathematics classes are being anticipated due to financial pressures. Faculty and student resistance is to be expected. The belief among most university students and faculty is that small classes are better (McKeachie, 1980). Do small and large mathematics classes result in differences in achievement and opinion? A university study done by Simmons (1959) with remedial algebra students resulted in a significant difference in achievement in favor of small class size. In general, however, research studies have drawn mixed conclusions with some suggesting small or large to be better and others concluding that there is no difference (Williams, Cook, Quinn, & Jensen, 1985). The literature involving large classes of mathematics students is almost nonexistent. This study, utilizing a faculty experienced with teaching large sections of mathematics extending from the freshman to the senior level, should be beneficial in assisting those involved with the teaching, scheduling, or administration of mathematics classes.

Statement of the Problem

The problem of this study was the association between
class size, achievement, and opinions of university mathematics students.

Purposes of the Study

The purposes of this study were to:

1. determine the relationship between class size and academic achievement among university mathematics students in first-semester calculus classes, and
2. compare opinions about the instructor, course, and classroom learning environment of university students in small first-semester calculus classes with those in large classes.

Hypotheses

To carry out the purposes of this study the following hypotheses were tested:

1. There is no significant difference in achievement between students in large and small first-semester calculus classes.

2. The percentage of students in small first-semester calculus classes earning a course grade of W or F is not significantly different from that of students in the large classes.

3. Student evaluations of the instructor and the course by first-semester calculus students in small classes is not significantly different from that of students in the large classes.
4. Students in small first-semester calculus classes respond more positively to the classroom learning environment than those in the large first-semester calculus classes.

Significance of the Study

Students and faculty prefer small classes (McKeachie, 1980). Can large classes be justified? As budgets tighten, college students, faculties, administrators, and others directly and indirectly involved will be confronted with the issue of large classes (Splaine, 1980). Should students expect lower grades and instructors anticipate lower student evaluations as class size increases? Can large classes be justified in the minds of administrators and others responsible for university funding decisions? Can student recruiters present a valid argument to questioning recruits or can parents be assured that their undergraduate children are being given equitable treatment in the classrooms? Based on the disproportionate amount of class size literature existing in the disciplines of psychology and education, the answer might be justifiable in these areas. However, the literature summary presented in Chapter II demonstrates that other subject areas, except perhaps in economics, are the beneficiaries of very little research on class size.

This study was significant in that examining the opinions of mathematics students in large classes toward the
instruction they are receiving and comparing the levels of achievement of students in small and large first-semester calculus classes may:

1. provide a rationale for making decisions involving class size in university level mathematics, and
2. contribute to class-size research in a critical subject area typically required of undergraduates.

Theoretical Frame of Reference

If students or instructors are asked whether learning is promoted better within the environment of a small class setting or that of a large class, the response will favor greatly that of the small class (McKeachie, 1980). At the University of Texas at Austin, a university-wide study involved students in class size experiments attempting to increase the effectiveness of teaching and management of large classes (Lewis, 1982, p. 1). One conclusion was that the students valued highly the instructor who was organized, knowledgeable, and could interact well with students (p. 54). Also students said they enjoyed most those classes in which the instructor used a lot of reinforcement and in which a large number of questions were asked by students (p. 50). It was discovered, however, that student participation is inversely proportional to the class size (p. 34). So given the same instructor using the same teaching method in both large and small classes, would the different levels of
student participation be enough to cause a significant difference in student achievement?

McConnell and Sosin (1984) drew the same conclusions as the University of Texas researchers concerning student participation, but they also found it more difficult for the students to be as attentive and free from distraction. In a survey of 737 classes, Feld and Grofman (1977) reported that students are less likely to attend large classes than small. If the achievement level remains the same for large and small classes, then there are several possible explanations.

The first and perhaps the most obvious explanation is that students in large classes compensate for the lack of learning in class. This would especially hold true for courses where achievement is measured by knowledge obtained from a textbook. Since the traditional criterion for student achievement in large classes is performance on some form of an objective test (Lewis, 1982, p. 34; McKeachie, 1978, p. 207), then it might seem reasonable to expect students to be able to perform at the same level whether in a large or small class. McKeachie (p. 207) believes this to be the case. He maintains that there is a criterion problem when using the objective final examination or other objective test as the sole achievement criterion.

Parsons, Ketcham, and Beach (1958) demonstrate this criterion problem in a dramatic way in a study that included groups of students that did not come to class at all. The
students that did best on the final examination were the ones who did not attend class. They theorized that since those not attending studied only the textbook and the test covered only information found in the textbook, then they were more prepared for this kind of objective examination than those involved in class discussions, listening to lectures on alternate viewpoints, and being exposed to supplementary materials.

A second explanation offered for no significant difference in achievement between large and small classes is that the students actually achieve at the same level for the defined objectives. The Parsons, Ketcham, and Beach study also illustrates this possibility. There is, however, no consensus of findings for the subject areas that have been dominant in class size research (McKeachie, 1978, p. 207). Williams et al. (1985, p. 316) believe that it is possible that class size has little, if any, effect on lower level educational outcomes, such as recall of facts, but class size could have an effect on achievement for higher level outcomes, such as the development of critical thinking and problem-solving skills. It is their belief that studies addressing this hypothesis should be conducted at the university level.

McKeachie (1980, p. 25) theorizes that since many learning theories suggest that because the effective lecture involves interaction between instructor and students, the
large class may be inferior even for lectures since the participation by students is less for large classes than for small. It may be that the participation encountered in small classes, highly valued by students and faculty, is significant and has an effect although often masked by an ineffective achievement measure. Social psychologists Thomas and Fink (1963) suggest that group participation utilizes a variety of skills, knowledge, and insights. As the group increases in size the number of needs of the group increases. There is a point reached whereby the resources of students normally utilized in group participation become extremely limited in use due to the unwillingness or impracticality of group interaction. Therefore, since the resources of the instructor and only a few, if any, participants are incapable of meeting this heightened number of group needs, the individual group members achieve at a lower level.

The Macomber and Siegel experiments at Miami University in Ohio (1957) attempted to measure achievement in skills, such as problem solving, and found students of small classes to outperform the larger classes of students. This finding would be consistent with the theory of Thomas and Fink. Moreover, Siegel and Siegel (1964) have suggested that student interaction and feedback are particularly important for students with low levels of motivation, lack of subject-matter sophistication, or the tendency to memorize facts.
It appears that these types of students find it most difficult to adjust their level of performance as a result of being in a large class, with limited interaction with the instructor, than the more capable students. Hence, these students might not achieve at the same level as their counterparts in a small class. This would indicate that lower level students in a small class might outperform the corresponding group of students in a large class. Especially this might be anticipated in a problem solving course such as calculus where analytic and algebraic skills are being utilized simultaneously in many of the problem-solving settings.

In summary, it appears that if a large and small class achieved at the same level on an objective test involving lower level educational outcomes, then it is possible that students in a large class might not achieve at the same level as those in a small class environment when a test involving problem solving skills is the achievement criterion. This indicates that class-size-research results in subject areas previously studied does not necessarily transfer to the subject area of calculus.

The primary focus of this experiment has been to lay out a different framework of analysis for measuring the criterion of achievement. Rather than using the traditional criterion of the final examination or course grade as the sole measure of achievement, separate analyses of covariance
were used for each of four periodic examinations along with the measures of the comprehensive final examination and final grade average. If McKeachie and others (Nachman & Opopchinsky, 1958) are correct in their assumption of a frequently occurring criterion problem, and if problem solving requires an added degree of interaction for full achievement, then it is anticipated that there will be a greater degree of achievement by the students in small classes. This difference in achievement might be evidenced by a significant difference for the first achievement measure while the difference lessens for the second measure as the better students adjust to the large classes and the more unsophisticated in the content area begin to drop out. Furthermore, the final or last two measures of achievement might be the same, revealing the criterion problem associated with final examination or course grade measures. If this line of reasoning is accurate, then it is anticipated that the combined drop and failure rates for the large classes will be greater than that of the small classes. That is, in large classes only the "hearty" survive. This study has explored this hypothesis as well as provide class size research in a subject area distinct from the more often researched subject areas of psychology, education, and economics/business.

Summary

Students and faculty prefer small classes when given
the choice. But presently, and for the foreseeable future, it appears that there will be an increase in the utilization of large classes. And, although the research literature involving large university mathematics classes is almost nonexistent, in general, university studies of class size associations with achievement have drawn mixed conclusions.

Williams, Cook, Quinn, and Jensen theorize that class size has an effect on achievement for higher level outcomes, such as problem solving skills. McKeachie, however, theorizes that the large class may be inferior even for lectures due to the lower level of student involvement, and this inferiority is masked by a criterion measurement problem. L. Seigel, L. C. Siegel, Thomas and Fink also offer explanations that suggest students in a small class should outperform a corresponding group of students in a large class.

The principal purpose of this study was to determine the effect of class size on achievement for university students involved in a mathematics course requiring a large degree of problem solving skill utilizing a less traditional framework of analysis for measuring the achievement. Theory suggests that students of small classes should outperform those in large classes and that the combined drop and failure rates for the large classes should be greater than that of the small classes.
CHAPTER BIBLIOGRAPHY


Parsons, T. S., Ketcham, W. A., & Beach, L. R. (1958, August). Effects of varying degrees of student
interaction and student-teacher contact in college courses. Paper presented at the meeting of the American Sociological Society, Seattle, WA.


CHAPTER II

SYNTHESIS OF THE LITERATURE

Research of Class Size and Achievement

In 1957, in response to the president of his institution, Wilkinson (1958) undertook a survey of published literature on class size and its effect on student achievement at the university or college level. His report of that survey listed a number of observations about the experimental studies he examined in addition to his conclusions about the effect of class size on achievement.

Wilkinson observed that while the differences for or against a given class size may in some cases be significant statistically, they were, for the most part, small. Another observation he made involved the meaning of student achievement. Almost all of the experimental studies evaluated used objective tests as the sole measure of achievement. In fact, one of the most common weaknesses promoted by large classes is the tendency to rely exclusively on this form of testing and to exclude the measurement of certain aspects of achievement which small classes promote but fail to be measured by the objective test. More specifically, the argument is that only a knowledge of facts is being tested while such factors as problem-solving and analytic skills are being ignored.
Another observation by Wilkinson was the near neglect of experimental studies outside the disciplines of education and psychology. He found that no significant attempt had been made to study subject-matter as a factor in achievement in large and small classes. Moreover, Wilkinson noted that one facet of his literature survey provided evidence that subject-matter can make a difference in performance on tests between classes. He believed that fundamental distinctions between attitudinal subject-matter, skills courses, and factual-content courses should be investigated as factors that affect the relationship between class size and achievement.

Wilkinson also underscored a factor that adds to the difficulty of evaluating the results of experimental studies of class size and achievement: the words "large" and "small" are not universally understood. Large may refer to a class of size 35 or to one involving several hundred students while small may refer to a class of size two to 35 or even larger.

Having weighed the result of his survey, Wilkinson concluded that class size has little, if any, effect on achievement if the objective test is the sole measure of achievement. But he decided that there was no evidence to justify the use of large classes in all subjects. Moreover, he concluded that most educators believe that there are a number of valuable intangibles that the student will not
acquire in a large class and that these, rather than the accumulation of facts, are the primary objectives of education.

According to Nelson (1959), class size studies at the college level began in 1923 with the work of Mueller (1924) and Edmondson and Mulder (1924). Mueller's well-controlled class size study (McKeachie, 1980) revealed a significant difference in achievement between a large class of 40 students and a small class of 20 students in an introductory psychology course. The difference was in favor of the small class.

Edmonson and Mulder's experiment (1924) utilized two classes of education students, mostly women. The small class consisted of 45 students, with 10 of these men, while there were 109 students in the large class, which included 12 men. Intelligence was tested and considered to be approximately the same for both classes and the same instructor taught both classes. Although there was no significant difference in achievement the researchers concluded that the students in the small class displayed a higher degree of mental alertness, interest, and class participation.

Encouraged by the results of Edmonson and Mulder (McKeachie, 1978, p. 203) a three-year program of experimentation was undertaken by the University of Minnesota to study the class size effect on achievement.
Nelson (1959) considers this study to be the best and most comprehensive of the early class size experiments. Hudelson (1928) reported on the resulting 59 experiments covering such subjects as physics, history, law, business, psychology, education, and sociology. However, approximately 78 percent of these studies involved psychology or education. The small classes varied in size from 12 to 60 students while the large classes ranged from 35 to 150 students. Using the final examination as the measure of achievement, 38 of the experiments resulted in no significant difference in achievement. Of the 21 recording significant differences 19 were in favor of large class size with only two favoring small class size.

McKeachie (1980), a frequent reviewer of class size literature (1962, 1963, 1978, 1980), points out that a countering base of support for small classes came from studies in the teaching of French conducted by Cheydleur (1945) at the University of Wisconsin between 1919 and 1943. Cheydleur studied the results of objective departmental examinations for several hundred classes. He found a consistent difference in results favoring the small class size. However, McKeachie (1980) believes that large lecture classes are not generally inferior to small lecture classes when the traditional objective test of knowledge is used as the measure of achievement. But when goals such as higher-level thinking, application, motivation, and attitudinal
change are considered, the smaller class is more effective.
Yet other reviewers of class size literature (Albritton, 1984; Connor, 1977) claim that with a thorough understanding of the problems associated with large classes together with appropriate teacher preparation and training, class size should not have an effect on student achievement.

According to Albritton (1984), there is no research, poll, or word of mouth query that contradicts the fact that teachers prefer small classes. This is based on their belief that smaller is better in terms of how energetic and creative they can be as teachers together with the reasoning that smaller classes provide a more desirable classroom environment. A study by Filby, Cahen, McCutcheon, and Kyle (1980) indicates that smaller classes experience fewer classroom management problems, fewer discipline problems, less teacher anxiety, and greater teacher accuracy in the diagnosis and monitoring of student progress and need.

More specifically, Noli (1980) lists a number of reasons for an expected improvement in achievement for decreased class size:

1. Teachers are more able to diagnose student needs, more effectively fit a cure to the diagnosed problem, and offer more substantive and helpful feedback.

2. Fewer discipline problems of smaller classes allow for more instruction time per class.

3. Smaller classes create a higher level of
expectation that a teacher has of students which often leads to a higher level of student morale and performance.

The Macomber and Siegel experiments (1957, 1960) at Miami University of Ohio are considered by McKeachie (1980) as being significant class size studies because, in addition to the traditional objective tests of knowledge, measures of critical thinking and problem-solving, scales measuring stereotypic attitudes, tests of students motivation, and measures of student attitudes toward instruction were used. The researchers were funded for a period of three and one-half years and included approximately 120 sections from a variety of courses. Subject areas involved were government, business, English, geography, psychology, French, chemistry, algebra, zoology, history, and physics.

A class size was considered large if it was at least two times the size of the section being compared. Most of the small classes were in the 25 to 35 category, but the large classes ranged in size from 50 to 225 students. The students in the large sections were matched with those in the small sections by the use of scores from standardized tests given all entering freshmen. Regular faculty members served as instructors, but neither the teaching method nor the course objectives were controlled.

There were 45 comparisons of large and small classes made using the objective portions of the final examinations
as the measure of achievement. Only four comparisons were found to be significant at the .05 level with three of these favoring small classes. Composition and literature, geography, and zoology favored small classes while one comparison of large and small sections of composition and literature favored large class size.

The researchers believe that the results of the studies indicate that the ability to solve problems or synthesize information is not generally affected by class size. Also, stereotypes and misconceptions were overcome about equally by students in the large and small sections. Moreover, when retention of knowledge was measured one to two years after completion of the courses, there were no significant differences. However, of the 3000 initially involved in the study, only 1277 were still enrolled in the university and able to participate in the retention study. Of the nine courses compared in the retention study, eight of the small differences favored the small class size (45).

Nelson (1959), caught in the typical enrollment surge of the 1950s, found it surprising that although it was assumed that the small class is the superior class size for teaching economics there was virtually no evidence to support this viewpoint. The only experiment he found that was designed to relate class size to teaching effectiveness involving economics was a comparison between a large television section of 141 students with a small lecture
section of 48 students (Macomber & Siegel, 1957). In this
instance no significant difference in student achievement
was found where achievement was measured by an objective
final examination.

In addition to the experiments directed by Edmonson and
Mulder (1924) and Macomber and Siegel (1957), Nelson found
an experiment involving four sections of American Government
reported by Rohrer (1957) and one involving two sections of
beginning geography (Perry, 1957). The experiment reported
by Rohrer involved large sections of sizes 309 and 332 and
small sections of sizes 23 and 31, while the experiment by
Perry involved large sections of approximately 125 students
and small sections of 30 or less. Using an objective final
examination as the achievement measure, the differences in
achievement were not found to be significant. From these
four studies Nelson concluded that large classes of from 40
to 250 students might be as effective as small sections of
less than 40 students if achievement is measured by an
objective final examination.

In order to test the hypothesis that the teaching
effectiveness of economics is not importantly related to
class size Nelson, (1959) designed an experiment involving
four instructors, each teaching one large and one small
section of elementary economics. The students in each
instructor's classes were matched according to major,
classification (freshman, sophomore), and gender. The seven
small classes varied in size from 16 to 20 while the seven large classes ranged in size from 85 to 141. Using a common objective final examination for all 14 sections and analyzing the scores by an analysis of covariance, with prior grade-point average as the covariant, there was no significant difference between the large and small classes. When the means of four major examinations were used as the measure of achievement, six of the seven comparisons resulted in no significant difference while one resulted in a significant difference in favor of the small section. Nelson concluded that the small section can not be assumed to be a superior form of organization for teaching elementary economics at the university level.

Craig, O'Neill, and Elfer (1979) conducted a two-year follow-up experiment designed to study the association between retention performance and class size for an introductory economics class. The large classes had a minimum size of 250 students while the small classes ranged in size from 40 to 75 students. No significant difference was found when testing recognition, understanding, and simple application. However, a significant difference in favor of the large class size was found when testing for complex application. But unexplained was a more positive attitude exhibited by the students from the small classes.

More recently, in a paper by Adams, Curtis, and Britton (1984) presented to the Southwestern Economics Association,
achievement results comparing large and small introductory economics principles classes were described. Measuring achievement by the Revised Macro Test on Understanding College Economics, a significant difference resulted in favor of large class size. A mixed conclusion was obtained for business management classes comparing an instructor's small class of 50 students with his other class of 196 students. Administering three periodic objective tests to both classes, the first two examinations resulted in higher scores for the large class while the small class scored higher on the final examination. Although it was anticipated that homework participation rates in the small class would be higher, no significant difference was found. However, the drop rate in the large class was significantly higher.

A study by Nachman and Opochinsky (1958) is perhaps worthy of special mention. Concerned about the measurement process for testing achievement, they conducted an experiment comparing two matched general psychology classes of sizes 21 and 140. They hypothesized that the degree of achievement is often masked by use of the traditional achievement measures and could account for the lack of significant differences in class size experiments. On surprise quizzes for which students were not allowed time to prepare, the small class performed significantly higher than the large class. Yet no significant difference was found on
examinations when testing the content covered by the surprise quizzes. They concluded that when given time to prepare, students are often able to compensate for the large class effect.

In 1970 Silver (1970) reported a significant difference in favor of large classes of English. Four small classes of size 30 to 35 were compared with two large classes of approximately 100 students each. The achievement measure used was the Iowa Test of Educational Development which was administered at the beginning and end of the semester. Moreover, no significant difference was found in the completion rates for the two class sizes.

No significant difference was found by Chrisp and Aven (1970) in a study involving 111 prospective elementary school teachers. The small class consisted of 40 students, while the large class was composed of 71 students. The two classes were tested for initial capability at the beginning of the semester and found to not be significantly different. The completion rates were also found to be comparable at 87.5 percent for the small and 90 percent for the large class of students. The achievement measure used was an objective final examination. The same conclusion was obtained by Feldhusen (1963) in a study involving educational psychology students and by Moore (1977) in an experiment comparing two sections of introductory psychology students. In each case both sections being compared were
taught by the same instructor, but there was no allowance made for initial ability.

Using an analysis of covariance for comparing the achievement of three pairs of large and small classes of general psychology students, Eash and Bennett (1964) found no significant difference. The large classes consisted of between 167 and 208 students while the small classes ranged in size from 31 to 46. The same instructor taught the classes that were being compared and an objective final examination was used as the achievement measure. However, the large classes were allowed the opportunity to participate regularly in small discussion groups of approximately 15 members each.

In an attempt to integrate class size research, Glass and Smith (1979), two of the most prominent reviewers of class size research, used a technique called meta-analysis to combine statistically the findings of 77 studies that represented data on nearly 900,000 students. However, the studies were restricted to a maximum class size of approximately 30 students with less than three percent of the sample involving students of age 19 or older.

Williams, David, Cook, and Jensen (1985) attempted to determine whether the meta-analysis findings by Glass and Smith apply to university settings by using the archives of test data at a large university. Since students at the university studied were required to pass several
examinations on a variety of topics to meet specified general education requirements and since courses designed to prepare students for these exams were taught in sections that vary widely in size, the records of student performance on these examinations were used to study the relationship between test performance and class size.

Sections ranging in size from 13 to 1006 students were included in the study. Only those sections using a lecture format and a common test across course sections were used. The final examination was used almost exclusively as the measure of achievement for courses covering the following 16 content areas: accounting, business management, child development and family relations, computer science, communications, economics, English, food science and nutrition, health, physical science, physics, religion, social science, statistics, theater, and cinematic arts. A total of 16,230 tests were analyzed for eight ranges of the class size variable: 13-20, 13-30, 13-40, 20-40, 30-40, 13-1006, 30-1006, and 40-1006.

Williams et al. (1985) concluded from their analysis of the archives data that "It appears that increasing class size from current levels of 30 to 40 or more, up to several hundreds, may not radically affect college student achievement" (p. 315). They reasoned that it is possible that class size has little effect on lower-level educational outcomes, but that there may be an impact on higher-level
educational outcomes such as critical thinking or problem-solving skills.

Three studies dealing with mathematics courses were found. One study, prompted by a university faced with a 60 percent increase in the number of student credit hours in mathematics courses, experimented with the use of large sections of a remedial algebra course (Simmons, 1959). Using 12 faculty members to teach 14 small sections of mean size 22 and only one faculty member to teach three large sections of mean size 85, results of the course grades were used to measure achievement. The results of an analysis of covariance revealed a significant difference between 200 students selected from the small classes and 200 students selected from the large classes in favor of the small class size. The drop rates for the two groups were also found to be significantly different. The combined drop and failure rate for the 200 students from the small classes was 39 percent as compared to 55 percent for the 200 students selected from the large classes.

Edgell (1981) also conducted a study involving four small classes and two large classes of remedial algebra students. The small classes ranged in size from 30 to 40 students whereas the two large sections were of sizes 121 and 129. The small sections were taught by graduate students and both large sections were taught by one regular faculty member. A common objective final examination
developed by a mathematics departmental committee was used as the measure of achievement. The results analyzed were based on scores from 61 percent of the students from the small sections and 54 percent of those in the large sections. The analysis of these scores resulted in no significant difference.

A third study, reported by Vest, Nunley, and Garner (1981), compared the achievement of students from four small sections of a mathematics course for elementary teachers with those in one large section. The small sections had a total of 70 students and the large section included 98 students. Using the Otis Quick-Scoring Mental Ability Test, the students were subgrouped into two groups and a 2x2 analysis of variance was used to analyze the results of the final examination. No significant difference was recorded for this achievement measure.

Research of Class Size and Student Opinion

Hudelson (1928) reported the results of a three-year experimental program undertaken by the University of Minnesota. Considered by Nelson (1959) to be the best and most comprehensive of the early class size studies, it involved 108 sections of courses covering a wide range of subjects and taught by 21 volunteer faculty members. Students and faculty preferred the small class over that of the large by 88 percent and 86 percent respectively. The experiment by Edmonson and Mulder (1924) also indicated a
preference by students for small classes over large ones.

Recently an experiment involving business and economics students by McConnell and Sosin (1984) was reported. From an initial enrollment of 1399 students, survey results of 961 students from two freshman-level, three sophomore-level, and three junior-level classes ranging in size from 140 students to 239 students were obtained. Individual class survey rates varied from 60 percent to 80 percent. Overall the results revealed a negative response to large classes. In particular, majors were found to be more discontent than non-majors, females more than males, and students with high grade point averages were more discontent than those with low grade point averages. Student opinions were also found to vary by the subject matter of the classes.

In a study by Macomber and Siegel (1957), as described above in the research of class size and achievement section of the literature synthesis, it was found that students generally preferred small to large classes. This study confirmed the results for high and low-ability students by McConnell and Sosin: the high-ability students were found to be more unfavorable toward large classes than were those of low-ability. The Macomber and Siegel results, although consistently unfavorable toward large classes, indicated a wide degree of unfavorability ranging form "... slight to considerable" (p. 225). However, students were found to prefer an instructor of known excellence with unknown
section size to that of a small section with an unknown instructor. Thus, the instructor factor is perceived by students as more important than class size. Vest et al. (1981) also found that 84 percent of the surveyed students participating in their class size experiment preferred a large section over a smaller one when the large section is taught by an instructor perceived as excellent and the small section is taught by an unknown instructor.

An experiment by Moore (1977) underscores the significance of the instructor factor. Large classes were unused at the university where the experiment was held and the students were registered for the experimental section before the decision was made to establish a large class using several small sections. Using a 14 statement Likert-type scale to measure student attitude toward large classes, pretest and posttest scores were compared. The attitude scores were found to reverse from a negative to a very favorable response toward large class size. The researcher interpreted this reversal in attitude as an indication that the quality of instruction is more significant to the students than the class size.

The Large Class Analysis Project (1982), conducted by the Center for Teaching Effectiveness at University of Texas at Austin, was designed to study teaching techniques used in a variety of large classes in order to help with the teaching and management of large classes. Nineteen classes
from the Colleges of Natural Science, Liberal Arts, Engineering, and Business were selected; the 19 faculty members were obtained on a volunteer basis. Of the 3,820 students enrolled, 56 percent responded to the post-semester survey.

Both students and faculty showed a preference for small classes. The reasons given for this preference for small classes were: The environment is more conducive to learning, students are able to interact with the instructor on a personal basis, students are able to know each other on a more personal basis, and students have more control over when they will participate in class. The reasons given by students preferring a large class size were: The instructor is more organized and qualified to teach, and the atmosphere is more casual and relaxed. For all but one of the instructors, the students indicated that they enjoyed the class less at the end of the semester. In particular, the students in the Colleges of Engineering and Business enjoyed their classes less than did those in Natural Science and Liberal Arts. The survey indicated no significant difference in attitude toward large classes due to gender or earned grade in the course, yet there was a more positive attitude by students in the non-required courses.

In a 1978 review, Feldman reported that approximately one-third of the studies reviewed indicated no relationship between class size and student ratings of the instructor.
The remaining studies implied a negative correlation; that is, the smaller the class size, the higher the student ratings. Yet another analysis of the studies indicated that the relatively small and relatively large classes tended to receive higher rankings than did the medium sized classes. But, when considering all studies reviewed, it is much more likely that a researcher will find an association between class size and student ratings of the instructor, and when such a relationship is found, it is most likely to be an inverse association.

In a more recent review, Feldman (1984) extended his topic of review to include course evaluations by the students. He concluded that a very weak inverse relationship exists between class size and overall student evaluation of the course and instructor. He also concluded that an inverse relationship between class size and the student evaluations of specific instructional dimensions pertaining to the instructor's interactions and interrelationships with students is typical. However, Kulik and McKeachie (1975) concluded from their review that student ratings of instructor and course may vary with the course taught. Moreover, there is a high correlation between student ratings and the experience of the instructor.

The Macomber and Siegel (1957) studies revealed lower course and instructor student evaluations for large classes
than for small classes. Adams and Britton (1984), Rohrer (1957), and Splaine (1980) reported no significant difference in student course evaluations. Sweeney, Siegfried, and Raymond (1983), and Wilkinson (1958) surveyed senior economics majors at 485 colleges and universities and reported a curvilinear relationship between class size and student course evaluations. Less than 35 students, between 35 and 125 students, and more than 125 students determined small, medium, and large class sizes for this study. The percentage of students ranking a course as being above average was 48 percent, 43 percent, and 64 percent respectively for these three class sizes of small, medium, and large.

Findings similar to Sweeney et al. (1983) have been found in studies of instructor evaluations by students. Gage (1961), Marsh, Overall, and Kesler (1979), and Wood, Linsky and Straus (1974) were studies confirming this curvilinear relationship. Wood, Linsky, and Straus theorize that the best instructors are sought to teach the very large classes while the problems that force the best instructors to be chosen for the very large classes have not become massive enough to force a change in instructor selection for large classes. However, a study by Crittenden, Norr, and LeHailly (1984), using 981 undergraduate classes categorized into eight class sizes, found no evidence of curvilinearity. Rather, they found a
smooth decline in mean ratings as class size increased through the lowest three categories to about 100 students. Then, as class size increased beyond this point, a declining trend continued, although the decline was not entirely smooth. The researchers offered several possible explanations for the contradictory findings:

(1) Some studies are based on a small number of classes.

(2) No general agreement on the definition of large class size has been obtained.

(3) Students may make allowance for large class size on their evaluations.

(4) Some universities carefully select instructors for large classes.

Studies by Eash and Bennett (1964), Hamilton (1980), Macomber and Siegel (1957), and Thibodeaux and Zuzan (1984) found no significant difference of class size effect on student evaluations of instructors. However, Haslett (1976), in a university-wide study, and Vest, Nunley, and Garner (1981) found the student ratings of instructors to be inversely related to class size. Whitten and Umble (1980) reported an inverse relationship, but they also found the instructor ratings for core courses to be significantly higher than those for non-core courses.

Summary

It is clear that students and teachers prefer small
classes, but any generalizations, concerning the relationship between class size and course or instructor evaluations by students appear unwarranted. It is also clear that most of the class size experiments find no significant difference for achievement when an objective final examination is used as the measure of achievement. However, there has been little research that uses tests for problem-solving or analytic skills as the measure of achievement.

Historically, most class size experimentation has been done in the disciplines of education and psychology with some of the more recent studies involving business or economics. Large and small class sizes have not been standardized and add to the degree of difficulty in evaluating the results of class size studies.

It appears that if the objective examination is the sole measure of achievement, then there is little, if any, association between class size and achievement. Also, since the final examination is the most common achievement measure used, the implications of class size studies are, for the most part, relegated to those students completing the course of study and are not applicable to the initial sample of students.
CHAPTER BIBLIOGRAPHY


Francisco: Far West Laboratory for Educational Research and Development.


Simmons, H. F. (1959). Achievement in intermediate algebra associated with class size at the University of Wichita. 
College and University, 34, 309-315.


University of Texas at Arlington (1986). Undergraduate catalog. (Available from Office of Admissions, University of Texas at Arlington, Arlington, 76019)


CHAPTER III

PROCEDURES FOR COLLECTION OF DATA

Introduction

This study was conducted at the University of Texas at Arlington (UTA). Arlington has a population of approximately 250,000 and is located midway between Dallas and Fort Worth. UTA is a comprehensive research university offering an array of baccalaureate, masters, doctoral, and special professional degrees and has an approximate enrollment of 23,000 students of which 85 percent are undergraduates. Nearly 70 percent of the students are over 22 years of age with an average age of 25 years, 55 percent of the students are full-time, and 57 percent are male. One-third of the undergraduate students are enrolled in the College of Business Administration, 29 percent in the College of Liberal Arts, 16 percent in the College of Engineering, and 12 percent in the College of Science. The Colleges of Architecture, Nursing, and Urban Studies account for the remaining 10 percent of the undergraduate enrollment.

Population and Sample Selection

This quasi-experimental study was composed of students enrolled in four sections of first-semester calculus during
the fall semester of 1987. This course is required for students majoring in the following disciplines: architecture, biology, chemistry, computer science, engineering, geology, mathematics, physics, predental, premedicine, prepharmacy, and psychology (University of Texas, 1986). Typically, 600 students register for first-semester calculus during the fall semester with approximately one-half of these enrolling in a lecture course. Moreover, the majority of the students enrolling in the lecture course are freshmen. However, approximately 25 percent are at or above the junior level. Those students not enrolling in the lecture course, primarily engineering students, take a combined course of analytic geometry and calculus utilizing televised instruction. The latter group did not participate in this study.

The fall semester classes met for 16 weeks and were chosen because it was thought to provide a less restrictive sample than those enrolling during the spring and summer semesters due to the sequential nature of the first three basic calculus courses. One small and one large section were offered at each of two different times before noon on a Monday-Wednesday-Friday basis and also on a Tuesday-Thursday basis. The large and small Monday-Wednesday-Friday sections met at 9:00 a.m. and 11:00 a.m. respectively for 50 minutes while the small and large Tuesday-Thursday sections met at 9:30 a.m. and 11:00 a.m. respectively for 80 minutes. The
enrolling students selected their class section, but neither the instructor nor the size of the class were known to the students during registration. This arrangement is standard for courses below the junior level. Moreover, the plans for the experiment were known only to the associate chair, who is responsible for class scheduling, and the two participating instructors. The small sections were restricted to a maximum of 30 students while the large sections were restricted to 90 students each. The Calculus Readiness Test (CR) was administered during the first week of the semester to aid in adjusting for possible significant variation in student ability between the classes.

Research Design

Large sections of mathematics have been taught at UTA for nine years and gradually its use has been extended to all levels of the undergraduate program. Two tenured instructors, experienced in teaching large sections of calculus, were each assigned one small and one large section that met on the same day. Each of these instructors has taught mathematics at all levels of the undergraduate program at the participating university for between 20 and 25 years and each has taught large classes for six or more years. One of the instructors has a Ph.D. in mathematics while the other is completing a dissertation for his Ph.D. in college teaching with a minor in mathematics.

The research design utilized was the quasi-experimental
nonequivalent control group design with class size as the independent variable and achievement and student opinion as the dependent variables. Borg and Gall (1983, p. 683) state that the main threat to the internal validity of this design is the possibility that group differences on the dependent variable are due to preexisting group differences rather than to the effect of the independent variable and that the analysis of covariance may be used to aid in the management of this threat. The CR was used as the covariate for this purpose.

Campbell and Stanley (1963, p. 48) suggest that the most common threat to the internal validity of nonequivalent control group experiments is due to the interaction of selection and maturation. They (p. 50) suggest, however, that unless the experimenter has reason to suspect differential recruitment related to the treatment, then the interaction of selection and maturation may be assumed to be controlled to a relatively high degree. It was assumed that this interaction was controlled by the natural conditions of the registration process and the procedure outlined for the collection of data. Campbell and Stanley (p. 48) also state that regression is a possible threat to internal validity. This possibility was controlled by the use of covariate analysis for data interpretation.

Interaction of testing and treatment is the greatest threat to external validity (Campbell & Stanley, 1963, p.
In this experiment, however, the CR is not viewed as having an effect of sensitization on either of the dependent variables. Campbell and Stanley (1963, p. 40) also list interaction of selection and treatment and reactive arrangements as possible external validity threats. The interaction of selection and treatment should be minimal due to the lack of disruption to the normal environmental setting. Care was also taken to generalize experimental results only to those of a similar university. Reactive arrangements were assumed to be controlled due to the natural setting of the experiment.

The Treatment

Daily problem assignments were prepared by the participating instructors in an attempt to standardize content objectives and were followed in all classes (see Appendix A). Four periodic tests were given along with the Cooperative Mathematics Test, Calculus (CMTC). The CMTC was used for the comprehensive final examination. Final grades were based solely on the average of the four periodic tests and the final examination. W was used to represent the course grade of any student failing to receive a course grade of A, B, C, D, or F. Excused absences were accepted and all missed tests were made up by the student.

Each section was taught in a classroom designed for its particular size. Lighting, sound, and visibility were
adequate for each of the classrooms selected. The instruction in both large and small classrooms utilized the chalkboard exclusively and did not require any amplification equipment. Each of the participating instructors used the lecture-discussion method of teaching in both large and small classes and conducted the classes as nearly identical as possible with no special equipment or preparation utilized for large classes.

During the first week of the semester, the following information was obtained from each student: age, gender, one-way distance traveled (if commuting), high school and college mathematics courses completed, major, classification, and, if employed, how many hours per week were spent at work (see Appendix B).

During the last two weeks of the semester, the semantic differential was used to measure each of the following opinions: (a) student evaluation of the instructor, (b) student evaluation of the course, and (c) student evaluation of the classroom learning environment (see Appendices C, D, and E).

**Delimitations**

This study was delimited by the following factors:

1. The use of intact groups were Calculus I sections formed during the regular registration process at UTA.
2. The use of students were from only one university.
3. The non-random nature of the sample.
Description of the Instruments

The semantic differential was used to measure the opinions of the students toward their instructor, course, and classroom learning environment. According to Nunnally (1970, p. 440), the evaluative factor almost serves to define opinion, and scales on the evaluative factor should serve as a valid measure of verbalized opinions. He (p. 443) states that the semantic differential is probably the most valid measure available for measuring the connotative meaning of opinions.

Osgood, Suci, and Tannenbaum (1957, pp. 193-195) report comparative studies indicating that the evaluative factor of the semantic differential measures "the same thing" that the Thurstone and Guttman scales measure "to a considerable degree". They conclude that the evaluative factor of the semantic differential is a valid general measure of verbalized opinions. Furthermore, they state that if care is given to selecting evaluative scales, then high correlations with standard attitude-measuring instruments would be expected.

The construction process followed for the development of the three semantic differentials used in this experiment was that suggested by Osgood, Suci, and Tannenbaum and detailed by Kerlinger (1964, pp. 567-571) with bright-dark, fragrant-foul, and pungent-bland being used for distractors. According to McCallon and Brown (1971, p. 70), unlike the
construction of Likert-type instruments, elaborate item analysis procedures and repeated revisions of the semantic differential instrument are not necessary; this constitutes a major advantage of the semantic technique.

The CR was administered at the beginning of the course for use as the covariate in the analysis of covariance to be performed on each of the four periodic tests, the final examination, and the final grade averages. The CR is a 20-question test of knowledge and skills needed by college students before entering first-semester calculus. Since 1977 it has been one of six tests for college students developed and published by the Mathematical Association of America (Mathematical Association, 1984, p. 3).

The CR developers contend that carefully devised procedures were followed to produce a test which will exhibit high content validity (Mathematical Association, 1984, pp. 10-11). Further, they (p. 12) contend that the construction procedures and extensive statistical analyses of items give the test a high level of reliability. A letter from John G. Harvey (see Appendix F), one of the directors of the testing program, states that the actual data on the completed form of the CR were lost. He, however, vouches for it as being a valid and reliable test.

Form A of the Cooperative Mathematics Test, Calculus (CMTC) was used for the final examination. Robinson (1972, p. 926) states that CMTC is a good achievement test for a
beginning college course in calculus and that it maintains a balance between theory and application. Kline (1972, p. 926), after reviewing the test, concluded that he knows of no standardized test designed to measure achievement in calculus that is better than CMTC. Moreover, the topics tested are those included in most first-year calculus courses and are weighted closely to the relative emphasis on the topics found in the popular textbooks. The Kuder-Richardson Formula 20 reliability coefficient for form A of CMTC was reported to be .87 in the testing of college students and the mean biserial correlation was .44 (Educational Testing Service, pp. 62-63). In summary, Kline (1972, p. 925) states that the CMTC provides a reliable score that is valid for measuring achievement of beginning college calculus.

Summary

General characteristics of the participating university and procedures for the collection of data were presented in Chapter III. Each of two tenured faculty members, experienced with teaching large calculus sections, taught one small and one large section of first-semester calculus. The CR was administered during the first week of classes for use as a covariate and four periodic tests were given to each pair of small and large sections. The CMTC was given at the end of the semester and used for the comprehensive final examination in all sections. During the last week of
the semester, three semantic differentials were administered to measure student opinion of the instructor, course, and classroom learning environment.

Each of the instruments used in the experiment was described and the validity and reliability of the instruments were documented. The quasi-experimental nonequivalent control group design was selected for this study; threats to the internal and external validity of this design were discussed and judged to be controlled.
CHAPTER BIBLIOGRAPHY


University of Texas at Arlington (1986). *Undergraduate catalog*. (Available from Office of Admissions, University of Texas at Arlington, Arlington, 76019)
CHAPTER IV

PRESENTATION AND ANALYSIS OF THE DATA

This chapter presents a description of the sample and the findings of the procedures utilized to test each of the four hypotheses. The chapter concludes with a listing of 12 findings of this study.

Description of the Sample

The data were collected from August 31 to December 17, 1987, during the fall semester at the University of Texas at Arlington. One instructor taught one large and one small section of first-semester calculus at 9:00 a.m. and 11:00 a.m., respectively, on a Monday-Wednesday-Friday schedule. The large section had an initial enrollment of 84 students while the small section contained 25 students. Each of these sections met for a 50-minute period. A second instructor taught one small and one large section of first-semester calculus at 9:30 a.m. and 11:00 a.m., respectively, on a Tuesday-Thursday schedule. Each of these sections met for 80 minutes. These two sections had initial enrollments of 29 and 87 students.

During the first week of the semester, scores from the CR and demographic data from the information sheet shown in Appendix B were obtained from 225 students in the four
first-semester calculus sections. These students were used as the sample for the investigation of final grades and the CR scores were used as the covariate in each analysis of covariance.

Of the 225 students involved in the study, 69.8 percent were male. Student age ranged from 17 to 47 years with 41.3 percent in the 17 to 19 age group, 30.4 percent in the 20 to 22 age group, and 28.4 percent above age 22. The average age was 21.4 years. Seventy-six percent of the students indicated that they were employed and only 6.2 percent of this group worked less than 15 hours per week. The average number of hours worked per week was 27.5.

Forty-seven percent of the students were found to live within a five-mile radius of the campus and 33.7 percent commuted at least 16 miles one-way to class on a daily basis. The number of science majors dominated that of engineering majors by 48 percent to 27 percent with 25 percent majoring in other academic areas. Approximately 55 percent of the students were freshmen, 22 percent sophomores, and 24 percent junior-level or above. Tables 1 and 2 contain more specific information for the demographic variables as reported by a limited portion of the total sample. This information is categorized for the sample and for each of the four sections of first-semester calculus separated by the large and small sections of each instructor.
Table 1
Description of the Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total</th>
<th>Instructor 0</th>
<th>Instructor 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>S (%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>68</td>
<td>30.2</td>
<td>31.0</td>
</tr>
<tr>
<td>Male</td>
<td>157</td>
<td>69.8</td>
<td>69.0</td>
</tr>
<tr>
<td>Age in years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-18</td>
<td>43</td>
<td>20.4</td>
<td>31.0</td>
</tr>
<tr>
<td>19</td>
<td>44</td>
<td>20.9</td>
<td>20.7</td>
</tr>
<tr>
<td>20</td>
<td>32</td>
<td>15.2</td>
<td>10.3</td>
</tr>
<tr>
<td>21-22</td>
<td>32</td>
<td>15.2</td>
<td>13.8</td>
</tr>
<tr>
<td>23-25</td>
<td>28</td>
<td>13.3</td>
<td>13.8</td>
</tr>
<tr>
<td>26-30</td>
<td>22</td>
<td>10.4</td>
<td>10.3</td>
</tr>
<tr>
<td>&gt; 30</td>
<td>10</td>
<td>4.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Employment in hours/week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>59</td>
<td>28.2</td>
<td>28.6</td>
</tr>
<tr>
<td>&lt; 15</td>
<td>13</td>
<td>6.2</td>
<td>21.4</td>
</tr>
<tr>
<td>15-20</td>
<td>52</td>
<td>24.9</td>
<td>10.7</td>
</tr>
<tr>
<td>21-30</td>
<td>38</td>
<td>18.2</td>
<td>21.4</td>
</tr>
<tr>
<td>&gt; 30</td>
<td>47</td>
<td>22.5</td>
<td>17.9</td>
</tr>
</tbody>
</table>

Note. S = small section; L = large section.
Table 2

Description of the Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total</th>
<th>Instructor 1</th>
<th>Instructor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>S (%)  L</td>
<td>S (%)  L</td>
</tr>
<tr>
<td>Commuting in miles one-way</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 6</td>
<td>97</td>
<td>46.6 51.9</td>
<td>65.2 37.2</td>
</tr>
<tr>
<td>6-15</td>
<td>41</td>
<td>17.7 20.3</td>
<td>17.4 25.6</td>
</tr>
<tr>
<td>16-25</td>
<td>39</td>
<td>18.8 28.6</td>
<td>4.3 23.1</td>
</tr>
<tr>
<td>&gt; 25</td>
<td>31</td>
<td>14.9 12.7</td>
<td>13.0 14.1</td>
</tr>
<tr>
<td>Classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>123</td>
<td>54.7 56.3</td>
<td>52.0 57.1</td>
</tr>
<tr>
<td>Sophomore</td>
<td>49</td>
<td>21.8 20.7</td>
<td>20.0 20.2</td>
</tr>
<tr>
<td>Junior or above</td>
<td>53</td>
<td>23.6 23.0</td>
<td>28.0 22.6</td>
</tr>
<tr>
<td>Major area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>108</td>
<td>48.0 51.7</td>
<td>24.0 50.0</td>
</tr>
<tr>
<td>Engineering</td>
<td>61</td>
<td>27.1 27.6</td>
<td>48.0 25.0</td>
</tr>
<tr>
<td>Liberal arts</td>
<td>22</td>
<td>9.8 11.5</td>
<td>8.0 9.5</td>
</tr>
<tr>
<td>Other</td>
<td>34</td>
<td>15.1 13.8</td>
<td>20.0 15.5</td>
</tr>
</tbody>
</table>

Note. S = small section; L = large section
Findings

The first hypothesis was: There is no significant difference in achievement between students in large and small first-semester calculus classes. To test this hypothesis the CR was given during the first week of class to each of the students participating in the four first-semester calculus sections. These scores were used as the covariate in each analysis of covariance that was applied to each of six achievement measures used to test hypothesis one.

The first four of the six achievement measures were periodic tests given approximately at the end of each quarter of the semester. The CMTC, a comprehensive final examination, was given during final examination week and served as the fifth achievement measure. The final grade averages were used as the sixth achievement measure. These final averages were obtained by averaging the periodic tests and the CMTC, with the CMTC weighted as the equivalent of two periodic tests. The analysis of the findings for each of the six achievement measures will be detailed.

First, as a preliminary to using the analysis of covariance, a scatter plot of the CR scores and the first periodic test scores was obtained for each of the four sections of first-semester calculus using a SAS plot computer program (SAS, 1986). Each of the scatter plots was judged to satisfy the linearity assumption needed as a
A prerequisite for application of the analysis of covariance (Huck, 1974, p. 144) to the periodic test one scores from each pair of large and small sections of first-semester calculus.

A second basic assumption necessary for the application of the analysis of covariance is that the regression lines of each pair of small and large sections have a common slope (Ferguson, 1981, pp. 361-362). Using the test for parallelism of slopes, the slopes for both pairs of small and large sections were calculated (BMDP, 1987). The small and large sections of instructor 0 had regression line slopes of 5.06 and 3.92, respectively. The regression line slopes of the corresponding pair of slopes for those sections taught by instructor 1 were 2.21 and 2.73. In both cases the slopes were not significantly different at the .05 level. Hence, the null hypotheses for equal regression slopes was accepted at the .05 level for each pair of regression lines.

According to Ferguson (1981, p. 245), homogeneity of group variance about the regression line is also an assumption which needs to be checked before using an analysis of covariance. The statistical test utilized was to test the hypothesis that the variances about the regression lines for each pair of small and large sections were equal.

For the pair of small and large sections of instructor
Mean square estimates of the regression variances of 452.62 and 379.25 were obtained, respectively, with corresponding degrees of freedom of 24 and 79. An experimental $F$-value of 1.19 was evaluated and found to be between the critical lower and upper limits of .49 and 1.82 at the .05 level (Ostle & Mensing, 1975, p. 547). Mean square estimates of 319.09 and 339.52 were obtained for the other pair of small and large sections, respectively, with corresponding degrees of freedom of 19 and 76. The experimental $F$-value, .96, fell within the critical limits of .45 and 1.94 at the .05 level (Ostle & Mensing, 1975, p. 557). Hence, homogeneity of group variance about the regression line was accepted.

Having found the assumptions of linearity, common regression slopes, and homogeneity of variance to be satisfactory, an analysis of covariance was applied to the adjusted means of periodic test one scores for each pair of small and large classes using CR as the covariate. Presented in Table 3 is a summary of this analysis computed using the statistical procedure contained in BMDP Statistical Software (1987).

The critical $t$-value at the .05 level for the two sections of instructor 0 was 1.99 using 96 degrees of freedom (Ferguson, 1981, p. 521). The experimental $t$-value was .14 (BMDP, 1981). The critical $t$-value of 1.99 was also calculated for the other pair of small and large sections at
Table 3

Analysis of Covariance Summary for Periodic Test One

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Adjusted mean</th>
<th>CR mean</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>21</td>
<td>55.38</td>
<td>59.31</td>
<td>10.57</td>
<td>0.14</td>
</tr>
<tr>
<td>Large</td>
<td>78</td>
<td>60.99</td>
<td>59.93</td>
<td>11.78</td>
<td></td>
</tr>
<tr>
<td>Instructor 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>27</td>
<td>64.07</td>
<td>59.07</td>
<td>13.96</td>
<td>1.48</td>
</tr>
<tr>
<td>Large</td>
<td>81</td>
<td>64.29</td>
<td>65.96</td>
<td>11.40</td>
<td></td>
</tr>
</tbody>
</table>

the .05 level using 105 degrees of freedom (Ferguson, 1981, p. 521). The corresponding experimental t-value was 1.48. Hence, for both pairs of small and large classes the adjusted means of the periodic test one scores were not significantly different at the .05 level.

The small section of instructor 0 had a CR mean of 10.57, whereas the CR mean for the large section was 11.78. These means account for the upward adjustment of the mean for the small section and the downward adjustment of the mean for the large section. The reverse was true for the two sections of instructor 1. In this case the small section had a CR mean of 13.96 while the large section had a CR mean of 11.40. Hence, the mean of the small section was
adjusted downward and the large section adjusted upward. For both instructors the unadjusted and adjusted means were larger for the large sections but the differences were not significant in either case. Therefore, hypothesis one was accepted at the .05 level for the first of four periodic tests used as a measure of achievement.

Next, the second achievement variable, periodic test two, and CR, the covariate, were given the preliminary analysis necessary for application of the analysis of covariance. A scatter plot of the CR scores and periodic test two scores was obtained for each of the four sections using a SAS plot computer program (SAS, 1986). Each of these scatter plots was judged to satisfy the linearity assumption (Huck, Cormier, & Bounds, 1974, p. 144).

Using the test for parallelism of regression line slopes, the regression line slopes for both pairs of small and large sections were calculated (BMDP, 1987). The small and large sections of instructor 0 had slopes of 2.12 and 2.43, respectively, while the slopes for the other corresponding pair of sections were 3.35 and 3.84, respectively. A significant difference was not found in either case at the .05 level.

The final preliminary, homogeneity of group variance about the regression line, was considered. For the small and large sections of instructor 0 mean square estimates of the regression variances of 362.69 and 321.62, respectively,
were calculated (SAS, 1986). The corresponding experimental F-value of 1.13 was calculated using degrees of freedom of 20 and 70, respectively, and found to be between the critical limits of .48 and 1.82 at the .05 level (Ostle & Mensing, 1975, p. 557). Mean square estimates of 808.38 and 508.71 were obtained for the other pair of small and large sections, respectively, with corresponding degrees of freedom of 23 and 67. The experimental F-value, 1.59, fell within the critical limits of .48 and 1.85 at the .05 level. Therefore, homogeneity of group variance about the regression line was accepted.

Having satisfied the three preliminary assumptions, an analysis of covariance was performed on the periodic test two scores for each pair of small and large sections using CR as the covariate. Presented in Table 4 is a summary of the analysis produced by the statistical procedure contained in BMDP Statistical Software (1987).

The critical t-value at the .05 level using 91 degrees of freedom was 1.99 (Ferguson, 1981, p. 521). The corresponding experimental t-values were 1.52 and 1.81 for the pair of adjusted means being tested (BMDP, 1987). In both comparisons the adjusted means were not significantly different at the .05 level. Hence, hypothesis two was supported at the .05 level for the achievement variable, periodic test two. The unadjusted and adjusted means, however, were greater for the large sections as was
Table 4
Analysis of Covariance Summary for Periodic Test Two

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Adjusted mean</th>
<th>CR mean</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>22</td>
<td>63.91</td>
<td>65.97</td>
<td>10.86</td>
<td>1.52</td>
</tr>
<tr>
<td>Large</td>
<td>72</td>
<td>73.37</td>
<td>72.75</td>
<td>12.00</td>
<td></td>
</tr>
<tr>
<td>Instructor 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>25</td>
<td>61.60</td>
<td>55.57</td>
<td>14.00</td>
<td>1.81</td>
</tr>
<tr>
<td>Large</td>
<td>69</td>
<td>63.87</td>
<td>66.05</td>
<td>11.78</td>
<td></td>
</tr>
</tbody>
</table>

the case in the comparison of means of the periodic test one scores. In all four sections the CR means were higher for the students taking the second periodic test than for those students taking the first periodic test.

The third of the four periodic tests was given and the preliminary checks for linearity, common slope, and homogeneity of variance about the regression lines were performed. The scatter plots of the periodic test three scores with the CR scores for each of the four first-semester calculus sections were obtained (SAS,1986) and were judged to satisfy the linearity assumption required for application of an analysis of covariance for adjusted means of periodic test three scores from each pair of the small
and large sections.

Then, after computation of the slopes of the regression lines for each pair of small and large sections was computed (BMDP, 1981), the test for parallelism of slopes was applied. The computed regression slopes of the small and large sections of instructor 0 were 3.34 and 2.78, respectively. The other corresponding pair of regression slopes were 1.93 and 2.23. Neither pair was found to differ significantly at the .05 level.

\( F \) -values for each pair of the large and small sections were calculated by evaluating the quotients of the mean squares (SAS, 1981) in order to test for homogeneity of variance about each pair of regression lines. \( F \) -values of 1.39 and 2.54 with corresponding degrees of freedom of 14, 61 and 22, 58 were calculated for the small and large sections of instructors 0 and 1, respectively. The first \( F \) -value was well within the critical limits of .378 and 2.02 at the .05 level (Ostle & Mensing, 1975, p. 545), whereas the second \( F \) -value, 2.54, was outside the critical lower and upper limits of .464 and 1.92 at the .05 level. Ostle and Mensing (1975, p. 353), however, indicate that lack of homogeneity of variances will have little effect on the usual tests and the resulting inferences. In particular, they state that the analysis of variance technique is quite robust and can be relied upon under most circumstances.

The analysis of covariance technique was applied to the
periodic test three scores for each pair of large and small sections using CR as the covariate (BMDP, 1987). A summary of this analysis is presented in Table 5.

Table 5
Analysis of Covariance Summary for Periodic Test Three

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Adjusted mean</th>
<th>CR mean</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Instructor 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>16</td>
<td>73.12</td>
<td>75.32</td>
<td>11.44</td>
<td>0.55</td>
</tr>
<tr>
<td>Large</td>
<td>63</td>
<td>73.65</td>
<td>73.09</td>
<td>12.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Instructor 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>24</td>
<td>74.67</td>
<td>71.51</td>
<td>13.92</td>
<td>-1.50</td>
</tr>
<tr>
<td>Large</td>
<td>60</td>
<td>79.38</td>
<td>80.65</td>
<td>11.92</td>
<td></td>
</tr>
</tbody>
</table>

The critical \( t \)-value was 1.99 for both comparisons of adjusted means (Ferguson, 1981, p. 521). The mean of the large section of instructor 0 was only slightly larger than the mean of the small class. When adjusted for the covariate the smaller class had a two-point higher mean, but the difference was obviously not significant as was indicated by the experimental \( t \)-value of .55 at the .05 level (BMDP, 1987).

Although the large section of instructor 1 revealed a mean 4.7 points higher for the large section and the
adjusted mean favored the large class by 9.1 points, a nonsignificant experimental $t$-value of -1.50 at the .05 level was obtained (BMDP, 1987). This rather large difference in the adjusted means was found to be highly related to the scores of two students in the small class. Each of these students made zeros on this test. When these scores were deleted and the mean and adjusted mean for the small class evaluated, the mean for the small class was 2.07 higher than that of the large class and the corresponding adjusted mean was only 2.46 points lower than that of the large class.

The fourth periodic test was given during the last week of the semester. For each of the four first-semester calculus sections scatter plots of the CR scores and the periodic test four scores were obtained (SAS, 1986) and judged to satisfy the linearity assumption. Then the regression line slopes for each pair of small and large sections was computed (BMDP, 1987) and the test for parallelism performed. The regression slopes of the small and large sections of instructor 0 were 3.47 and 1.40, respectively. The corresponding pair of slopes for the other two sections of Math 1326 were 3.35 and 2.56. Both pairs of slopes were nonsignificant at the .05 level.

Next, mean square estimates of the regression variances of 501.97 and 473.96 were calculated (SAS, 1986) for the small and large sections, respectively, of instructor 0.
The experimental $F$-value of 1.06 was obtained by calculating the quotient of these mean square estimates using degrees of freedom of 12 and 52, respectively. Then critical $F$-value limits of .35 and 2.20 were evaluated at the .05 level (Ostle & Mensing, 1975, p. 557). Mean square estimates of the regression variances of 481.45 and 369.57 were evaluated for the large and small sections of instructor 1 (SAS, 1986) providing an experimental $F$-value of 1.30 with 55 and 18 degrees of freedom, respectively. Using linear interpolation, critical $F$-value limits of .50 and 2.36 were calculated at the .05 level (Ostle & Mensing, 1975, p. 555). For both pairs of small and large classes the experimental $F$-value was inside the interval determined by the lower and upper critical $F$-values. Homogeneity of group variance about the regression line, therefore, was accepted.

Having satisfied the three preliminary assumptions of linearity, regression slope parallelism, and homogeneity of group variance about the regression line, an analysis of covariance was performed on the periodic test four scores for each pair of large and small sections using CR as the covariate. Presented in Table 6 is a summary of this analysis computed using the statistical procedure contained in BMDP Statistical Software (1987).

The critical $t$-value at the .05 level for both pairs of sections shown in Table 6 was 2.00 using 65 and 74 degrees of freedom, respectively (Ferguson, 1981, p. 521). The
### Table 6

#### Analysis of Covariance Summary for Periodic Test Four

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Adjusted mean</th>
<th>CR mean</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>14</td>
<td>72.14</td>
<td>73.86</td>
<td>11.86</td>
<td>.27</td>
</tr>
<tr>
<td>Large</td>
<td>54</td>
<td>76.11</td>
<td>75.67</td>
<td>13.00</td>
<td></td>
</tr>
</tbody>
</table>

Instructor 0

Small | 20 | 70.55 | 65.48 | 14.30 | .47   |
Large | 57 | 66.39 | 68.17 | 11.81 |       |

corresponding experimental t-values were .27 and .47, neither of which is significant at the .05 level. Hence, for both pairs of small and large sections the adjusted means of the periodic test four scores were not significantly different at the .05 level.

The comprehensive final examination, CMTC, was administered to the four first-semester calculus sections and a scatter plot of these scores with the CR scores was obtained (SAS, 1981) for each section. Each of these scatter plots was judged to satisfy the linearity assumption.

Using the test for parallelism of regression line slopes, the regression line slopes for both pairs of small
and large sections were calculated (BMDP, 1987). The small and large sections of instructor 0 had slopes of 2.11 and 2.19, respectively, while the slopes for the other corresponding pair of sections were 2.32 and 2.76, respectively. In both cases the differences were nonsignificant at the .05 level.

The final preliminary, homogeneity of group variance about the regression line, was checked. For the large and small sections of instructor 0 mean square estimates of the regression variances of 293.97 and 194.80, respectively, were calculated (SAS, 1986). The corresponding experimental F-value of 1.51 was calculated using degrees of freedom of 53 and 10, respectively, and found to be between the critical F-value limits of .43 and 3.21 at the .05 level (Ostle & Mensing, 1975, p. 553). Mean square estimates of 266.97 and 209.58 were obtained for the other pair of small and large classes, respectively, with corresponding degrees of freedom of 18 and 54. The experimental F-value, 1.27, fell within the critical F-value limits of .42 and 2.03 at the .05 level. Homogeneity of group variance about the regression line was accepted.

An analysis of covariance was performed on the final examination scores for each pair of small and large classes using CR as the covariate. A summary of the analysis produced by the statistical procedure contained in BMDP Statistical Software (1987) is presented in Table 7.
Table 7

Analysis of Covariance Summary for Final Examination

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Adjusted mean</th>
<th>CR mean</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>12</td>
<td>67.75</td>
<td>68.46</td>
<td>12.58</td>
<td>.03</td>
</tr>
<tr>
<td>Large</td>
<td>55</td>
<td>68.78</td>
<td>68.62</td>
<td>12.98</td>
<td></td>
</tr>
<tr>
<td>Instructor 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>20</td>
<td>68.40</td>
<td>63.80</td>
<td>14.30</td>
<td>1.48</td>
</tr>
<tr>
<td>Large</td>
<td>56</td>
<td>68.12</td>
<td>69.77</td>
<td>11.95</td>
<td></td>
</tr>
</tbody>
</table>

The critical \( t \)-value at the .05 level for the two sections of instructor 0 was 2.00 using 64 degrees of freedom (Ferguson, 1981, p. 521) and the corresponding experimental \( t \)-value was .03 (BMDP, 1981). The critical \( t \)-value of 2.00 was also calculated at the .05 level for the second pair of classes using 73 degrees of freedom (Ferguson, 1981, p. 521). The corresponding experimental \( t \)-value was 1.48. Hence, neither of the experimental \( t \)-values was significant at the .05 level. Therefore, the adjusted means of final examination scores were not significantly different at the .05 level.

The last of the six achievement measures, final grade average, was coordinated with the covariate, CR, for each of
the four first-semester calculus sections using four separate scatter plots (SAS, 1986) and judged to satisfy the linearity assumption. Then, using the test for parallelism of regression line slopes, the regression line slopes for each pair of small and large sections was calculated (BMDP, 1981). The small and large sections of instructor 0 had slopes of 1.45 and 1.77, respectively, while the slopes for the other corresponding pair of sections were 1.97 and 2.814, respectively. The difference in each pair of slopes was nonsignificant at the .05 level.

The final preliminary for an application of the analysis of covariance, homogeneity of group variance about the regression line, was considered. The mean square estimates of the regression variances for the large and small sections of instructor 0 were 231.28 and 124.10, respectively (SAS, 1986). The experimental $F$-value of 1.86 was evaluated using the quotient of the mean square estimates with the corresponding degrees of freedom, 53 and 10. The critical $F$-value limits of .43 and 3.21 were calculated at the .05 level (Ostle & Mensing, 1975, p. 557). Mean square estimates of the regression variances of 184.67 and 172.68 were obtained for the other pair of small and large sections, respectively, with corresponding degrees of freedom of 18 and 55. The respective $F$-value critical limits of .43 and 2.01 were calculated at the .05 level (Ostle & Mensing, 1975, p. 557). Each of these experimental
F-values fell within the corresponding pairs of critical limits at the .05 level. Homogeneity of group variance about the regression line, therefore, was accepted.

Having satisfied the three preliminary assumptions, an analysis of covariance was performed on the final grade averages for each pair of small and large sections using CR as the covariate. A summary of the analysis produced by BMDP Statistical Software (1987) is presented in Table 8.

Table 8
Analysis of Covariance Summary for Final Grade Averages

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Adjusted mean</th>
<th>CR mean</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>12</td>
<td>73.58</td>
<td>74.15</td>
<td>12.58</td>
<td>0.50</td>
</tr>
<tr>
<td>Large</td>
<td>55</td>
<td>71.95</td>
<td>71.82</td>
<td>12.98</td>
<td></td>
</tr>
<tr>
<td>Instructor 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>20</td>
<td>74.90</td>
<td>70.15</td>
<td>14.30</td>
<td>1.10</td>
</tr>
<tr>
<td>Large</td>
<td>57</td>
<td>72.42</td>
<td>74.09</td>
<td>11.86</td>
<td></td>
</tr>
</tbody>
</table>

The critical t-value at the .05 level for the two sections of instructor 0 was 2.00 using 64 degrees of freedom (Ferguson, 1981, p. 521). The critical t-value of 2.00 was also calculated for the other pair of small and large sections at the .05 level using 74 degrees of freedom.
(Ferguson, 1981, p. 521) with corresponding experimental $t$-value of 1.10 (BMDP, 1987). Since each of the experimental $t$-values was smaller than the corresponding critical $t$-value, the adjusted means of the final grade averages were nonsignificant at the .05 level. Hence, this sixth achievement measure is consistent with hypothesis one.

**Hypothesis Two**

Hypothesis two stated: The percentage of students in small first-semester calculus classes earning a course grade of $W$ or $F$ is not significantly different from that of students in large classes. The grades of $A$, $B$, $C$, $D$, and $F$ were used in the traditional way and $W$ represents the grade received by a student not completing the course. Logistic regression was utilized to test this hypothesis so that the possible effect of the covariate, $CR$, would not be lost.

Attrition rate is defined as the percentage of students in a given class receiving a course grade of $W$ or $F$. The attrition rates for each pair of small and large classes is summarized in Table 9. The category of students receiving a course grade of $W$ or $F$ is designated as fail and those students receiving a course grade of $A$, $B$, $C$, or $D$ is designated as pass.

The attrition rates as displayed in Table 9 may be misleading since initial student capability as measured by the covariate, $CR$, is not considered in the evaluation of these rates. This possible bias is the reason Chi-square
Table 9

Attrition Rates for Each Pair of Small and Large Classes

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Pass</th>
<th>Fail</th>
<th>Attrition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Instructor 0</td>
</tr>
<tr>
<td>Small</td>
<td>25</td>
<td>12</td>
<td>13</td>
<td>52.0</td>
</tr>
<tr>
<td>Large</td>
<td>84</td>
<td>53</td>
<td>31</td>
<td>36.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Instructor 1</td>
</tr>
<tr>
<td>Small</td>
<td>28</td>
<td>18</td>
<td>10</td>
<td>35.7</td>
</tr>
<tr>
<td>Large</td>
<td>87</td>
<td>48</td>
<td>39</td>
<td>44.8</td>
</tr>
</tbody>
</table>

Note. Pass represents the group of students receiving a grade of A, B, C or D. Fail represents those students receiving a grade of W or F.

was not used to test hypothesis two.

Logistic regression is a technique for modeling the probability, \( p \), that a dichotomous dependent variable having values 0 and 1 will be 1 as a function of one or more independent variables (Netter & Wasserman, 1974, pp. 320-338). The independent variables used were class size and the covariate, CR.

Using the logistic regression computer procedure (SAS, 1986), a logistic regression analysis of the attrition rates of each pair of large and small classes was performed. The regression equations for each pair of large and small
classes was defined by \( \log(p/1-p) = B_0 + B_1(\text{size}) + B_2(\text{CR}) \) where \( p \) is the probability of fail and \( B_0, B_1, \) and \( B_2 \) are the parameters (Beta).

The equality of slopes between the size and CR variables for each pair of large and small classes was checked using the Proc Logist Printc statement of SAS (1986) and found to be satisfactory. The estimates necessary for comparing attrition rates and analyzing the CR effect on the rates for each pair of large and small classes is summarized in Table 10.

Table 10
Logistic Regression Analysis of Attrition Rates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>Std. Error</th>
<th>Chi-square</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.24</td>
<td>1.03</td>
<td>16.93</td>
<td>.00</td>
</tr>
<tr>
<td>Size</td>
<td>-.50</td>
<td>.53</td>
<td>.89*</td>
<td>.34</td>
</tr>
<tr>
<td>CR</td>
<td>-.38</td>
<td>.08</td>
<td>20.35</td>
<td>.00</td>
</tr>
</tbody>
</table>

Instructor 1

| Intercept | 2.36 | .94        | 6.35       | .01  |
| Size      | -.06 | .49        | .01*       | .91  |
| CR        | -.23 | .06        | 12.07      | .00  |

\* \( p < .05 \).
The Chi-square statistic tests the hypothesis that a Beta parameter is zero. The Chi-square values of 20.35 and 12.07 for the variable CR of instructors 0 and 1, respectively, are significant at the .05 level. Hence, the covariate, CR, is accepted as having a significant effect on both pairs of attrition rates. If the CR had not been significant, a normal use of the Chi-square statistic could have been used to test hypothesis two. Moreover, the Chi-square estimates of .89 and .01 for instructors 0 and 1, respectively, are not significant at the .05 level. Hence, class size was not found to have a significant effect on the attrition rates of each pair of large and small classes. By ignoring the effect of initial ability as indicated by the CR scores, Chi-square estimates for each pair of small and large classes were evaluated and found nonsignificant at the .05 level.

Hypothesis two was accepted at the .05 level. There was no significant difference in the percentage of students in small first-semester calculus classes earning a grade of W or F from that of students in large classes.

Although a statistical analysis of subgroups within the large and small classes was not possible because of their small sizes, a comparison of the percentages of students in two subgroups within each small and large class receiving a course grade of W or F is presented in Table 11. The two complementary subgroups are those students scoring below 10
Table 11

Attrition Rates for Subgroups Within Two Pairs of Large and Small Classes

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Class size</th>
<th>n</th>
<th>W/F (%)</th>
<th>n</th>
<th>W/F (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Instructor 0</td>
<td></td>
<td>Instructor 1</td>
</tr>
<tr>
<td>CR &lt; 10</td>
<td>Small</td>
<td>9</td>
<td>77.8</td>
<td>4</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>22</td>
<td>81.8</td>
<td>33</td>
<td>66.7</td>
</tr>
<tr>
<td>CR &gt; 9</td>
<td>Small</td>
<td>16</td>
<td>37.5</td>
<td>24</td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>62</td>
<td>21.0</td>
<td>54</td>
<td>31.5</td>
</tr>
</tbody>
</table>

Note. W/F represents those students receiving a course grade of W or F.

An immediate observation of Table 11 is that over 66 percent of those students scoring below 10 on the CR received W or F except for the small class of instructor 1. The corresponding observation is that at most 37.5 percent of those students with a CR score of 10 or above received a course grade of W or F. In summary, class size appears to have no effect on the attrition rates for these two subgroups of students.

Hypothesis Three

Hypothesis three stated: Student evaluations of the instructor and the course by first-semester calculus
students in small classes is not significantly different from that of students in large classes. This hypothesis was tested by an analysis of two semantic differentials consisting of 16 scales each (see Appendixes C and D). Three additional bipolar adjective pairs, bright-dark, fragrant-foul, and pungent-bland, were used as distractors.

Each of the instruments was given to the students in each of the four calculus sections during the last week of the semester. A summative score for each student was obtained for each instrument by assigning values of one to seven to the seven possible responses of each of the 16 scales. The strongest positive response was given the value of seven, and the next strongest a value of six. Continuing in this manner each of the responses is rated from seven to one. Hence, the maximum favorable summative score by a student on the instructor survey is 112 and the least favorable score is 16. To analyze these summative scores a two-tailed $t$-test for significance of the difference between two means for independent samples was utilized (Ferguson, 1981, pp. 177-179).

As a preliminary to the use of the $t$-test on the summative scores for the semantic differential on the instructor by the students in each pair of large and small sections, the homogeneity of group variances was tested (Huck, 1974, pp. 57-58). Experimental $F$-values and $t$-values for each pair of large and small classes are shown in Table
Table 12
Summary of Statistics for Tests of Homogeneity of Variance and Equality of Means for Instructor Survey Scores

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Variance</th>
<th>F</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Instructor 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>10</td>
<td>100.10</td>
<td>80.28</td>
<td>2.50</td>
<td>2.94</td>
</tr>
<tr>
<td>Large</td>
<td>44</td>
<td>86.34</td>
<td>199.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>20</td>
<td>97.90</td>
<td>199.09</td>
<td>3.41</td>
<td>1.62</td>
</tr>
<tr>
<td>Large</td>
<td>54</td>
<td>102.04</td>
<td>58.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

≠
p < .05.

The experimental $F$-value, 2.50, for the quotient of variances associated with the pair of large and small classes of instructor 0 was found to exceed the critical $F$ value of 2.10 (Ferguson, 1981, p. 527) at the .05 level. Hence, the variances of the two samples of unequal size are significantly different. But when samples of unequal size have significantly different variances, the results of the $t$-test will be biased. In particular, when the larger sample has the larger variance, the $t$-test will yield too few significant results (Ferguson, 1981, p. 182). Now, the
experimental $t$-value, 2.94, associated with the sections of instructor 0 exceeds the critical $t$-value of 2.01 at the .05 level (Ferguson, 1981, p. 521). But since the variance of the large sample is larger than the variance of the small sample, the $t$-test results will be biased against a significant difference conclusion. Therefore, since the $t$-value is significant at the .05 level, the $t$-test result of significance is accepted and the means of the large and small classes of instructor 0 are significantly different at the .05 level with the small class having the larger mean.

The same testing procedure was applied to the small and large classes of instructor 1. The statistics necessary for this test evaluation is given in Table 12. The experimental $F$-value, 3.41, associated with the variances of this pair of classes was found to exceed the critical $F$-value of 2.00 at the .05 level (Ferguson, 1981, p. 526). Hence, the variances are not homogeneous. But when the large sample has the smaller variance, as is true in this instance, the tendency is to yield too many significant results. But since the experimental $t$-value of 1.62 is smaller than the critical $t$-value of 2.00 at the .05 level, the $t$-value is not significant. Hence, the $t$-test results of nonsignificance is acceptable. Therefore, the small and large classes have means that are not significantly different at the .05 level.

Next, as a preliminary to the use of the $t$-test on the
summative scores for the semantic differential on the course by the students in each pair of large and small classes, the homogeneity of group variances was tested (Huck, Cormier, & Bounds, 1974, pp. 57-58). The experimental $F$-values and $t$-values for each pair of large and small classes are shown in Table 13.

Table 13

Summary of Statistics for Tests of Homogeneity of Variance and Equality of Means for Course Survey Scores

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Variance</th>
<th>$F$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>10</td>
<td>90.70</td>
<td>149.57</td>
<td>1.04</td>
<td>.84</td>
</tr>
<tr>
<td>Large</td>
<td>44</td>
<td>82.70</td>
<td>156.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>20</td>
<td>85.90</td>
<td>172.40</td>
<td>1.38</td>
<td>1.04</td>
</tr>
<tr>
<td>Large</td>
<td>54</td>
<td>89.96</td>
<td>237.47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The experimental $F$-value of 1.04 associated with the variances of the large and small classes of instructor 0 was not significant since it was smaller than the critical $F$-value, 2.81, at the .05 level (Ferguson, 1981, p. 527). Therefore, homogeneity of group variances was accepted at
the .05 level. The corresponding experimental $t$-value was 2.01 (Ferguson, 1981, p. 521) at the .05 level. Hence the assumption of equal means for the large and small classes of instructor 0 on the course survey was accepted at the .05 level.

Testing the pair of large and small classes of instructor 1 for homogeneity of variance resulted in an experimental $F$-value of 1.77 at the .05 level (Ferguson, 1981, p. 528). Hence, homogeneity of group variances was accepted at the .05 level. The corresponding experimental $t$-value of 1.04 was smaller than the critical $t$-value of 2.00 (Ferguson, 1981, p. 521) at the .05 level. Hence, the means of the summative scores on the course survey for the pair of classes of instructor 1 were not significantly different at the .05 level.

Summarizing, the means of the instructor and course surveys for the large and small classes of instructor 0 were larger for the small class. Although the means were not significantly different at the .05 level for the course survey, the means were found to be significantly different for the instructor survey at the .05 level. The course survey result, therefore, is consistent with hypothesis three whereas the instructor survey result contradicts this hypothesis.

For both the instructor and course surveys of students in classes of instructor 1 the means were larger for the
large class. In neither case was the difference significant at the .05 level. The results of the instructor and course surveys for instructor 1, therefore, are consistent with hypothesis three.

Hypothesis Four

The fourth hypothesis tested was: Students in small first-semester calculus classes will respond more positively to the classroom learning environment. The testing process was identical to that of hypothesis three except that a directional one-tailed $t$-test was performed at the .05 level. The statistics necessary for this testing process is displayed in Table 14.

Testing the first pair of scores for homogeneity of variance, the experimental $F$-value was 4.16. This evaluation exceeded the critical $F$-value, 2.10 (Ferguson, 1981, p. 527), indicating that homogeneity of variance could not be assumed. Since the larger sample, however, had the larger variance, the ordinary $t$-test was biased with the tendency of yielding too few significant results (Ferguson, 1981, pp. 182-183). But in this instance the experimental $t$-value, 3.30, exceeds the critical $t$-value of 1.67 (Ferguson, 1981, p. 521) at the .05 level. The $t$-value, therefore, is significant. Hence, the significant result is acceptable since the bias derived from lack of homogeneity of variance only increases the degree of significance required for a significant $t$-value. Hence, the mean of the
Table 14

Summary of Statistics for Tests of Homogeneity of Variance and Equality of Means for Classroom Learning Environment Scores

<table>
<thead>
<tr>
<th>Section</th>
<th>n</th>
<th>Mean</th>
<th>Variance</th>
<th>F</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>10</td>
<td>98.40</td>
<td>60.53</td>
<td>4.16</td>
<td>3.30</td>
</tr>
<tr>
<td>Large</td>
<td>44</td>
<td>81.30</td>
<td>251.86</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Instructor 1 |
| Small   | 19 | 87.26 | 223.80   | 1.07  | .21 |
| Large   | 54 | 88.09 | 209.09   |       |     |

* p < .05.

small class of instructor 0 is significantly larger than the mean for the large class at the .05 level.

When the classroom learning environment survey scores for the second pair of large and small classes was analyzed the experimental F-value of 1.07 was determined and was found to be below the critical F-value of 2.04 at the .05 level (Ferguson, 1981, 526). Therefore, homogeneity of variance was accepted at the .05 level. An experimental t-value of .21 also fell far below the corresponding one-tailed critical t-value of 1.66 at the .05 level (Ferguson,
1981, p. 521). Hence, the mean of the smaller class was not found to be significantly larger than the mean for the larger class. Rather, the mean of the larger class was slightly larger than the mean of the smaller class.

In summary, the result of the directional one-tailed t-test for the scores on the classroom learning environment survey for the pair of classes of instructor 0 was consistent with hypothesis four, whereas, the test result for the other set of scores contradicted this hypothesis.

Summary of Findings

The following list of statements is used to summarize the findings of this study:

1. Out of a sample population of 225 students, 69.8 percent were males, 71.9 percent were less than 23 years of age with an average age of 21.4 years, 76 percent were employed and averaged 27.5 hours per week, 33.7 percent commuted 16 or more miles one-way to class on a daily basis, 77 percent were engineering or science students, and 77 percent were freshmen or sophomores.

2. The covariate, CR, was judged to be linearly related to each of the six achievement measures.

3. For both pairs of large and small first-semester calculus classes there was no significant difference in the adjusted means for each of the four periodic tests.

4. There was no significant difference in the
adjusted means of the final examination scores for each pair of large and small first-semester calculus classes.

5. There was no significant difference in the adjusted means of the final grade averages for each pair of large and small classes of first-semester calculus. The reverse was true for the other pair of classes.

6. There was no significant difference in the attrition rates of the small and large first-semester calculus classes.

7. The covariate, CR, had a significant effect on the attrition rates of each pair of small and large first-semester calculus classes.

8. Class size appears to have no association with the attrition rate of those groups of students scoring below 10 on the CR or above nine on the CR.

9. The student evaluation of one instructor by students in the small class was significantly different from that of students in the large class of first-semester calculus for one of the instructors and nonsignificant for the other. The evaluations were more favorable by students in the small class when the result was significant.

10. The student evaluation of the course by students in small and large first-semester calculus classes was not significantly different.

11. In one comparison of classes student evaluation of the classroom learning environment by students in the
small first-semester calculus class was significantly greater than that of students in the corresponding large class, but the result was not significantly different for the other comparison of small and large classes.
CHAPTER BIBLIOGRAPHY


CHAPTER V

SUMMARY OF THE STUDY

This final chapter is utilized to summarize the study, discuss the findings and conclusions, and list recommendations for further study.

Summary

The problem of this study was the association between class size, achievement, and opinions of university mathematics students. The typical opinion of students and faculty is that learning is promoted better within the environment of a small class setting (McKeachie, 1980). And since large classes are characterized by less student participation (Lewis, 1982), more classroom distraction (McConnell & Sosin, 1984), and higher rates of absentism (Feld & Crofman, 1977), learning theorists suggest (McKeachie, 1980; Thomas & Fink, 1963) that students in large classes achieve at lower levels than those in smaller classes.

The purposes of this study were to:

1. determine the relationship between class size and academic achievement among university mathematics students in first-semester calculus classes, and

2. compare opinions of university students in small
first-semester calculus classes with those in large classes.

A survey of related literature revealed only three class size studies dealing with university mathematics courses. In one of these studies the achievement of students selected from large classes was significantly less than that of those selected from the small classes. In the other two studies no significant difference in achievement was found.

Historically, most class size experimentation has been done in the disciplines of education and psychology with some of the more recent studies involving business or economics. The final examination has been the most common achievement measure used. Large and small class sizes have not been standardized and add to the degree of difficulty in evaluating the results of class size studies. Most of the class size experiments resulted in no significant difference in achievement when an objective final examination was used as the measure of achievement. There has been, however, little research that uses tests for problem-solving or analytic skills as the measure of achievement. Finally, generalizations concerning the relationship between class size and course or instructor evaluations by students appear unwarranted.

This quasi-experimental study was composed of 225 university students enrolled in two large and two small sections of first-semester calculus at the University of
Texas at Arlington during the fall semester of 1987. Each of two tenured faculty, both with six or more years experience teaching large sections, taught a large section of approximately 85 students and a small section of less than 30 students. Demographic and mathematical-preparation data were collected on each of the 225 students in the sample and presented in summary form in Tables 1 and 2.

The CR was used as the covariate in an analysis of covariance that was applied to each of six achievement measures. Logistic regression was utilized to analyze the attrition rates of each pair of large and small first-semester calculus classes. Student opinion of the instructor, course, and classroom learning environment was measured using the semantic differential (Osgood, Suci, & Tannenbaum, 1957) and analyzed using the t-test on the summative scores of the semantic differential.

The following list of statements is used to summarize the findings of this study:

1. Out of a sample population of 225 students, 69.8 percent were males, 71.9 percent were less than 23 years of age with an average age of 21.4 years, 76 percent were employed and averaged 27.5 hours per week, 33.7 percent commuted 16 or more miles one-way to class on a daily basis, 77 percent were engineering or science students, and 77 percent were freshmen or sophomores.

2. The covariate, CR, was judged to be linearly
related to each of the six achievement measures.

3. For both pairs of large and small first-semester calculus classes there was no significant difference in the adjusted means for each of the four periodic tests.

4. There was no significant difference in the adjusted means of the final examination scores for each pair of large and small first-semester calculus classes.

5. There was no significant difference in the adjusted means of the final grade averages for each pair of large and small classes of first-semester calculus.

6. There was no significant difference in the attrition rates of the small and large first-semester calculus classes.

7. The covariate, CR, had a significant effect on the attrition rates of each pair of small and large first-semester calculus classes.

8. Class size appears to have no association with the attrition rate of those groups of students scoring below 10 on the CR or above nine on the CR.

9. The student evaluation of one instructor by students in the small class was significantly different from that of students in the large class of first-semester calculus for one of the instructors and nonsignificant for the other. The evaluations were more favorable by students in the small class when the result was significant.

10. The student evaluation of the course by students
in small and large first-semester calculus classes was not significantly different.

11. In one comparison of classes the student evaluation of the classroom learning environment by students in the small first-semester calculus class was significantly greater than that of students in the corresponding large class, but the result was not significantly different for the other comparison of small and large classes.

Discussion of the Findings

There were four hypotheses in the study. Hypotheses one and two related to the purpose of determining the relationship between class size and academic achievement among university mathematics students in first-semester calculus. Hypotheses three and four related to the purpose of comparing opinions of university students in small and large first-semester calculus classes. A discussion of these four hypotheses follows.

Hypotheses One and Two

There was no significant difference in the adjusted means for the traditional achievement measure of the objective final examination (Educational Testing Service, 1964) for both pairs of large and small sections of first-semester calculus. This result is consistent with the results of most class size experiments in other disciplines but contradicts the belief of Williams, Cook, Quinn, and Jensen (1985) that although class size has little, if any,
effect on lower-level educational outcomes, such as recall of facts, it does have an effect on the achievement of higher-level outcomes such as problem solving skills. This common belief expressed by Williams et. al is based on the theory that the extreme limitations of group participation and feedback together with classroom distraction result in a lower level of achievement. The final examination result, however, is consistent with the belief of McKeachie (1978) that students in large classes are able to compensate for the lack of in-class learning obtained in a small class when the objective final examination is the sole achievement criterion.

There was also no significant difference in the attrition rates for both pairs of small and large classes. Since the covariate, CR, was utilized in the logistic regression analysis of attrition rates, the effect of student ability of the individual large and small calculus classes was taken into account. It was theorized by Siegel and Siegel (1964) that student interaction and feedback are particularly important for students with low levels of motivation, lack of subject-matter sophistication, or the tendency to memorize facts. In particular, when problem-solving skills were being taught, Siegel and Siegel anticipated that students in small classes would out-perform their counterparts in large classes. Based on this line of reasoning it was the anticipation of this researcher that in
a large class only the "hearty" survive resulting in a higher attrition rate for the large class. Since the attrition rate analysis took the student ability into account the lack of a significant difference in attrition rate was unexpected.

McKeachie (1978) believes the objective final examination is not always a valid measure of achievement when it is the only measure used. This reasoning is based on the belief that students are able to compensate for in-class learning or that the objective final examination is an ineffective achievement measure. In anticipation of this criterion problem four periodic tests, given approximately at the end of each quarter of the semester, and the final grade averages were included as achievement measures. For both pairs of large and small classes there was no significant difference in the adjusted means for each of these five achievement measures. Of the 10 comparisons nine of the adjusted group means were in favor of the large classes.

Apparently, if students are adversely affected by large class size, then they are able to adjust and achieve at or above the level that they would in a small class. This adjustment, moreover, appears to occur early in the semester since the students in both large classes achieved at a higher level for each of the periodic tests than those of the corresponding students in the small classes. This
researcher anticipated that the students in the small class would achieve at a higher level than those in the corresponding large class until the higher attrition rate of the large class resulted in a higher achieving group of students remaining in the large class. This change in the remaining student population would result in the large class achieving at or above the level of the small class. This reasoning was believed to account for the large number of experiments resulting in no significant difference when the objective final examination was the sole measure of achievement used.

Hypotheses Three and Four

The results of the student evaluation of the instructor were mixed. There was a significant difference in the means of the instructor evaluation by students for one pair of large and small classes. In this case the students in the small class rated the instructor significantly higher than those in the large class. The other instructor was rated higher by students in the large class, but the difference was not significant. Perhaps it is of value to note that all of the mean ratings were high and the lowest rating was 34 percent above a neutral rating. These contradictory results typify the mixed research findings in the literature relating class size and student opinion. Feldman (1978) reported that approximately one-third of the studies he reviewed indicated no relationship between class size and
student ratings of the instructor. The remaining studies implied that the smaller the class, the higher the student rating of the instructor.

The means of the student evaluation of the course were not significantly different for each pair of large and small classes. The literature on student evaluations of the course is extremely varied. Only one of these results is consistent with a recent major survey of the literature by Feldman (1984). His review indicated a very weak inverse relationship between class size and student evaluation of the course.

The student evaluation of the classroom learning environment appears to support the results of the student evaluation of the instructor. The students in the small class that rated the instructor significantly greater than the corresponding large class also rated the classroom learning environment significantly greater than those in the corresponding large class. For the other pair of classes the students in the large class rated the instructor greater than those in the small class, although the difference was not significant. Similarly, this same large class of students rated the classroom environment greater than the corresponding small class of students. The consistency of the first pair of results was anticipated by this researcher, but the second pair of results was not. Almost all of the literature reviewed by this researcher indicated
a negative opinion toward the large class. In all four classes the mean classroom learning environment rating was well above the neutral level. The lowest mean rating was over 17 percent higher than the neutral rating.

Demographics

An analysis of the demographic data collected during the first week of the semester led to the following student profile: the "typical" student is male, 18 to 21 years of age, freshman or sophomore, science or engineering major, employed 27.5 hours per week, and lives within 15 miles of the university. Over 90 percent of the students had credit for a high school or college course in analytic geometry and the mean Calculus Readiness score for the sample of 225 students was 11.73.

Conclusions

The findings of the present study led to the following conclusions:

1. The main purpose of this study was to compare the academic achievement of university students in large first-semester calculus classes with those in small classes. A different framework of analysis for measuring the criterion of achievement was used. In addition to the traditional achievement measures of an objective final examination and final grade averages, four periodic tests given approximately at the end of each quarter of the semester
were used.

For each of the six achievement measures each pair of comparisons of large and small classes resulted in no significant difference. Since the attrition rate between each pair of large and small classes was not significantly different, the major conclusion of this study was that there is no difference in the achievement of university students in small first-semester calculus classes of size 25 to 30 and large classes of size 85 to 90.

2. The conclusion that there is no difference in achievement for small and large first-semester university calculus classes, where small is defined as 25 to 30 students and large is defined as 85 to 90 students, is consistent with the majority of results of research done in the subject areas of education, psychology, and business/economics. The testing of problem-solving skill did not account for a difference in achievement as was hypothesized by McKeachie (1978) and Williams et al. (1985). It can also be concluded that the lack of student interaction and feedback does not result in a significant effect on the level of achievement as was hypothesized by social psychologists, Thomas and Fink (1963), and researchers Siegel and Siegel (1964). Since the objective final examination was found to be consistent with the other five measures of achievement, there is no reason to believe that the objective final examination is an unreliable
instrument for measuring achievement.

There is little doubt the conclusion that the class sizes considered in this study make no difference in the achievement of university students in first-semester calculus is not a popular one. There are two factors, however, that may be related to this conclusion. The first is the factor that being in a large class may motivate a significant number of students to work harder and more independently than they might if they were in a small class. The second factor is that in a small class some of the students who have had calculus in high school are intimidating to some of those who have not had this experience. This, in turn, might serve to inhibit substantially the group participation that many believe is a significant factor in achievement.

3. A second purpose of the study was to compare the opinions of first-semester calculus students in large classes toward the instructor, course, and classroom learning environment with those in small classes. The semantic differential (Osgood, Suci, & Tannenbaum, 1957) was the instrument used for these comparisons.

The student evaluation of the instructor by students in the small class was significantly different from that of students in the large class for one of the instructors. The result of the student evaluation of the other instructor was not significant. Since the results were not consistent it
appears that instructors are unable to assume that students in a small class will rate them more favorably than those in a large class. Even though it is true generally that students prefer small classes to large, it was concluded that this factor is not significant enough that instructors should anticipate lower instructor ratings from large classes than from small classes. This conclusion is consistent with the mixed conclusion by Feldman (1978): one-third of the studies reviewed indicated no relationship between class size and student ratings of the instructor, whereas, the remaining studies implied a negative correlation.

4. The result of the course evaluation was very similar to that of the instructor evaluation. The students of the large class gave a more favorable opinion of the course than those of the corresponding small class for one of the instructors and the reverse result occurred in the student evaluation of the course by those in the other pair of large and small classes. The conclusion, therefore, was the same: It can not be anticipated that the student evaluation of a course by those in a small class will be more favorable than by those in a large class. The conclusion by Feldman (1984) of a very weak inverse relationship between class size and student evaluation of the course was not substantiated.

5. The result of the classroom-learning-environment
student evaluation was identical with that of the student evaluation of the instructor. The students in the small class that evaluated their instructor significantly more favorably than those in the large class also evaluated their classroom learning environment significantly more favorably than did those in the large class. And in the other comparison the large class evaluated the classroom learning environment more favorably than those in the corresponding small class just as they evaluated the instructor more favorably than the small class. This leads to the conclusion that the student evaluation of the classroom learning environment is independent of class size but there may be an association between the student evaluation of the instructor and the student evaluation of the classroom learning environment.

6. The CR was used as the covariate in each analysis of covariance of the six achievement measures and in the logistic regression analysis of the attrition rate. In every case it related significantly to the criterion being measured and, because of its strong relationship to attrition rate, it was concluded that it could be used as a placement examination for calculus students at the participating university.

Since the scores of the students taking the CMTC were consistent with their final grade averages, this test was judged to be an effective comprehensive final examination
for first-semester calculus. The three semantic
differentials (Osgood, Suci, & Tannenbaum, 1957) also were
judged to be effective instruments for evaluating student
opinion.

Recommendations

Based on the findings and conclusions of this study,
the following recommendations are made:

1. A replication of this study in the same content
area should be made at a variety of institutions of higher
education for comparative purposes.

2. A similar study should be made using a large class
size exceeding 150 students to investigate the association
of achievement and the upper limits of large class size.

3. Replications of this study should be made in other
entry level mathematics courses, such as business
mathematics and college algebra, since the mathematical
profile of the students population for these courses is
different from those taking first-semester calculus.

4. An investigation should be made to determine the
extent of effect on students taking first-semester calculus
by those students in the classroom who have credit for a
high school course in calculus.

5. This researcher recommends the instruction of
first-semester calculus of class size not exceeding 100
students when taught by willing and experienced faculty.
Recitation sections are not recommended as a necessary
instructional resource.

6. A study should be made to investigate the association between student evaluation of the instructor and student evaluation of the classroom learning environment.

7. A study should be made to investigate the association between study habits of university students of small and large mathematics classes.
CHAPTER BIBLIOGRAPHY


APPENDIX A

ASSIGNMENT SHEET FOR FIRST-SEMESTER CALCULUS
ASSIGNMENT SHEET FOR FIRST SEMESTER CALCULUS

Assignment (Calculus with Analytic Geo., Swokowski, 3rd. ed)

1. p. 32 1, 3, 37, 41, 43, 45, 53, 55, 58, 60, 67, 69, 70, 73
2. p. 45 odds 1-11; p. 60: 1, 3, 13, 15, 17, 21, 25, 27, 33
3. p. 64 1, 3, 21-23, 25, 27, 29
4. p. 72 1, 3, 5, 7, 9; p. 161: 1-12
5. p. 79 1, 3, 4, 5a, b, 6a, b
6. p. 85 4-7, 9, 11, 13, 3
7. p. 93 1-6, 13-17, 22, 25, 27, 35, 42, 43, 47a, b, c, d
8. p. 93 9, 11, 19, 29, 31, 33, 39, 41, 47e, 53, 54
9. p. 101 1-3, 5, 7, 11, 13, 15
10. p. 108 odds 1-17, 21, 23, 27, 29, 45, 47, 49
11. p. 113 9-15, 17, 19, 21, 23
12. p. 115 odds 1-17, odds 21-31, 35, 37
13. p. 118 1, 3, 4, 9, 11, 15, 21; p. 131: odds 9-23
14. p. 131 1, 3, 4; p. 141: 3, 5, 9, 11, 15
15. p. 149 1, 3, 7, 9, 13, 19, 21
16. p. 169 1, 3, 5, 6, 7, 9
17. p. 170 11, 13, 17, 19, 24, 31
18. p. 177 1, 3, 5, 17, 23, 25, 29, 30
19. p. 182 1, 3, 7, 9, 13, 15, 17
20. p. 183 2, 4, 8, 10, 19, 21
21. p. 189 odds 1-15, 21, 23, 25, 41, 43
22. p. 227 summation handouts
23. area-by-limits handouts
24. integration handouts
25. p. 229 odds 1-15, 21, 25
26. p. 237 1-21
27. p. 245 1, 5, 9; p. 247: 11, 15, 17, 19, 23
28. p. 256 1, 5, 7, 9, 13, 15, 16
29. p. 256 3, 11, 17, 19, 8, 10, 12
**STUDENT INFORMATION CARD**

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>SEX</th>
<th>MAJOR</th>
<th>CLASSIFICATION</th>
<th>EMPLOYED</th>
<th>COMMUTE</th>
<th>HOURS/WEEEK AT WORK</th>
<th>TELEPHONE</th>
<th>DISTANCE</th>
<th>HIGH SCHOOL MATHEMATICS COMPLETED (circle)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fr. Soph. Jr. Sr. Grad.</td>
<td>Yes No</td>
<td>Yes No</td>
<td>(one way)</td>
<td>( )</td>
<td>(one way)</td>
<td>Algebra 1 year 2 years Trigonometry 1/2 years Analytic Geometry 1/2 years Analysis 1/2 years 1 year Other (in years)</td>
</tr>
</tbody>
</table>

**COLLEGE MATHEMATICS COMPLETED** (list courses by title)
APPENDIX C

INSTRUCTOR SURVEY
Directions: The purpose of this survey is to measure your opinion toward your instructor. Complete the rating sincerely, fairly rapidly, and spontaneously. If you feel that your opinion toward the instructor is very closely related to either end of a particular scale, then mark the appropriate end. If you feel that your opinion toward the instructor is neutral, then mark the middle space. The other spaces represent various degrees of feeling.

<table>
<thead>
<tr>
<th>INSTRUCTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>good</td>
</tr>
<tr>
<td>sensitive</td>
</tr>
<tr>
<td>valuable</td>
</tr>
<tr>
<td>pleasant</td>
</tr>
<tr>
<td>honest</td>
</tr>
<tr>
<td>dark</td>
</tr>
<tr>
<td>relaxed</td>
</tr>
<tr>
<td>knowledgeable</td>
</tr>
<tr>
<td>nice</td>
</tr>
<tr>
<td>foul</td>
</tr>
<tr>
<td>fair</td>
</tr>
<tr>
<td>reputable</td>
</tr>
<tr>
<td>effective</td>
</tr>
<tr>
<td>organized</td>
</tr>
<tr>
<td>bland</td>
</tr>
<tr>
<td>positive</td>
</tr>
<tr>
<td>commendable</td>
</tr>
<tr>
<td>enthusiastic</td>
</tr>
<tr>
<td>calm</td>
</tr>
</tbody>
</table>
APPENDIX D

COURSE SURVEY
Directions: The purpose of this survey is to measure your opinion toward this math course. Complete the rating sincerely, fairly rapidly, and spontaneously. If you feel that your opinion toward this course is very closely related to either end of a particular scale, then mark the appropriate end. If you feel that your opinion is neutral, then mark the middle space. The other spaces represent various degrees of feeling.

<table>
<thead>
<tr>
<th>COURSE</th>
<th>115</th>
</tr>
</thead>
<tbody>
<tr>
<td>good</td>
<td>---</td>
</tr>
<tr>
<td>full</td>
<td>---</td>
</tr>
<tr>
<td>valuable</td>
<td>---</td>
</tr>
<tr>
<td>pleasant</td>
<td>---</td>
</tr>
<tr>
<td>foul</td>
<td>---</td>
</tr>
<tr>
<td>strong</td>
<td>---</td>
</tr>
<tr>
<td>relaxed</td>
<td>---</td>
</tr>
<tr>
<td>acceptable</td>
<td>---</td>
</tr>
<tr>
<td>nice</td>
<td>---</td>
</tr>
<tr>
<td>fair</td>
<td>---</td>
</tr>
<tr>
<td>bland</td>
<td>---</td>
</tr>
<tr>
<td>reputable</td>
<td>---</td>
</tr>
<tr>
<td>useful</td>
<td>---</td>
</tr>
<tr>
<td>successful</td>
<td>---</td>
</tr>
<tr>
<td>dark</td>
<td>---</td>
</tr>
<tr>
<td>positive</td>
<td>---</td>
</tr>
<tr>
<td>fresh</td>
<td>---</td>
</tr>
<tr>
<td>pleasing</td>
<td>---</td>
</tr>
<tr>
<td>rich</td>
<td>---</td>
</tr>
</tbody>
</table>
Directions: Classroom learning environment is the set of classroom conditions, physical or psychological, that enhance or inhibit learning. The purpose of this survey is to measure your opinion toward this classroom as a learning environment. Complete the rating sincerely, fairly rapidly, and spontaneously. If you feel that your opinion toward this classroom learning environment is very closely related to either end of a particular scale, then mark the appropriate end. If you feel that your opinion is neutral, then mark the middle space. The other spaces represent various degrees of feeling.

CLASSROOM LEARNING ENVIRONMENT

calm                 ---- ---- ---- ---- ---- ---- disruptive
appropriate          ---- ---- ---- ---- ---- ---- inappropriate
pleasant            ---- ---- ---- ---- ---- ---- unpleasant
bland                ---- ---- ---- ---- ---- ---- pungent
tasteful             ---- ---- ---- ---- ---- ---- distasteful
happy                ---- ---- ---- ---- ---- ---- sad
relaxed              ---- ---- ---- ---- ---- ---- tense
foul                 ---- ---- ---- ---- ---- ---- fragrant
nice                 ---- ---- ---- ---- ---- ---- awful
fair                 ---- ---- ---- ---- ---- ---- unfair
enjoyable            ---- ---- ---- ---- ---- ---- distasteful
reputable            ---- ---- ---- ---- ---- ---- disreputable
dark                 ---- ---- ---- ---- ---- ---- bright
commendable          ---- ---- ---- ---- ---- ---- unacceptable
comfortable          ---- ---- ---- ---- ---- ---- uncomfortable
positive             ---- ---- ---- ---- ---- ---- negative
encouraging          ---- ---- ---- ---- ---- ---- discouraging
good                 ---- ---- ---- ---- ---- ---- bad
stimulating          ---- ---- ---- ---- ---- ---- distracting
APPENDIX F

LETTER FROM JOHN G. HARVEY
February 10, 1987

Prof. Eddie Warren, Chair  
Department of Mathematics  
University of Texas at Arlington  
Arlington, TX 76019

Dear Eddie:

I enjoyed talking with you on the phone and am happy to supply information about the MAA Placement Testing Program (PTP) Calculus-Readiness (CR) test.

The PTP CR test is generally regarded as the most successful of the PTP tests. It is extensively used; our last survey of PTP subscribers showed that 13 institutions used Part I and 42 institutions used both Part I and Part II. These institutions tested a total of 17,943 students in the year the survey was completed. CR is used by a variety of institutions; for example, it is used at the University of Arkansas-Fayetteville, at Colgate, and at St. Olaf (in Minnesota).

I am sorry that I cannot give you item reliabilities, difficulty levels, or discriminant analyses for CR. We had those data from tryouts of the completed CR, but those data were lost. I can tell you that CR, according to those data, is psychometrically valid; that is, a majority of the items are of medium difficulty, there are a few difficult and a few easy items, the items do discriminate between students, and the mean score (on the groups tested) was about 50%. CR is also content valid; that is, it does test knowledge and skills that students need before they can enter calculus.

I have looked through the past issues of the PT Newsletter and have selected a few papers that give some indication of the effectiveness of CR in placement. One of these papers (by Ernest Manfred) shows how CR is used with other data for placement while the other two papers report its use alone.

Give me a call (at (608) 262-3746) if you need further information.

Sincerely yours,

John G. Harvey  
MAA Committee on Placement Examinations
BIBLIOGRAPHY


class size on selected variables in a collegiate principles of management class: An exploratory study.


University of Texas at Arlington (1986). Undergraduate catalog. (Available from Office of Admissions, University of Texas at Arlington, Arlington, 76019)


