A Study of the Effects of Various Teaching-Testing Frequencies on Cognitive Gains.

Stan Andrew Wilkins

Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation


https://digitalcommons.lsu.edu/gradschool_disstheses/3363
INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in “sectioning” the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University Microfilms International
300 N. ZEEB ROAD, ANN ARBOR, MI 48106
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND
WILKINS, STAN ANDREW
A STUDY OF THE EFFECTS OF VARIOUS
TEACHING-TESTING FREQUENCIES ON COGNITIVE
GAINS.

THE LOUISIANA STATE UNIVERSITY AND
AGRICULTURAL AND MECHANICAL COL., PH.D., 1979
A STUDY OF THE EFFECTS OF VARIOUS TEACHING-TESTING FREQUENCIES ON COGNITIVE GAINS

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

The Department of Education

by

Stan Andrew Wilkins
B.S., Louisiana Tech University, 1970
M.Ed., Louisiana State University, 1972
May, 1979
ACKNOWLEDGMENTS

The author gratefully acknowledges the members of his committee, Dr. J. B. Gremillion, Dr. Sam Adams, Dr. Helen Cookston, Dr. James Firnberg and Dr. Robert Coon for their cooperative guidance. Sincere appreciation is expressed to Dr. R. C. Von Brock, the author's major professor, for his advice, direction and personal sacrifice during the study.

Acknowledgment is made to Dr. Douglas Peterson, Dean of Bossier Parish Community College, for allowing the study to be conducted at the school. Thanks are extended to other staff members who helped collect the data and whose cooperation was essential to the success of the study.

Special gratitude is expressed to the introductory psychology students at Bossier Parish Community College in the Spring, Summer and Fall Semesters for serving as subjects in the study and without whose participation the research could not have been possible.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>STATEMENT OF THE PROBLEM</td>
<td>2</td>
</tr>
<tr>
<td>THEORETICAL FRAMEWORK</td>
<td>3</td>
</tr>
<tr>
<td>PROCEDURE</td>
<td>6</td>
</tr>
<tr>
<td>DELIMITATIONS</td>
<td>7</td>
</tr>
<tr>
<td>DEFINITION OF TERMS</td>
<td>8</td>
</tr>
<tr>
<td>SOURCE OF DATA</td>
<td>9</td>
</tr>
<tr>
<td>ORGANIZATION OF THE STUDY</td>
<td>10</td>
</tr>
<tr>
<td>2. REVIEW OF RELATED LITERATURE</td>
<td>11</td>
</tr>
<tr>
<td>USING TESTS TO TEACH</td>
<td>12</td>
</tr>
<tr>
<td>RESEARCH SUPPORTING FREQUENT TESTING</td>
<td>13</td>
</tr>
<tr>
<td>RESEARCH NOT SUPPORTING FREQUENT TESTING</td>
<td>17</td>
</tr>
<tr>
<td>ABILITY LEVELS</td>
<td>17</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>20</td>
</tr>
<tr>
<td>3. CONDUCT OF THE STUDY</td>
<td>22</td>
</tr>
</tbody>
</table>

iii
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESEARCH DESIGN</td>
<td>22</td>
</tr>
<tr>
<td>RESEARCH SETTING</td>
<td>22</td>
</tr>
<tr>
<td>RESEARCH SAMPLE</td>
<td>23</td>
</tr>
<tr>
<td>STUDENT DATA</td>
<td>23</td>
</tr>
<tr>
<td>ADMINISTRATION OF THE INSTRUMENT</td>
<td>23</td>
</tr>
<tr>
<td>ASSIGNMENT OF SUBJECTS</td>
<td>24</td>
</tr>
<tr>
<td>TEACHING-TESTING FREQUENCIES</td>
<td>25</td>
</tr>
<tr>
<td>ADDITIONAL COMPARISONS</td>
<td>26</td>
</tr>
<tr>
<td>TREATMENT OF THE DATA</td>
<td>27</td>
</tr>
<tr>
<td>4. PRESENTATION AND ANALYSIS OF DATA</td>
<td>29</td>
</tr>
<tr>
<td>TESTING FREQUENCY</td>
<td>29</td>
</tr>
<tr>
<td>ANALYSIS OF COVARIANCE</td>
<td>30</td>
</tr>
<tr>
<td>THE DUNCAN'S TEST</td>
<td>31</td>
</tr>
<tr>
<td>INTERACTION EFFECTS</td>
<td>33</td>
</tr>
<tr>
<td>MALE-FEMALE DIFFERENCES</td>
<td>35</td>
</tr>
<tr>
<td>DAY-NIGHT CLASS ENROLLMENT</td>
<td>35</td>
</tr>
<tr>
<td>AGE</td>
<td>36</td>
</tr>
<tr>
<td>COLLEGE EXPERIENCE</td>
<td>36</td>
</tr>
<tr>
<td>5. SUMMARY AND CONCLUSIONS</td>
<td>38</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>38</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>39</td>
</tr>
<tr>
<td>RECOMMENDATIONS</td>
<td>40</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>42</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>APPENDICES.</td>
<td>45</td>
</tr>
<tr>
<td>A. STUDENT INFORMATION QUESTIONNAIRE.</td>
<td>46</td>
</tr>
<tr>
<td>B. SUBJECTS' RESPONSES TO THE STUDENT INFORMATION QUESTIONNAIRE.</td>
<td>47</td>
</tr>
<tr>
<td>VITA.</td>
<td>48</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>ADJUSTED MEAN POST-TEST SCORES FOR THE CONTROL AND EXPERIMENTAL GROUPS.</td>
<td>30</td>
</tr>
<tr>
<td>2.</td>
<td>SUMMARY OF ANALYSIS OF COVARIANCE: POST-TEST ACHIEVEMENT SCORES, EXPERIMENTAL AND CONTROL GROUPS.</td>
<td>32</td>
</tr>
<tr>
<td>3.</td>
<td>ADJUSTED POST-TEST MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS BY GENDER.</td>
<td>34</td>
</tr>
<tr>
<td>4.</td>
<td>ADJUSTED POST-TEST MEANS OF ALL SUBJECTS BY GENDER.</td>
<td>35</td>
</tr>
<tr>
<td>5.</td>
<td>ADJUSTED POST-TEST MEANS OF ALL SUBJECTS ACCORDING TO ENROLLMENT IN DAY AND NIGHT CLASSES.</td>
<td>36</td>
</tr>
<tr>
<td>6.</td>
<td>ADJUSTED POST-TEST MEANS OF ALL SUBJECTS GROUPED ACCORDING TO AGE.</td>
<td>37</td>
</tr>
<tr>
<td>7.</td>
<td>ADJUSTED POST-TEST MEANS OF ALL SUBJECTS GROUPED ACCORDING TO COLLEGE HOURS EARNED.</td>
<td>37</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTERACTION OF SEX AND TESTING FREQUENCY</td>
<td>34</td>
</tr>
</tbody>
</table>
This study was conducted to determine whether or not certain combinations of classroom testing frequencies facilitated the achievement of students enrolled in introductory psychology classes. The semester testing frequencies selected for study were, three tests, weekly tests, or no tests. The major objective of the study was to determine which of these teacher-made examination combinations resulted in higher achievement test results at the end of the semester, as measured by a standardized test.

Additional student characteristics were included in the study to ascertain what effect they would have on achievement scores. These concomitants included, student gender, age, previous college experience, and day-night class enrollment.

The following main hypothesis was tested. There is no significant difference in achievement test scores between subjects taught using the various testing frequencies.

Procedure

The study used 161 students, for whom all pertinent data were available, in ten classes of introductory psychology at Bossier Parish Community College, Bossier City, Louisiana. All classes were taught by the researcher using the lecture-discussion method, and the duration of the experiment was three semesters.

During the experimental study, three of the classes received
weekly tests, three of the classes received three tests, and four of
the classes received no tests at all except for the achievement pre-
and post-tests which were given to all groups. The classes receiving
three tests served as the control group while the classes receiving
weekly and no tests served as the experimental groups. Questions
given to the groups receiving examinations were from the lecture and
the text. The weekly quizzes consisted of fifteen multiple-choice
items, while the examinations administered three times a semester con­
tained from fifty-five to seventy multiple-choice items each.

During the first and last week of each semester, subjects were
administered the Cass-Sanders Psychology Test as pre- and post-test
measures of their knowledge of psychology. During the semester, the
experimental treatments were applied with the post-test achievement
score serving as the dependent variable in this study. In order to
control for previous knowledge of psychology and academic aptitude,
the pre-test score and the American College Test composite score of
each subject were included as covariables.

An F test was used to determine the significance of the differ­
ence between the means of the various groups. In addition, the
Duncan's Multiple Range Test was used to determine which specific
groups were significantly different.

Findings

An analysis of covariance revealed that the main effect of
testing frequency was statistically significant. The Duncan's Multi-
ple Range Test showed that subjects in the groups tested, whether three
times per semester or weekly, had significantly higher test scores than those who received no examinations \((p < 0.05)\).

Females scored higher than males in all groups, and analysis of the interaction of testing frequency with gender revealed that females in the weekly tested group and the group tested three times per semester scored significantly higher than all groups of males as well as the group of females receiving no examinations \((p < 0.05)\). The American College Test and the pre-test scores were significant factors in predicting student achievement in the study \((p = 0.0001)\).

No significant differences in achievement test scores were revealed because of enrollment in day or night classes, because of student age, or because of previous college experience between the experimental groups and the control group.

Student attitudes toward the testing frequencies indicated that those students in the classes receiving no examinations were disappointed that their final semester grades would depend on a single examination.

These results suggest that the use of quizzes is an effective technique for promoting increased learning in a course of this type.
CHAPTER 1

INTRODUCTION

Each semester teachers utilize time testing pupils in their classrooms. From the primary grades through college, testing is considered to be an important, though time consuming, activity that absorbs the effort and attention of the student and teacher.

A fundamental question is raised: "What are the motives of the teacher in giving a test?" (Gremillion and Adams, 1977). In offering an answer to this question Phillips (1968) suggests that tests can be used to determine how much students have learned, they can be used to assign grades, and they are helpful in determining needed adjustments in instruction. In addition, tests can provide feedback, improve programs and curricula, be used in placement and selection decisions, in diagnosis, and in remediation (Sax, 1974).

While research studies give evidence supporting the use of testing in accomplishing the above, there is perhaps another reason for testing which has been little emphasized, but may be just as important. It is that "... evaluation, properly used, can aid learning..." (TenBrink, 1974).

The use of testing procedures as teaching devices has been investigated by several researchers:

Phillips (1968) has reported that teacher-made tests are the most frequently used devices for gathering data about pupils.
Stanley and Hopkins (1972) have stated that students consistently report greater study and learning with periodic testing, and pursue mastery of objectives more diligently if they expect to be evaluated. In taking a test a student is engaging in behavior that is instructionally valuable, apart from the evaluation function being served by the test. Scheduled examinations stimulate review, relearning and overlearning.

In a study involving test questions, Von Schilling (1972), in quoting Grossier, said that "to question well is to teach well," and that "from the time of Socrates, history's greatest teachers have generally been penetrating questioners."

Gremillion and Adams (1977) have stated that studying for and preparing to take a test can be a good learning experience, since "the test gives the student individual review and a means to fill in the gaps in his own learning." According to one author, all the reasons for testing feed into this one: "to enhance student learning and growth" (Tuckman, 1975).

Statement of the Problem

The purpose of this study was to determine if a significant difference existed in the achievement test scores of students taught in various classes utilizing different numbers of teacher-made tests during a semester.

An answer to the following specific question was sought:

Is there a significant difference in achievement test scores of students taught using three teacher-made tests, no teacher-made tests,
and weekly teacher-made tests during a semester?

The effects of the different teaching and testing frequencies on achievement test scores with respect to other student characteristics were investigated. Sex, age, day-night class enrollment, and college hours previously earned were the student characteristics that were evaluated in terms of their interaction with the testing frequencies.

The following hypotheses were tested:

A. There is no significant difference in achievement test scores between subjects in the experimental groups and the control group.

B. There is no significant difference in achievement test scores between subjects grouped according to gender in the experimental groups and the control group.

C. There is no significant difference in achievement test scores between subjects grouped according to age in the experimental groups and the control group.

D. There is no significant difference in achievement test scores between subjects grouped according to the number of college hours they have earned in the experimental groups and the control group.

E. There is no significant difference in achievement test scores between subjects grouped according to their enrollment in day and night classes in the experimental groups and the control group.

THEORETICAL FRAMEWORK

Significance of the Study

Several studies have been conducted with respect to the effect of
testing procedures on the test scores of students in various grade levels. Researchers have dealt with various facets of testing procedures including frequency of testing (Shapiro, 1975; Stokes, 1973; Marso, 1968), ability levels (Rievman, 1974), feedback (Townsend, 1972), and pre-class tests (Curo, 1963).

These studies have used different experimental subjects with various experimental designs and have revealed mixed conclusions.

Curo (1963) found that there was no significant difference in achievement between two classes of high school American History, one given daily pre-class tests and one not given daily pre-class tests, while research by Marso (1968) revealed that students who were given six brief unit examinations in an introductory psychology class achieved greater on post-test measures than did those given three, one-hour unit examinations.

Research by Stokes (1973) revealed that without concern to ability levels, neither frequent nor infrequent testing results in greater achievement with college chemistry students. However, low ability students benefit more from frequent examinations while the number of examinations makes no difference with high ability students. Another investigation (Shapiro, 1975) concluded that high ability students do benefit from more examinations per semester.

A study of the effects of previous academic history on the frequency of testing with introductory educational psychology students (Rievman, 1974) concluded that the main effect of testing frequency was not statistically significant, but the interaction of frequency and ability was.
In research by Townsend (1972), comparing daily, weekly, three per semester, and one per semester testing frequencies, it was found that those receiving daily quizzes achieved significantly higher than those receiving only one examination per semester in college analytic geometry and calculus. All other comparisons were not significant.

Another study pointed out "... that when tests are given every day in a unit on sixth grade fractions and division, students achieve significantly higher," providing evidence that "... sixth grade students can learn from tests themselves" (Proger, 1968).

A survey of research concerning how testing procedures affect achievement test scores reveals a void with respect to comparing classes using frequent tests and those using no tests at all.

Can testing time be better used for teaching? Can the time used by teachers to make, grade and return test papers be better utilized in lesson preparation? Sparse evidence seems to exist regarding whether no tests at all might be a good, or even best, method of conducting a class.

Additional emphasis needs to be placed on the idea that tests themselves can be motivational and teaching devices. This is evidenced by the fact that less research exists in this area than in some of the areas involving the more conventional uses of testing. This study was conceived as a means of providing additional information on how classroom testing frequencies affect achievement test scores.

Specifically, the present study sought to examine the relationship between the number of tests taken during a semester and student achievement. It strove to determine if tests can serve as teaching
devices by comparing student achievement in classes where a variety of teaching-testing frequencies were used including classes in which no tests were used. Emphasis was placed on different in-class teaching and testing frequencies with concomitant comparisons made concerning several student and class characteristics.

PROCEDURE

The following steps were used in conducting this study:

1. Classes of Psychology 201, Introduction to the Study of Psychology, at Bossier Parish Community College, Bossier City, Louisiana, served as subjects in the study.

2. The subjects were taught during the Spring, Summer, and Fall Semesters of 1978. Students registered in the various class sections of their own volition. Ten classes were used in the study, five taught as day classes and five taught as night classes, and no attempt was made to place a particular student into any section.

3. During the three-semester study, four classes were taught without examinations (N = 74), three classes were taught using weekly examinations (N = 51) and three classes were taught in which three examinations were administered during a semester (N = 36).

4. At the beginning and at the end of each semester all classes were administered the Cass-Sanders Psychology Test as a pre-measure and post-measure of their knowledge of psychology.

5. Post-test scores, adjusted with the use of covariance analysis, were used as achievement scores for the subjects and was the dependent variable in the study. The pre-test achievement scores and
the American College Test scores were included as covariables to control for pre-existing differences in student aptitude and knowledge of psychology.

6. The Cass-Sanders Psychology Test, administered as pre- and post-tests, had no bearing on student semester grades. After the research was concluded all students were given study guides to help prepare them for the "teacher-made" final examination. Each class was evaluated independently in assigning semester grades so that any differences in achievement due to different testing frequencies did not work to the disadvantage of any student in assigning final semester grades.

7. Data cards containing all necessary information were key-punched, the necessary F-ratios were obtained and all null hypotheses were tested at the .05 level of significance.

DELIMITATIONS

The sample for this study was limited to those students who enrolled in the researcher's classes of Psychology 201 at Bossier Parish Community College during the Spring, Summer, and Fall Semesters of 1978. It consisted of ten classes.

Only those students who enrolled in the course, took the pre-test remained in the course and took the post-test, and for whom American College Test results were available, were included in the study. The number of subjects contained in the sample was greatly reduced because of these limitations.
DEFINITION OF TERMS

**Achievement Test Scores.** For the purposes of this study, this term refers to the student scores on the **Cass-Sanders Psychology Test.** These raw scores were used as a quantification of the achievement and knowledge of the student in psychology.

**Cass-Sanders Psychology Test.** A paper and pencil examination for use in psychology classes on the senior high school and college levels. It is an objective, multiple-choice examination covering the basic facts, rules, and principles of a first course in psychology, and it is used for determining pupil achievement.

The content of the test is based on the common content of leading contemporary textbooks and courses in introductory psychology. It is a timed test of fifty minutes consisting of one-hundred and twenty-five items.

Norms for the test are available based on 4180 students used in standardizing the examination, and its reliability was determined by the split-half method (Cass and Sanders, 1964).

**Psychology 201.** A course in introductory psychology taught at Bossier Parish Community College. It is a required course in most curricula, and it is listed as a sophomore level course, though it is open to any student regardless of his college level.

**Bossier Parish Community College.** Located in Bossier Parish, Louisiana, this school is one of two junior colleges in Louisiana funded through the Louisiana State Minimum Foundation Program. It is under the jurisdiction of the Bossier Parish School Board and the Board of Elementary and Secondary Education of the State of Louisiana.
**Day Class.** A class that meets during the hours 8:30 A.M. to 2:30 P.M. The classes meet in one of the following three combinations. Mondays, Wednesdays, and Fridays; Tuesdays and Thursdays; or Mondays and Wednesdays. The classes meet for a total of three hours a week.

**Night Class.** A class that meets during the hours 6:30 P.M. to 9:30 P.M., once a week, on any of the nights Monday through Thursday. The classes meet for a total of three hours a week.

**Teacher-Made Test.** A locally constructed test, of either objective or subjective type, for local classroom use (Phillips, 1968).

**American College Test.** A college admissions examination comprised of four tests of educational development and scholastic ability. The composite score of the academic tests was used. This score is a composite of the areas of English usage, mathematical usage, social studies reading, and natural sciences reading. The most crucial characteristic of this test is its predictive validity (Buros, 1972).

**Different Teaching and Testing Combinations.** Procedure for teaching different sections of the same course during a semester by having some classes engaged in testing and others in classroom study while all cover the same amount of subject-matter content.

**SOURCE OF DATA**

Data for this study was derived from student responses on the Cass-Sanders Psychology Test. In addition, American College Test scores and several other student characteristics were obtained for study.
ORGANIZATION OF THE STUDY

Chapter 1 presents an Introduction to the study and includes the Problem, Delimitations, Significance of the Study, Definition of Terms, and the Procedure followed in conducting the study.

Chapter 2 contains a Review of the Related Literature.

Chapter 3 presents the Method of Procedure used in conducting the study.

Chapter 4 presents an analysis of the Data obtained for the study.

Chapter 5 presents a summary of the study, includes implications and conclusions concerning the data and recommendations for further study. A bibliography of resources and an appendix is included.
CHAPTER 2

REVIEW OF RELATED LITERATURE

From the elementary classroom through graduate school great emphasis is placed on paper and pencil examination. Anxieties develop in students because of their anticipation in taking examinations and because so much depends on the outcome of the examination scores. Many different sources pressure students to do well on tests.

Why do tests exist and why are they such a large part of our educational atmosphere? The answers to these and related questions may be overlooked or tacitly assumed by the instructor in the classroom. Teachers may not often ask themselves what part tests serve in the teaching-learning process. The testing-evaluating program in a class is more likely seen as separate from instruction rather than part of it.

Frequently, there is an unstated "understanding" between the instructor and the student regarding the purpose of tests. An unspoken agreement seems to exist that in-class examinations are for the purpose of assigning grades or that they measure "achievement" of subject-matter. Students have learned to assume, without question, that they will be tested, and they do not ask, nor does the instructor explain, the purpose of the classroom test.

If a general objective of education is to enhance learning, then it is important that all phases of education have a purpose in
relation to this goal. It is also imperative for the teacher to know how a testing program can help achieve this goal and fit into the teaching process. Only when the purposes of examinations are clear in the mind of the instructor can they be communicated to the student. Examinations can fit logically and cogently into the instructional process while helping to achieve the goals and aims of education.

USING TESTS TO TEACH

Although testing serves a variety of purposes, one that receives little attention is that tests themselves can be teaching devices and learning aids (Sax, 1974; Phillips, 1968).

The examination should be regarded as a valuable instrument for motivating learning. It should fit into the teaching-learning atmosphere and further the goals of education. Sometimes this is not realized by the classroom teacher in designing a testing program. However, if students know that their achievement is going to be measured they will try to accomplish more (White, 1932).

Concerning the use of examinations in the teaching-learning atmosphere, Jersild (1929) took the definite position that the value of classroom examination as an aid to learning is two-fold. One effect of examination is to compel the student to rehearse and recapitulate what he has already learned. Another advantage arises from the influence which examination has on the learner's attitude. When put to a practical test, the learner is called upon to work under pressure, to participate more actively in the learning situation.

Among the many decisions facing the instructor concerning the teaching value of testing procedures, one which may be overlooked, is the number of examinations to be given during a course or semester.
This is a decision all instructors must make from the grade school years through graduate school.

A specific question arises: does the extent to which tests serve as teaching-learning devices depend upon the frequency of testing? A number of investigators have raised this question in the past, but fewer inquiries have been made into this purpose of testing than into some of the other, more popular and well known justifications for testing.

Research in this area has yielded multifarious experimental designs in different settings with findings somewhat inconsistent. Additional research is warranted to investigate the teaching ability of tests and to establish experimentally whether or not it makes a difference if learning can be maximized by utilizing a certain number of tests during a semester. Varying test frequency may have no effect, a slight effect, or a significant effect on learning (Monk and Stallings, 1971).

On the following pages a number of different studies are cited relating to the topic. No two are exactly in their design, and the various studies cited used different age levels of experimental subjects in various kinds of subject-matter classes. All of the studies reviewed included in some manner the topic of testing frequency, though many sought answers to slightly different questions with the main effect of testing frequency not always primary in the experiment.

RESEARCH SUPPORTING FREQUENT TESTING

The same tests administered to students in classes of educational psychology in the form of weekly tests, as opposed to monthly tests,
revealed a mean performance which was higher by twelve per cent in favor of the weekly tests, and which had high statistical significance (Keys, 1934). Further research showed that the retention of the weekly tested group as revealed by "pop-quizzes" administered from five to thirteen weeks after the corresponding periodic tests.

Four sections of introductory psychology were used in a study by Ross and Henry (1939). Each researcher taught two classes, an experimental group and a control group, and they found that students achieve better results when weekly tests are used than when they are tested twice per semester.

While research in the general psychology class was being conducted, one of the instructors also conducted the same experiment in two sections of educational psychology and found contrary results: the students receiving the fewer number of tests scored higher than those receiving weekly tests. The authors viewed this finding with skepticism however, because of the small number of students in the study (40) and the fact that the reliability of the final examination was not known.

Stanlee and Popham (1960) concluded, in a study involving college students in four sections of educational psychology, that the use of weekly quizzes, as opposed to no quizzes, tended to increase students' subject-matter achievement in a lecture discussion course, but the significance of the increase is lost by the end of the course. It was suggested that the reason for this result was that the pre-test examination provided sufficient extrinsic motivation, knowledge of results, structure, and enforced activity, with subject-matter to overbalance,
to a degree, the quiz effects.

A study on testing frequency evolved from the interest of some researchers in teaching machines. The question raised in this context was how beneficial frequent testing might be in the classroom for improving grades in high school chemistry classes. Research showed that the use of a weekly testing program, compared to no tests, did improve significantly, examination scores of high school Chemistry II students during a six-weeks experimental period (Pikunas and Mazzota, 1965). It was suggested that this finding boosted support for the teaching machine, a primary feature of which is frequent evaluation.

The researchers recommended additional investigations to determine whether this finding also applied to subjects other than science and whether it applied on each level of education.

Daily, weekly, and three per semester levels of testing were investigated by Townsend (1972) in college classes of analytic geometry and calculus. An analysis of the data revealed that those classes receiving daily quizzes achieved significantly higher scores on a criterion measure than those classes receiving only one midterm examination. Other differences between test frequency levels were not significant. This suggests that a possible optimum frequency for tests is daily.

A recent study was designed, in part, to determine whether several brief unit examinations better facilitate student achievement than do longer unit examinations (Marso, 1968). The researcher used students enrolled in an introductory educational psychology course and
the criterion measures were two achievement tests which were constructed from items appearing on a departmental examination.

The results were that "the students who were given six brief unit examinations achieved greater on the post-test measures than did those students given three one-hour unit examinations."

Research reported by Proger (1968) revealed a significant difference between groups tested daily, every other day, once per week, and not at all. Using sixth grade arithmetic students, he revealed that students tested daily recorded superior achievement on post-tests. Additional comparisons revealed that "on the immediate post-test, girls achieved significantly higher than boys at the .025 level, while on the delayed post-test, the girls achieved significantly higher than boys at the .005 level."

In a study designed to determine what effect the use of daily quizzes would have on the learning of college algebra students in a junior college, Nystrom (1969) found that the group receiving daily quizzes recorded a significant increase in learning over the group not tested. Using achievement gain scores as criterion measures he concluded that "the use of daily quizzes in an algebra class at the junior college level did result in greater mathematical achievement than in a similar course without the quizzes."

Dropout rates were significantly lower for the frequently tested group and an attitude questionnaire given the last week of the semester indicated no difference toward the course between the group given the daily quizzes and those not given daily quizzes.

The author recommended "that the use of frequent quizzes is an
effective technique for promoting increased learning in a course of this type."

RESEARCH NOT SUPPORTING FREQUENT TESTING

Although research has tended to suggest that students profit more from frequent testing than non-frequent testing, there is evidence that this is not always the case.

In a simple experiment involving high school American History students, some taught using daily, pre-class tests and some taught not using daily, pre-class tests, there was no significant difference in achievement, at the .05 level, between the two types of classes for a six-week period (Curo, 1963).

In a similar study, twelve "pop-quizzes" were given to college classes in American national government. There were no significant differences in achievement on desired outcomes between these classes and those tested only three times during a semester (Selakovich, 1962). The researcher concluded that the advisability of frequent testing of the sort he employed be decided by further research.

ABILITY LEVELS

A number of studies involving testing frequency have also looked at other closely related variables. Many of them have considered what effect frequent or infrequent testing had on achievement scores of students of varying abilities.

Generally, high previous achievers excel low previous achievers (Proger, 1968). Low ability students, admitted to an urban community college under an open admissions policy, benefited more from five tests
per semester (Shapiro, 1975).

A related study revealed that teacher ratings increased as the test frequency increased and that the findings were produced by the independent variable, frequency of testing, not by differences in subject-matter, teacher, or methodology (Shapiro and Stern, 1972).

Stokes (1973) researched two testing frequencies in college chemistry classes: one involving seven tests during a semester and the other comprised of four tests taken during the semester. Among several findings was "without concern to ability levels, neither testing method was more likely than the other to result in significantly greater achievement" on the criterion measures. However, frequent testing of low ability students was more likely than conventional testing of low ability students to result in significantly greater achievement.

Additional study has concurred in this finding. Rievman (1974) studied the effects of testing frequency upon the achievement of college students of three levels of ability enrolled in introductory psychology. Among subjects with low ability and medium ability, those tested weekly had the highest performance on the criterion measures of acquisition and retention while on a relearning measure, weekly and two and one-half week testing groups were equal. Those tested every five weeks had the lowest scores. It was concluded that infrequent testing most handicaps those students with poor academic backgrounds.

Rievman also reported that weekly testing was related to higher performance for high ability students as well, but cautioned that further study with larger numbers of students with higher academic
ability was needed.

Using low, middle, and high ability groupings with minimal, medial, and maximal testing frequencies in classes of sophomore level calculus, Mach (1963) revealed a significantly higher average achievement for the "middle ability" group than for the "low" or "high" ability groups among students tested frequently as opposed to those tested non-frequently.

He further revealed, using a previous mathematics grade-point index and a pre-test score as covariables, that the middle ability students who were the recipients of the "maximal" evaluation program attained significantly higher average achievement than did the middle ability students who participated in the "medial" evaluation program. In addition, "frequent testing did not produce significantly higher average homework scores than the nonfrequent testing" within any ability level group.

The "minimal" evaluation program consisted of a pre-test, a midterm examination and a final examination. The "medial" evaluation program consisted of all the testing frequencies in the minimal program plus two additional hour examinations and three half-hour quizzes. The "maximal" evaluation program consisted of all the testing frequencies in the "minimal" and "medial" programs plus a ten-minute quiz given during every class meeting when half-hour quizzes or examinations were not scheduled.

In other findings of the Ross and Henry study (1939), cited previously, the achievement gain in the experimental groups, who received more testing, was much more marked for the weaker student, and frequent
testing was less effective with higher ranking students.

Research involving retesting until mastery of subject-matter found that grade-point average was significantly correlated with final retention and test performance in introductory educational psychology classes. Students with low grade-point averages profited most from retesting (Komaridis, 1971). This finding was a replication of an earlier study by Summer and Brooker (1944).

Contingency grading, where test scores count or do not count toward a student's grade, seems to be a more significant variable on academic achievement than test frequency (Robinson, 1972).

SUMMARY

After a review of studies related to testing frequency, the following summary may be made:

1. Most studies were conducted using college age experimental subjects and a majority of those were enrolled in educational and introductory psychology classes;

2. The number of experimental subjects in the individual studies ranged from a low of thirty-eight in one study to a high of 442 in another study, and the number of experimental and control groups ranged from a low of one to a high of sixteen;

3. The maximum length of any study was one semester;

4. Frequent testing generally consisted of weekly or daily examinations;

5. Infrequent testing generally consisted of monthly, two per semester, or no examinations;
6. Variables concomitant with testing frequency were included in most studies. These consisted of knowledge of results, delay of feedback, anxiety levels, teacher popularity, contingency grading, retention scores, sex differences, previous achievement, and ability levels;

7. Research generally supported the argument for frequent testing as opposed to infrequent testing;

8. With respect to ability levels, students of low ability seem to profit more from frequent testing than those of high ability;

9. No study reviewed concluded unequivocally that students receiving fewer tests achieved significantly greater than those tested frequently;

10. Neither frequent nor infrequent testing was found to be superior for a particular subject or with a certain age student.
CHAPTER 3

CONDUCT OF THE STUDY

This study was conducted for the purpose of investigating the relationship between student achievement test scores and testing frequency. The relationship was further examined to determine the interaction effects of testing frequency with a student's age, sex, previous college experience, or day-night class enrollment. Covariates in the study included American College Test composite scores and a pre-test measurement of knowledge of the subject-matter.

This chapter is subdivided to include details of the research design, the research setting, the research sample, the assignment of subjects, the teaching-testing frequencies, and the data analysis.

RESEARCH DESIGN

The research design employed in this study was experimental and followed the paradigm of a pre- and post-test procedure.

The experimental groups consisted of four classes receiving no examinations during a semester and three classes receiving weekly examinations during a semester. The control group consisted of three classes receiving three examinations during a semester.

RESEARCH SETTING

Research for the present study was conducted at Bossier Parish Community College. The school conducts both Day and Evening Divisions.
with approximately three-fifths of the enrollment of 2600 attending the Evening Division.

The college serves a student population with varying academic goals. Some students are not pursuing a degree while others are seeking to earn credits satisfying curricular requirements at four-year degree-granting institutions. Others are working on associate degrees and some are taking courses in the non-academic division of the college. Credits earned applicable to a degree are transferred to other schools in the area and to other parts of the state.

RESEARCH SAMPLE

The research sample consisted of those students who registered in the researcher's classes of Psychology 201, Introduction to the Study of Psychology, during the Spring, Summer, and Fall Semesters of 1978, and who remained in the course for the entire semester. The sample was limited to those students for whom all pertinent data were available.

STUDENT DATA

At the beginning of each semester students were asked to fill out a brief questionnaire listing personal characteristics to be used in the study. A copy of the questionnaire is presented in the appendix.

ADMINISTRATION OF THE INSTRUMENT

Students enrolled in each of the ten sections of introductory psychology used in the study were administered the Cass-Sanders Psychology Test as a pre-test measurement of their knowledge of psychology.
at the beginning of each semester. They were instructed to do their
best on the test and not to guess haphazardly. They were told to
answer those questions they could, but not to spend unnecessary time
on questions concerning subject-matter unfamiliar to them.

Approximately one week before the end of each semester, the
students were again administered the Cass-Sanders Psychology Test as
a post-test measurement of their knowledge of psychology. The instruc­
tions given for taking the test were the same as for the pre-test.

Answers were recorded by placing the appropriate number on the
answer sheets provided. Both pre- and post-tests were administered
and scored by the researcher, and no student was told that he was
participating in an experiment.

The students were instructed that the pre-test was an attempt
by the researcher to obtain some information on where the class stood
in its knowledge of psychology, and that the post-test was designed to
measure how much progress had been made during the semester.

Further instructions to the class members at the beginning and
at the end of each semester explained that the pre-test and post-test
would have no bearing on their semester grades.

ASSIGNMENT OF SUBJECTS

Students enrolled in Psychology 201 during the college registra­
tion period at the beginning of each semester, and were free to select
any section offered. No attempt was made to counsel any student into
a particular section. This procedure made possible a random assignment
of subjects in each section and in each treatment modality.
There were ten sections of Psychology 201 used in the study and the treatments were randomly assigned according to the following schedule:

1. One class each semester the study was conducted received weekly, teacher-made quizzes; one met three times each week during the day, one met twice each week during the day, and one met twice each week at night. These three classes made up one of the experimental groups. The quizzes were administered at the last class meeting of the week in each section.

2. During the Spring Semester, two classes received no examinations. One was taught during the day meeting twice each week and one was taught at night meeting once each week. During the Fall Semester, two classes received no examinations. One was taught during the day meeting three times each week and one was taught at night meeting once each week. These four classes served as the second experimental group.

3. One class each semester the study was conducted received three tests. One of the classes met four times each week during the day and two met once each week at night. These three classes served as the control group.

TEACHING TESTING FREQUENCIES

All of the experimental subjects were taught by the researcher. Each class utilized the same amount of in-class time studying the subject-matter, with the following exceptions: during the times when testing was taking place in the classes receiving three tests per
semester and the classes receiving weekly tests, additional teaching was taking place in the non-tested classes. All classes were given the same reading assignments from the text and lecture topics were identical in all classes. Additional class discussion and more in-depth coverage of the subject-matter consumed the additional teaching time available in the non-tested classes.

The teacher-made tests consisted of multiple-choice items. Some were written by the researcher while others were used that were provided by the company publishing the text. The weekly examinations consisted of fifteen multiple-choice items. The classes receiving three examinations each semester were administered tests having from fifty-five to seventy multiple-choice items each.

After scoring by the researcher all teacher-made examinations were briefly returned to the students. An opportunity was given for asking questions about the examinations.

ADDITIONAL COMPARISONS

The following student characteristics were investigated with respect to their effects on testing frequency:

1. Gender - subjects' achievement test scores in the experimental and control groups were divided into male and female categories.

2. Day-Night Class Enrollment - subjects' achievement test scores in the experimental and control groups were divided into day and night categories based on their enrollment in a day or a night class.

3. Age - subjects' achievement test scores in the experimental
and control groups were divided into three categories, based on their chronological age at the end of the semester in which they participated in the study. The age categories were, 17-21, 22-29, and above 29. This age breakdown provided approximately the same number of students in all three groups.

4. **College Hours Previously Earned** - subjects' achievement test scores in the experimental groups and the control group were divided into three categories based on the number of semester hours they had earned: 0-15 hours, 16-30 hours, and more than thirty hours. These divisions were selected in order that a "beginning" student, a "second semester" student, and an "experienced" student might be studied. This categorization provided approximately the same number of students in all three groups.

**TREATMENT OF THE DATA**

Data for this study were derived from scores on the pre- and post-test of the Cass-Sanders Psychology Test, student American College Test composite scores, and information obtained from a student information questionnaire. After the tests were scored and all pertinent data used in the study had been collected, they were coded by the researcher and then key-punched by personnel in the Computer Center at Louisiana State University.

Analysis of covariance was used to test the difference between means of the three groups based on different testing frequencies. In order to equate the three groups statistically, the pre-test score and the American College Test composite score were included as covariables.
The analysis of data yielded adjusted means for the post-test scores, the dependent variable in the study. The resulting F-ratios were used to test the hypotheses of no significance difference between groups at the .05 level of confidence.

Comparisons were made using the several variables under study. All relevant data were analyzed and displayed in tabular form.
CHAPTER 4

PRESENTATION AND ANALYSIS OF DATA

The purpose of this study was to determine if a significant difference existed in the achievement test scores of students taught in various classes utilizing different numbers of teacher-made tests during a semester.

An answer to the following specific question was sought:

Is there a significant difference in achievement test scores of students taught using three tests, no tests, and weekly tests during a semester?

Additional student characteristics were included in the study to ascertain what effect they would have on achievement scores. These concomitants included student gender, age, previous college experience, and day-night class enrollment. In order to ascertain these characteristics the subjects were asked to complete a Student Information Questionnaire designed for the study. A copy of the questionnaire is included in Appendix A and a summary of all of the subjects' responses to the questionnaire is presented as Appendix B.

The above question and student characteristics were restated as five null hypotheses and tested with the analysis of covariance.

TESTING FREQUENCY

The null hypothesis stated that there is no significant difference
in achievement test scores between subjects in the experimental groups and the control group.

For the purpose of analyzing the scores, the group receiving three examinations per semester was labeled the control group (C), while the group receiving weekly examinations was labeled the first experimental group (E₁), and the group receiving no examinations was labeled the second experimental group (E₂). An analysis of the adjusted post-test scores of the subjects in the experimental and control groups revealed that those subjects receiving tests, whether weekly or three per semester, had a significantly higher adjusted mean score on the post-test than those receiving no examinations (p < .05). These data are presented in Table 1.

Table 1
ADJUSTED MEAN POST-TEST SCORES FOR THE CONTROL AND EXPERIMENTAL GROUPS

<table>
<thead>
<tr>
<th>GROUP</th>
<th>N</th>
<th>ADJUSTED MEAN SCORE</th>
<th>POSITIVE DIFFERENCES BETWEEN GROUPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>36</td>
<td>46.03</td>
<td>C - E₂ = 3.96*</td>
</tr>
<tr>
<td>E₁</td>
<td>51</td>
<td>45.11</td>
<td>E₁ - E₂ = 3.04*</td>
</tr>
<tr>
<td>E₂</td>
<td>74</td>
<td>42.07</td>
<td></td>
</tr>
</tbody>
</table>

* Significant at the .05 level.

ANALYSIS OF COVARIANCE

The data derived from the analysis of covariance for the dependent
variable, testing frequency, are presented in Table 2. An analysis of the scores for the total group revealed that the main effect of testing frequency was significant beyond the .05 level. Females had a significantly higher score than males beyond the .01 level while the only significant interaction was between testing frequency and gender. All other interactions of testing frequency with the variables under study failed to meet the test for significance at the .05 level of confidence.

Therefore the first two null hypotheses concerning the main effect of testing frequency and its interaction with gender were rejected while the remaining three null hypotheses failed to be rejected.

Further examination of the data revealed that the American College Test scores and the Cass-Sanders Psychology Test scores were significant predictors of achievement (p = .0001).

Although the analysis of covariance revealed a statistically significant difference between testing frequencies and a statistically significant difference between the interaction of testing frequency and gender, it was left to further analysis to show which of the groups differed significantly from the others.

**THE DUNCAN’S TEST**

For the purpose of determining which groups had significantly different scores, the Duncan's Multiple Range Test was chosen. It can be used to identify which groups are contributing to any overall differences if a significant F value is obtained from an analysis of covariance (Senter, 1969).
Table 2
SUMMARY OF ANALYSIS OF COVARIANCE POST-TEST ACHIEVEMENT SCORES, EXPERIMENTAL AND CONTROL GROUPS

<table>
<thead>
<tr>
<th>SOURCE OF VARIATION</th>
<th>DEGREES OF FREEDOM</th>
<th>MEAN SQUARES</th>
<th>F VALUE</th>
<th>PROBABILITY OF F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing Frequency</td>
<td>2</td>
<td>178.24</td>
<td>3.14</td>
<td>.0469</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>1127.61</td>
<td>19.83</td>
<td>.0001</td>
</tr>
<tr>
<td>Day-Night Classes</td>
<td>1</td>
<td>6.95</td>
<td>.12</td>
<td>.7271</td>
</tr>
<tr>
<td>Hours Earned</td>
<td>2</td>
<td>18.38</td>
<td>.32</td>
<td>.7244</td>
</tr>
<tr>
<td>Age</td>
<td>2</td>
<td>4.75</td>
<td>.08</td>
<td>.9199</td>
</tr>
<tr>
<td>Testing Frequency Gender Interaction</td>
<td>2</td>
<td>248.53</td>
<td>4.37</td>
<td>.0146</td>
</tr>
<tr>
<td>Testing Frequency Day-Night Class Interaction</td>
<td>2</td>
<td>85.62</td>
<td>1.51</td>
<td>.2258</td>
</tr>
<tr>
<td>Testing Frequency-Hours Earned Interaction</td>
<td>4</td>
<td>38.56</td>
<td>.68</td>
<td>.6083</td>
</tr>
<tr>
<td>Testing Frequency-Age Interaction</td>
<td>4</td>
<td>112.37</td>
<td>1.98</td>
<td>.1021</td>
</tr>
<tr>
<td>American College Test</td>
<td>1</td>
<td>3967.80</td>
<td>69.79</td>
<td>.0001</td>
</tr>
<tr>
<td>Pre-Test</td>
<td>1</td>
<td>2538.94</td>
<td>44.66</td>
<td>.0001</td>
</tr>
</tbody>
</table>
The use of the Duncan's Multiple Range Test in determining which group of subjects had the highest scores revealed that those subjects who were tested, weekly or three times per semester, had significantly higher scores than those who received no examinations ($p < .05$). There were no significant differences between those subjects tested three times per semester and those tested weekly. These data are also contained in Table 1. Since the analysis of covariance revealed no other significant interactions, further application of the Duncan's Test was unnecessary.

**INTERACTION EFFECTS**

A revealing finding of this study was the interaction of testing frequency with gender. After an overall significant F value was obtained, the Duncan's Test revealed that the interaction of testing frequency and gender was due to a significance difference between the females ($p < .05$). An analysis of the data showed that females scored higher than males in all groups, and females in the weekly tested group and in the group tested three times per semester scored higher than all groups of males, as well as the group of females not tested. These data are presented in Table 3.

Since no significant differences existed between the males at the .05 level of significance, it was concluded that the significant interaction of the main effect of testing frequency with gender was due only to the significant difference existing between the females. A graphical representation of this finding is presented in Figure 1.
Table 3
ADJUSTED POST-TEST MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS BY GENDER

<table>
<thead>
<tr>
<th>GROUP</th>
<th>ADJUSTED MEAN SCORES OF MALES</th>
<th>ADJUSTED MEAN SCORES OF FEMALES</th>
<th>POSITIVE DIFFERENCES BETWEEN SEXES</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>44.03 (N = 20)</td>
<td>48.03 (N = 16)</td>
<td>4.00*</td>
</tr>
<tr>
<td>E₁</td>
<td>39.70 (N = 17)</td>
<td>50.53 (N = 41)</td>
<td>11.83*</td>
</tr>
<tr>
<td>E₂</td>
<td>40.85 (N = 33)</td>
<td>43.27 (N = 34)</td>
<td>2.42</td>
</tr>
</tbody>
</table>

* Significant at the .05 level.

Figure 1. Interaction of Sex and Testing Frequency
Male-Female Differences

The null hypothesis stated that there is no significant difference in achievement test scores between males and females in the experimental groups and the control group. An examination of the scores, with students grouped according to gender, revealed that female subjects had higher mean scores than males. These data are presented in Table 4.

Table 4

<table>
<thead>
<tr>
<th>GENDER</th>
<th>ADJUSTED POST-TEST MEAN</th>
<th>POSITIVE DIFFERENCE BETWEEN SEXES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>41.07</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>47.30</td>
<td>6.23*</td>
</tr>
</tbody>
</table>

* Significant at the .05 level.

This difference was of sufficient magnitude to be statistically significant (p < .05).

Day-Night Class Enrollment

The null hypothesis stated that there is no significant difference in achievement test scores between day and night classes in the experimental groups and the control group. An examination of the scores revealed that there was no significant difference at the .05 level of confidence in achievement test scores between students in day
and night classes. Thus, the null hypothesis failed to be rejected. These data are presented in Table 5.

Table 5

ADJUSTED POST-TEST MEANS OF ALL SUBJECTS ACCORDING TO ENROLLMENT IN DAY AND NIGHT CLASSES

<table>
<thead>
<tr>
<th>TIME</th>
<th>ADJUSTED POST-TEST MEAN</th>
<th>POSITIVE DIFFERENCE BETWEEN DAY-NIGHT CLASSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>43.91</td>
<td></td>
</tr>
<tr>
<td>Night</td>
<td>44.46</td>
<td>.55</td>
</tr>
</tbody>
</table>

Age

The null hypothesis stated that there is no significant difference in achievement test scores between students age 17-21, 22-29, and 30 years of age and older, between the experimental groups and the control group. An examination of the scores revealed that there was no significant difference at the .05 level of confidence between the various age students in the experimental groups and the control group. Thus, the null hypothesis failed to be rejected. These data are presented in Table 6.

College Experience

The null hypothesis stated that there is no significant difference in achievement test scores between students with 0 to 15, 16 to 30, and more than 30 college hours between the experimental groups and the control group. An analysis of the data revealed no significant
Table 6
ADJUSTED POST-TEST MEANS OF ALL SUBJECTS
GROUPED ACCORDING TO AGE

<table>
<thead>
<tr>
<th>AGE</th>
<th>ADJUSTED POST-TEST MEAN</th>
<th>POSITIVE DIFFERENCE BETWEEN GROUPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 - 21</td>
<td>43.70</td>
<td>.80</td>
</tr>
<tr>
<td>22 - 29</td>
<td>44.50</td>
<td>.14</td>
</tr>
<tr>
<td>30 and above</td>
<td>44.36</td>
<td></td>
</tr>
</tbody>
</table>

difference at the .05 level of confidence in achievement test scores between these students in the experimental groups and the control group. Thus, the data failed to reject the null hypothesis. These data are presented in Table 7.

Table 7
ADJUSTED POST-TEST MEANS OF ALL SUBJECTS GROUPED ACCORDING TO COLLEGE HOURS EARNED

<table>
<thead>
<tr>
<th>HOURS EARNED</th>
<th>ADJUSTED POST-TEST MEAN</th>
<th>POSITIVE DIFFERENCE BETWEEN GROUPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 15</td>
<td>43.25</td>
<td>1.39</td>
</tr>
<tr>
<td>16 - 30</td>
<td>44.64</td>
<td>.02</td>
</tr>
<tr>
<td>31 and above</td>
<td>44.66</td>
<td></td>
</tr>
</tbody>
</table>
SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter contains a summary of the findings of the study, the conclusions arrived at from the research are presented, and recommendations for further study are included.

Summary of the Findings

One of the more cogent findings derived from this study was that subjects involved in the group tested weekly and in the group tested three times per semester scored significantly higher on their achievement tests than did those subjects in the non-tested group ($p < .05$). There were no significant differences between the group tested weekly and the group tested three times per semester ($p > .05$).

An analysis of the data showed that females scored higher than males in both experimental groups and in the control group. Further examination of the results revealed that females in the weekly tested group and the group tested three times per semester scored higher than all groups of males, as well as the group of females not tested. All other comparisons and interactions were insignificant.

In conducting this study student attitudes toward the course were obtained formally and informally. These attitudes were not essential to any finding of the study, but are discussed here to help summarize the research.

A brief questionnaire was presented to the subjects at the end of
each semester which solicited student opinion about the testing frequencies used in their classes. In general those students in the classes receiving no examinations did not feel that this was the best method to use in determining semester grades. They did not desire for their grade to be based on the final examination only.

The students in the classes tested weekly seemed to like this method much better at the end of the semester than the beginning. Informal verbal opinion sampling revealed that studying for the comprehensive final examination did not seem as difficult for this group.

Another anecdote to the experiment concerned student reaction to the standardized pre- and post-test instruments. They were somewhat discouraged by their low scores, even though it was pointed out that these scores had no effect on their grade, and even though they were encouraged not to be overly apprehensive about these scores.

**Conclusions**

A major conclusion derived from this study was that testing frequency does make a difference with respect to student achievement.

Based on the statistical treatment and analysis of the data, the following conclusions were drawn relative to the problem investigated in this study:

1. Students who took weekly examinations during a semester had significantly higher achievement test scores than students who took no examinations during a semester.

2. Students who took three examinations during a semester had significantly higher achievement test scores than those students who took no examinations during a semester.
3. Females had significantly higher achievement test scores than males in all groups, regardless of testing frequency.

4. Females in the group who took three examinations and those who took weekly examinations during a semester had significantly higher achievement test scores than all groups of males, regardless of testing frequency.

5. Females in the group who took examinations, regardless of number, had significantly higher achievement test scores than females who took no examinations.

6. The statistically significant interaction of testing frequency and gender was due to the fact that there was a statistically significant difference between the females in the experimental groups and the control group.

Recommendations

Previous research indicated mixed conclusions concerning whether frequent testing during a semester increased student achievement. The present study revealed that there were differences in means between the experimental groups and the control group with respect to achievement test scores. These differences were of significant disparity to be statistically significant. Accordingly, the following suggestions were recommended as points to be considered in future studies regarding the number of tests to use in maximizing student achievement:

1. Additional research should be conducted using a teacher-made evaluation instrument in a post-test only experimental design with American College Test scores included as a single covariable.
2. Additional testing frequencies might be considered for study including daily quizzes, monthly examinations, and combinations of these.

3. Other dependent variables might be added to a study of this type including study habits and student attitudes toward the course.

4. Dependent variables which might be omitted from an experiment of this type include gender, college experience, and day-night class enrollment.
BIBLIOGRAPHY


APPENDIX A

STUDENT INFORMATION QUESTIONNAIRE

1. Name ________________________________

2. Total number of college hours earned (at all institutions) prior to this semester ________.

3. Total number of hours taking this semester ________.

4. Check the following courses that you have taken:

   High School Psychology __________
   Introductory College Psychology __________
   Child Psychology __________
   Adolescent Psychology __________
   Other Psychologies (list) __________

5. Sex: Male ______ Female ______

6. Have you taken the ACT test? Yes ______ No ______

7. If no will you take it this semester? ______

8. Degree Seeking: Yes ______ No ______

9. Which of the following kinds of high schools did you graduate from? (check one).

   A Bossier Parish High School ______
   A Caddo Parish High School ______
   Other Louisiana High School ______
   Out of State High School ______
   G.E.D. ______

10. Do you work? Yes ______ No ______

11. If yes, Full Time ______ Part Time ______
APPENDIX B

SUBJECTS' RESPONSES TO THE STUDENT INFORMATION QUESTIONNAIRE

1. Name ____________________________________________

2. Total number of college hours earned (at all institutions) prior to this semester. $X = 15.83$

3. Total number of hours taking this semester. $X = 9.35$

4. Check the following courses that you have taken:

   High School Psychology 7.45% (12)
   Introductory College Psychology 2.48% (4)
   Child Psychology 2.48% (4)
   Adolescent Psychology (0)
   Other Psychologies (developmental) .62% (1)

5. Sex: Male 43.48% (70)  Female 56.52% (91)

6. Have you taken the ACT test? Yes 100% (161)  No (0)

7. If no will you take it this semester?

8. Degree Seeking: Yes 98.14% (158)  No 1.86% (3)

9. Which of the following kinds of high school did you graduate from? (check one).

   A Bossier Parish High School 29.19% (47)
   A Caddo Parish High School 18.01% (29)
   Other Louisiana High School 9.94% (16)
   Out of State High School 27.95% (45)
   G.E.D. 14.91% (24)

10. Do you work? Yes 67.08% (108)  No 32.92% (53)

11. If yes, Full Time 47.83% (77)  No 19.25% (31)
VITA

Stan Andrew Wilkins was born in Vicksburg, Mississippi. He attended Tallulah Elementary School and graduated from Tallulah High School in 1966. In 1970, he received his Bachelor of Science degree in Mathematics Education at Louisiana Tech University in Ruston, Louisiana.

His teaching career began in 1970 in Shreveport, Louisiana where he taught high school mathematics for three years. In 1972, he received his Master of Education degree from Louisiana State University, Baton Rouge, and began part-time work on his doctorate.

From 1973 to 1979 he pursued his doctoral studies, intermittently, at LSU, full and part-time, serving as a graduate assistant in the same institution in 1973-74 and again in 1977. While serving as a graduate assistant he had both research and teaching responsibilities in addition to serving as a supervisor of student teachers. In 1974, he became an instructor of psychology at Bossier Parish Community College, Bossier City, Louisiana, where he is currently employed.
Candidate:     Stan Andrew Wilkins

Major Field:   Education

Title of Thesis:  A Study Of The Effects Of Various Teaching-Testing Frequencies On Cognitive Gains

Approved:

\[\underline{\text{Major Professor and Chairman}}\]

\[\underline{\text{Dean of the Graduate School}}\]

EXAMINING COMMITTEE:

\[\underline{\text{H. C. Brown}}\]

\[\underline{\text{James B. Traylor}}\]

\[\underline{\text{Helen M. Cochran}}\]

\[\underline{\text{Dorothy Adams}}\]

\[\underline{\text{L. S. Linsberg}}\]

\[\underline{\text{Robert C. Cooper}}\]

Date of Examination:

April 19, 1979