5-1976

An Analysis of Oral and Written Quizzes As Teaching Techniques

K. Anthony Edwards
Utah State University

AN ANALYSIS OF ORAL AND WRITTEN QUIZZES AS TEACHING TECHNIQUES

by

K. Anthony Edwards

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

of

DOCTOR OF PHILOSOPHY

in

Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1976

ACKNOWLEDGMENTS

There are many to thank for the completion of this work, but I alone can be blamed for its faults. Those who have contributed most include Dr. Richard B. Powers, with whom I have worked quite closely over many projects in past years, and Elki Powers, who endured it all without, it is hoped, irreversible effects. Dr. Edward K. Crossman has been a special kind of positive influence for which he is greatly appreciated. Dr. Thomas Johnson is thanked for his encouragement. And, Dr. Michael Bertoch is greatly appreciated for his role as counselor. Dr. Frank Ascione is also thanked for his late arrival and support.

Others who deserve a great many thanks include Drs. Carl D. Cheney and J. Grayson Osborne. They will remain an influence on my work for many years to come. Dr. James Shaver has also been appreciated for a favorable influence on my attitude. Mr. John Williams of Student Services has been of great assistance and the help provided through his office has been deeply appreciated. Several others of my peers have been greatly supportive. They need not be listed; they know who they are.

The greatest appreciation and thanks of all are extended to my close companion, critic, and colleague, Ms. Penelope Gottula. Gratitude of the highest magnitude is extended for her encouragement and support.

K. Anthony Edwards

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ii

LIST OF TABLES vi

LIST OF FIGURES viii

ABSTRACT ix

Chapter

I. INTRODUCTION 1

The Problem 2
The Purpose 5

II. REVIEW OF THE LITERATURE 6

The Personalized System of Instruction 6

Keller's Method 7
Ferster's Method 7
Michael's Method 8
Summary and Discussion 8

Preparing for a Personalized Course 10

Written Materials 10
Proctors 11
Grading 11
Lectures 12
Work Load 12
Costs 13
Other Sources 13

Lectures and Attendance 14
Learning and Retention 16
Work Patterns 17
Bonus Points for Early Work 19
Withdrawals 21
Attitudes 22
Proctoring in PSI 24
Summary and Discussion 26

III. EXPERIMENT I 32

Methods 32

TABLE OF CONTENTS (Continued)

Page

Subjects 32
Apparatus and Setting 33
Procedures 33

Results 42

Learning 42
Attitudes and Interest 46
Summary of Results 52

Discussion 53

IV. EXPERIMENT II 57

Methods 57

Subjects 57
Materials 58
Course Procedures 60

Results 69

Reliability and Validity 69
Learning 71
Attitudes and Interest 79
Summary of Results 79

Discussion 82

V. GENERAL DISCUSSION 86

On Rating and Raters 89
On Self-Paced Instruction 91
Implications for Further Research 92

VI. SUMMARY AND CONCLUSIONS 95

Summary 96
Discussion and Conclusions 97

REFERENCES 99

APPENDIXES 106

Appendix A. Experiment I: Raw Test Scores 107

TABLE OF CONTENTS (Continued)

Page

Appendix B. Experiment I: Tabled Analysis of Covariance . 115

Appendix C. Experiment I: Student Evaluation Form . . . 121

Appendix D. Experiment I: Course Evaluation Form . . . 122

Appendix E. Experiment I: Tabled Survey Data 124

Appendix F. Experiment II: Means of Pre-Course Test, Post-Course Test, and Difference Scores 130

Appendix G. Experiment II: Sample Quiz 134

Appendix H. Experiment II: Reliability Data 135

Appendix I. Experiment II: Rating Quiz Instructions to Raters 139

Appendix J. Experiment II: Student Performance Evaluation Data 142

Appendix K. Experiment II: Course Evaluation Form . . . 144

VITA 146

LIST OF TABLES

Table		Page
1.	Temporal Conditions for Each Section in Each Course	36
2.	Means and Standard Deviations of Test Scores on Test 1, Test 2, and Test 3 for Each Class Section in Each Course	45
3.	Course Procedures and Conditions for Each Section of the Psychology of Adjustment Class	60
4.	Correlations of Quiz Rating With Difference Scores on Pre-Course and Post-Course Tests	70
5.	Pre-Course and Post-Course Test Means, Standard Deviations, and Variances for the Psychology of Adjustment Class	74
6.	Raw Scores on Tests During the Term	107
7.	Percent Test Scores, Quiz Conditions, and Percent Differences for Individual Students in Two Courses	111
8.	Analysis of Covariance: Introductory Psychology Course (Test 1)	115
9.	Analysis of Covariance: Introductory Psychology Course (Test 2)	116
10.	Analysis of Covariance: Introductory Psychology Course (Test 3)	117
11.	Analysis of Covariance: Psychology of Adjustment Course (Test 1)	118
12.	Analysis of Covariance: Psychology of Adjustment Course (Test 2)	119
13.	Analysis of Covariance: Psychology of Adjustment Course (Test 3)	120
14.	Majors, Minors, and Students Not in Psychology	124
15.	Sex of Students Enrolled	124
16.	Attrition From Each Class at the Term's End	125

LIST OF TABLES (Continued)

Table		Page
17.	Ratings of Oral and Written Procedures and Resultant Preferences Collected at Test 3	125
18.	Number of Students and Hours Spend Studying In Each Course by Section	126
19.	Number of Students Reporting a Particular Method of Reading Assigned Material	127
20.	Number of Students Within a Self-Estimated Grade-Point-Average Interval	127
21.	Number of Students Expecting a Particular Grade From the Course	128
22.	Final Grade Distributions for Introductory Psychology and Psychology of Adjustment Courses	129
23.	Final Grade Distributions for the Oral/Written and Written/Oral Sections	129
24.	Raw Scores Means for Pre-Course, Post-Course, and Difference Scores From Pre-Course to Post-Course Tests	131

LIST OF FIGURES

Figure		Page
1.	Flow-chart diagram of experimental course procedures . .	34
2.	Mean percent of test scores for tests following lectures (L), oral quizzes (O), and written quizzes (W) in the introductory psychology sections (101), and the psychology of adjustment sections (201)	43
3.	Distribution of student's reported comparisons of the introductory psychology class with other classes taken in the same term	47
4.	Distribution of student's reported comparisons of the psychology of adjustment class with other classes taken in the same term	50
5.	Flow-chart diagram of instructor-paced procedures in the psychology of adjustment course	61
6.	Mean correct answers for each unit on the pre-course and post-course tests in two sections of a psychology of adjustment course	72
7.	Mean difference scores for students in both sections: pre-course test scores were subtracted from post-course test scores	75
8.	Distributions of summed ranks based on difference scores between pre-course and post-course tests and the condition to which students were exposed: Oral quizzes (O), Written quizzes (W), or No quizzes (NQ)	77
9.	Distribution of student's reported comparisons of the psychology of adjustment class with other classes taken in the same term	80
10.	Distribution of student's reported comparisons of the summer psychology of adjustment class with other classes taken in the same term	87

ABSTRACT

An Analysis of Oral and Written Quizzes

As Teaching Techniques

by

K. Anthony Edwards, Doctor of Philosophy

Utah State University, 1976

Major Professor: Dr. Richard B. Powers
Department: Psychology

This study was conducted to determine whether any differences existed between the effectiveness of oral and written quizzes as teaching techniques. In the first of two experiments, 130 students enrolled in two sections of an introductory psychology class and two sections of a psychology of adjustment class served as subjects. The course was taught using Michael's method of instruction, a contingency-managed technique. In the first half of the term, one section from each course was taught by written quizzes while the other section was taught by oral quizzes. Following four weeks, teaching assignments were reversed. The dependent variables were scores on a test following one week of lectures, socres on tests following each condition, and attitude and interest ratings at the end of the course. For the two introductory psychology sections, there was no significant difference between oral and written methods. For the psychology of adjustment sections, a significant difference was found favoring the oral method in one of two comparisons. From surveys, it was shown that students were interested in and reported favorable attitudes toward

the class in both courses. However, a preference for the oral method was shown only in the introductory psychology course.

In a second experiment, 70 students enrolled in two sections of psychology of adjustment completed the requirements for the study. After each of 14 taped lectures, students were quizzed orally, wrote quiz answers, rated oral quizzes, rated written quizzes, or took no quiz. Inter-rater reliability checks were made by an item-by-item analysis of paired rater's scores of student's performances on quizzes. The raters' reports were judged reliable. Validity was examined by a correlational analysis of quiz ratings and unit test performances. Validity was poor. An analysis of rank sums for difference scores obtained from pre-course and post-course test performances showed no differences between any two of the variables studied. Respectively, ranked sums of scores for each variable from high to low was: oral quiz rating, oral quiz taking, written quiz taking, no quizzes, and written quiz rating. However, the differences were not significant. Students interests in, and attitudes toward, the course in psychology of adjustment were found to be on a par with other classes taken in the same term.

From Experiment I, it was concluded that test performances following oral quizzes were not significantly different from test performances following written quizzes. From Experiment II, rating and taking oral quizzes may result in higher test scores than rating and taking written quizzes, but in the present study the results leading to such a conclusion are only suggestive. Finally, judgments of untrained raters on

student quiz performances should be viewed with caution since, in the examination of rater's validity, the raters failed to match quiz performance with test performance.

(163 pages)

CHAPTER I

Introduction

Keller's Personalized System of Instruction (PSI) has been termed "The brightest new star in the constellations of college teaching methods" (McKeachie, 1974, p. 171). This is exemplified by an increase in the application of PSI to teaching psychology and other areas (Johnston, 1975; Keller & Sherman, 1974; Kulik, Kulik, & Carmichael, 1974; Ryan, 1974; Sherman, 1974). Increased interest in the research of PSI lends further support to McKeachie's view (Johnston, 1975; Sherman, 1974).

PSI is characterized by a unit mastery requirement, small unit steps, self-paced performance, an emphasis on the written word, and the use of proctors (Keller, 1963, 1968, 1971). Unit mastery requires that the student demonstrates articulation of the material before passing to future units. Units are brief enough for nearly all students to learn the unit material quickly. Students are allowed to pace their reading or studying, test-taking, and interviewing performances as their schedule permits. The instructor emphasizes the use of written instruction rather than lectures. Lectures are used as motivating devices first and as teaching devices second. More advanced students, or proctors, are used to deliver immediate feedback to students.

It can be seen that PSI is a new concept. Yet, its elements have been used in one form or another for centuries. Proctors, or "monitors," were used in the 16th century to assist in instruction (Hager, 1959).

The Bell-Lancaster system of monitored teaching became quite popular in the U.S. in the early 1800's, but enthusiasm was later directed more toward teacher-preparation (Hager, 1959; Meyer, 1967, 1972). Keller (1967) reported that self-pacing has been used in military centers and in some areas of vocational training. Emphasis on the written word rather than lecturing has been the case with correspondence studies. Brief units are often used in conventional classes in which weekly quizzes are the rule. Unit mastery is required for training medical students. Keller (1967) remarked that a method similar to his own had been used to teach elementary students (see Washburn & Marland, 1963).

The Problem

"Contingency management in an introductory psychology course produces better learning," according to McMichael and Corey (1969, p. 79). Questions have been raised concerning the value of any one teaching method over another prior to the advent of PSI (Dubin & Taveggia, 1968; Gage, 1968a, 1968b). McKeachie (1963) stated that it is clearly unjustified to conclude that there is no one best method; but, it is no simple task to detail the so-called best method in a few words. McKeachie (1974) has suggested more recently that PSI may be that "best" method.

Early studies of teaching methods have generally failed to include surveys of such contingency-managed courses as Keller's PSI. Therefore, whether personalized instruction is the "one best method" has not yet

been determined. Keller's PSI, however, has a long history of research beginning largely with Skinner's (1938) early statements about the behavior of organisms. Keller's theory based on the experimental analysis of behavior (Sherman, 1974) has yet to be subjected to a complete analysis. To attempt a thorough experimental analysis of Keller's system would be an overwhelming task for a single research program. But, few teaching methods prior to Keller's allowed a complete analysis. The major problem has been the lack of a teaching method which allows research of its component parts leading to its support. Strong evidence, as indicated in the studies reviewed below, now exists for McMichael and Corey's (1969) conclusion that "contingency management produces better learning" (p. 79).

Keller's (1968) classic paper stressed the use of proctors who examined the results of student's written performance. The proctors made the results available to the student, showed him where to find a correct answer, and allowed the student to retake the test later (if the performance was not satisfactory). Ferster (1968), in the same year, reported a procedure quite similar to Keller's. The student in Ferster's class was first required to demonstrate an oral mastery of the unit material and then required to take a written quiz over several units.

Whether an oral quiz and a written quiz differ as teaching techniques remains unknown. Although Farmer, Lachter, Blaustein, and Cole (1972) found that some proctoring was better than no proctoring, it was not examined whether the proctoring procedure might have been done as

well by oral quizzes without interaction. Whitehurst (1972) compared oral tutorials with written quizzes. No differences between oral procedures and written procedures were seen on weekly test scores. Johnston (1971) and Johnston and Pennypacker (1971) using rate of responding as a measure of performance showed equivalence between the two procedures with equivalent testing procedures. Edwards and Gottula (1973), however, showed an opposite effect on final exam scores. Comparisons of group means between the oral and written groups showed results approaching significance, but neither weekly exams nor difference scores obtained from pre-course and post-course test scores were used to measure performance.

Some experimental oversights raise some additional questions. For example, Johnston (1971) and Johnston and Pennypacker (1971) may have created a ceiling effect, where higher scores were not possible, with respect to performance measures over test retakes. Calhoun (1974) found no differences between oral and written methods, but the procedures used were not specified clearly enough to enable the location of any "cause." Calhoun (1974) used carefully selected proctors from past courses while Edwards and Gottula (1973) used proctors from within the class.

The present research is based upon the following question: Does teaching with the use of oral quizzes differ from teaching with the use of written quizzes? An answer to this question would aid instructors in the design of a teaching procedure. It is easier to monitor written quizzes, but the oral quizzes produce greater student interaction.

There may be further advantages to each technique. The ultimate answer may be that both techniques in combination are necessary for maximum learning; but, the rationale for the present research is the determination of whether the two techniques differ. If so, answers to the above question will provide additional courses of action in the technology of instruction.

The Purpose

The purpose of the present research was to determine whether differences on examination performance follow instruction by oral or written quizzes. An additional purpose was to determine the effects of proctoring (i.e., rating) oral and written quizzes on the test performances of the proctors (i.e., raters).

CHAPTER II

Review of the Literature

According to Keller (1967), the traditional system of instruction by lecturing has changed in ways important to PSI. Development of reinforcement theory, derived from experimental results, has provided the basis for the instructional techniques characteristic to PSI (see also Keller & Sherman, 1974; Sherman, 1974). Skinner's (1958) programmed system of instruction which antedated PSI has been successful in many respects; but, as Keller (1963) noted, something was missing which involved the personal touch. As Keller (1967) was quick to add, however, problems still remained to solve.

This chapter reviews the literature relevant to the problem under study: Does teaching using oral quizzes differ from teaching using written quizzes with respect to learning? Basic to the present research are the contingency-managed instructional techniques developed by Keller, Ferster, and Michael. These techniques will be described in greater detail below. The review will provide a basis for an understanding of the Keller method and an understanding of the rationale for the research methods used in the present study. For a recent review of the system as a whole, see Keller and Sherman (1974).

The Personalized System of Instruction

Contingency-managed instruction (CMI) has been described as a term most appropriate to studies that apply operant conditioning

procedures to teaching college courses (Cooper, 1973). There are some clear differences among at least three methods falling under the heading of CMI. These three methods have often been described as PSI or The Keller Method, but PSI is not an appropriate term for most variations of Keller's method. Each method of primary significance to the present research is described in detail below.

Keller's Method

Keller (1971) has summarized the five major characteristics of PSI. The defining properties of Keller's method are student-paced learning, small steps, unit perfection, emphasis on the written word, and the use of proctors. Students self-pace by arranging for testing as soon as each feels prepared to demonstrate mastery of the material. Mastery implies that the student demonstrates knowledge of the unit material with a grade of 90% or better on the unit test. The units are small enough to allow most students to demonstrate mastery on the first attempt, but often more than one attempt is necessary. Emphasis on the written word means that the instructor presents coursework through written material rather than through lecture. Proctors are selected from students who have demonstrated unit mastery. They have had previous experience in the course or they have moved through the coursework more rapidly than their peers.

Ferster's Method

Ferster's (1968) early paper described his personalized method which included the use of oral interviews with introductory psychology

students. Ferster and Perrott (1968) detailed the procedures for conducting oral interviews in their book, Behavior Principles. Briefly, students are first given a set of study guide materials. The student studies the unit material, writes notes on the study guide forms, and practices orally presenting the material. The student then reports his readiness to the instructor. An interviewer (i.e., a listener) is assigned to listen to the student (i.e., the speaker) and to rate his performance. If the student demonstrates mastery of the material, he is allowed to proceed to the next unit. If mastery is not demonstrated, the student is asked to review the unit material and retake the interview later.

Michael's Method

Michael (1971) has compared his method with Keller's. In Michael's procedure, a lecture, demonstration, or film is given in the first part of each week. Next, the unit exam is given to the class from text- and lecture-related study questions issued earlier. Students who pass the unit exams are excused from class for the remainder of the week. Those students who fail to pass the unit exam are allowed to recover all, or at least most, of their lost points by attending a remediation lecture and by passing the remediation test later in the week.

Summary and Discussion

In Keller's method, students are allowed to pace their own test taking. They are not required to attend lectures. But, they must make contact with the written word and maintain a high level of mastery of

the material in small steps in order to proceed through the course. They receive assistance from their proctors who are knowledgeable in the subject matter. In Ferster's method, the student must demonstrate oral mastery of the material before he may proceed. Keller's method uses written quizzes with some oral interaction; Ferster's method uses oral quizzes with some written examinations. In most other respects, the method described by Ferster (1968) is similar to the method described by Keller (1968). Michael's method includes a combination of teacher-paced and student-paced instruction. Students may eliminate about half of their class contact time by simply passing the unit exams on the first attempt; or, if they are more daring, they could simply pass all of the remedial quizzes on the first attempt. But, they are restricted by a limited-hold characteristic: the exam must be passed on the second attempt, or by the end of the week. Michael's method also requires attendance at lectures or other class activities at least once each week.

The results of studies comparing Michael's method with traditional methods (see Cooper & Greiner, 1971) resemble the results of studies comparing Keller's method (see McMichael & Corey, 1969) and Fester's method (see Sheppard & MacDermot, 1970) with traditional methods. Thus, with respect to learning, attitudes, and withdrawals from the class, contingency management techniques produce similar results, at least in the three studies cited above.

Preparing for a Personalized Course

McKeachie (1974) has stated: "The Keller plan, well done, requires a great deal of planning and effort" (p. 172). At least two published manuals illustrate this notion. Corey and McMichael (1970) have prepared a procedural manual for the operation and planning of a course using Keller's method. Born (1970a, b) has published a full and complete outline of Keller's procedures including a detailed description of pre-class preparation and maintenance information. In addition, Green (1974) has published, through the Center for Personalized Instruction, a handbook which teaches prospective teachers how to use PSI. The handbook uses a PSI technique. A summary of some particularly important characteristics follows.

Written Materials

The text selected must be chosen with care. Small and explicit units of material should comprise the text. Study guides need to be prepared. These guides should consist of an introduction to the unit, a set of procedures to follow through the unit, a set of study questions to answer, and a set of post-test questions following study of the unit. The introduction should motivate the student to further reading. The procedures should guide the student through the necessary material. The study questions should be pertinent to the procedures and to the test questions. Study guides can be used to present new material or expand issues in the text. Students should be expected to demonstrate mastery of the study guide material but they should not

be expected to master the entire text on the unit quizzes. Quizzes should be watched closely, as the self-paced course proceeds, to correct ambiguities and difficult areas. By attending to the students who are proceeding ahead of others, deficiencies in the material can be corrected with few or no errors made in teaching a unit.

Proctors

Proctors are important to PSI courses and instructors usually select them with great care. Proctors should have demonstrated above average grades, obtained satisfactory recommendations from other instructors, and have a reputation for honesty. Often it is necessary to use proctors concurrently enrolled in the class. This practice is most common at the beginning of a new school year and when a new class is formed. One way to insure reliable performance from proctors is to replace those who are unsatisfactory with students who are eager to serve. Close supervision of proctors is always necessary.

Grading

Some instructors have used an A grade for students who complete all units and lower grades for completion of fewer units. Other instructors have used final exam scores to determine the letter grade for the course and have required completion of all units in order to take the final examination. An incomplete is often assigned to those who fail to complete all units by the end of the course. In all PSI courses, grades are determined by individual performance on pre-defined objectives and not by a class curve. The instructor should always specify the grade requirements at the beginning of the term.

Lectures

Lectures are difficult to give in self-paced courses because student progress at different rates through the subject material. Usually only students who have progressed to a certain point in the course at the time scheduled for the lecture are permitted to attend. One problem with lectures is low student attendance. Even when attendance is allowed only for those who are progressing rapidly through the course, few of these students attend. Several studies have pointed toward locating the source of the problem with lectures, and some solutions have been proposed. But lectures should not be rejected out-of-hand because they may serve as a form of instruction as useful as any other mode of presenting information. Self-pacing simply makes lectures more difficult to use due to "individual differences" in moving through the course.

Work Load

The implementation of a PSI course is time-consuming. Keller (1971) has recommended that an instructor not admit more than 100 students, he should not accept responsibility for other tasks at the same time, and the material should be prepared well ahead of the first meeting. Before the beginning of the course, rooms should be selected and prepared, and proctors should meet with the instructor and be assigned specific roles.

Costs

Costs of PSI courses depend largely upon the equipment used and whether the help is paid. Students enrolled in teacher-training practicums for class credit can eliminate some of the financial outlay for the class. Some instructors have used students concurrently enrolled in the class, as mentioned above. More paper and other materials are also needed in PSI courses. Some instructors have used a lab fee from all students enrolled in the class to pay for the needed supplies. Although no detailed analysis of cost is available, the increase in quality of instruction is probably worth the increase in cost (Gallup, 1969).

Other Sources

Born's (1970a) manual provides a set of instructions for the proctor and the instructor. An appendix is included which details the preparations for the course. In another manual, Born (1970b) further described the development of a PSI course for the instructor and included several useful tools for conducting the class. Additional information regarding the implementation of PSI courses can be found in Malott and Svinicki (1969) and Sherman (1974). No detailed manual has yet been provided for implementing Michael's method. However, Cooper and Greiner (1971) and Michael (1971) provide some useful suggestions for such a course procedure. Ferster's method is detailed in Ferster and Perrott (1968) with an emphasis on the interviewing technique.

Lectures and Attendance

Lectures in PSI courses were originally used by Keller (1968) as a motivational device. Lectures were used to stimulate interest in psychology for students who had already mastered the material. Those who were eligible to attend lectures consisted only of those students who demonstrated mastery of a minimum number of units by some specified time. Even under the stringent requirements, only about half of those who were eligible to attend lectures did so. There were no consequences for absence from the lectures and the only consequence for attending was the opportunity to listen to the lecture.

Lloyd, Garlington, Lowry, Burgess, Euler, and Knowlton (1972) examined three contingencies for class attendance. Students were sometimes given credit toward the final grade, sometimes they were given information relevant to a subsequent quiz, and one class was allowed admission to the lecture contingent upon assignment completion. The results showed clearly that most students attended class when discussions and quizzes were held in the same session. Class attendance gradually declined when there were no contingencies attached. Other reports have also noted gradual reductions in class attendance when lectures were not required (Hess, 1974; Powers & Edwards, 1971, 1974). In a recent report (Edwards, 1975), attendance at instructor lectures, guest lectures, and films gradually decreased as the term continued regardless of the reinforcement magnitude. Attendance, in Lloyd et al.'s (1972) study increased to above 90% when points for attendance were given toward the final examination. When questions on forthcoming

quizzes were answered in the lecture, attendance was functionally related to the number of questions answered. The results of the study clearly showed that attendance at class activities was related to its consequences.

The major method of instruction was most closely related to Michael's, although characteristics of Keller's method were present in Lloyd et al.'s study. Lectures, quizzes, and class attendance were teacher-paced, but some projects were self-paced and could be performed at any time by the students (see also Lloyd & Knutzen, 1969). The laboratory portion of the courses offered by Lloyd and his colleagues were close approximations to Keller's method.

Some designers of teaching systems have used variations of the lecture method successfully (Hergenhahn, 1972; Postlethwait, Novak, & Murray, 1969). Postlethwait and his colleagues refer to an audio-tutorial method in which students make much of their contact with the course material through cassette tape-recorded instructions. A bright spot in the resurrection of the lecture method as a possibility in self-paced courses has been a suggestion made by Hergenhahn (1972). Hergenhahn divided his theories of learning course into 10 equal segments. Each segment contained a taped lecture; a reading assignment; a list of names, terms, and concepts; and a set of discussion questions. Each of the segments contained a quiz and an alternative quiz form for use in the self-paced course. The results from his method using an added lecture were quite similar to those reported by Keller (1968); most of the grades at the end of the term were A's and B's.

Learning and Retention

McMichael and Corey (1969) attempted to show the applicability of Keller's method to general subject matter and its superiority to lecture methods. Keller's method was implemented for the experimental group while three control classes were used in which traditional lecture methods predominated. The results showed Keller's method superior in terms of student's final examination scores and their ratings of the course procedures. Fewer withdrawals occurred in the experimental class than in the control classes. Corey and McMichael (1974) and Corey, McMichael, and Tremont (1970) noted in follow-up studies that retention of the material was superior for PSI classes.

Cooper and Greiner (1971) examined retention using Michael's method in one class and traditional lecture methods in another class in introductory psychology. The procedures in the experimental class using Michael's method involved a test on the first class day of the week, a lecture or film on the second day, and repeated opportunities to retake the test on the third and fourth class days. Proctoring was done in class by other students concurrently enrolled. Test scores were checked by teaching assistants. In the control class, students were given lectures, demonstrations, or films during three class days each week. Each four weeks, all students were tested over the material covered during the four-week period. At the end of the term, a comprehensive examination was given to all students in both classes. Five months following the end of the term, a comprehensive examination was given to members of both classes as a retention test. Results showed that the

experimental students spent twice as much time preparing for class each week, liked the class better, scored higher on the exams, and retained the material longer than the control students.

Work Patterns

Ferster (1968) described the work patterns of individual students in his self-paced course. Attendance averaged 60% to 70% daily. An average of 10% to 20% of the students who attended on a particular day did not take quizzes on the same day. The frequency of interviews taken by the students increased as the course progressed. According to Ferster, the typical work performance of the student was "scalloped." That is, performance was slow to start and gradually the rate of work increased as the course progressed. Of 91 students enrolled, 79 completed the course for credit. Thus, 13% of the students failed to complete the course requirements. Final grades reported for those who completed the course were 90%--A's, 4%--B's, and 6%--C's.

Lloyd and Knutzen (1969) used a variation of Keller's method which in many respects also approximated Michael's method. Several activities were scheduled for completion by the end of the term, each with a certain point value. For a D grade, the student was required to obtain 310 points; for a C grade, the student needed 410 points; for a B grade, 510 points were required; and for an A grade, the student needed to secure 600 points. Class attendance was one way to gain points, but activities other than lecture attendance were usually self-paced. In an analysis of weekly progress, records indicated a

tendency toward "break-and-run" performance by individual students. That is, students seemed to work at a high rate once work began, but some did not begin to work until it was too late to finish at the scheduled time. This result differed slightly from Ferster's (1968) report where scalloped performance was reported. It is important to note that the procedures differed for each investigation. Lloyd and Knutzen (1969) required a large amount of outside work while Ferster (1968) required oral interviews. It should also be recognized that the individual performance in Lloyd and Knutzen's study appeared as break-and-run performance, but grouped data appeared as a scallop. Final grades reported by Lloyd and Knutzen appeared more rectangular in distribution as opposed to the U-shaped distributions reported by Keller (1968) and others.

Sheppard and MacDermot (1970) described student performance in a course designed along the lines of Ferster's method. An indication of scalloped performances was shown for grouped data from the A, B, and C students. Withdrawals from the PSI class were 17% which was higher than in the control lecture-taught classes. It is notable that the students with the lower grades started work later than did the students with A grades. This result is similar to the grade data reported by Lloyd and Knutzen (1969).

Powers and Edwards (1974) reported some of the characteristics of self-paced student performance in a PSI-taught introductory psychology course. Most notable in their report was the variety of test-taking performances for individual students. The method used resembled

Ferster's. About 40% of the students showed a smooth linear rate of performance through the course; 25% of the students showed scalloping; and, 19% of the students showed break-and-run performance. Since mastery of unit material was equivalent for all students, the variable under question was the time to completion and the pattern of test-taking in the self-paced course. As Flammer (1970) noted, in PSI the time is the variable and learning is the constant. As Powers and Edwards (1974) showed, individual performance is not predictable as a single class of behaviors (see also Sutterer & Holloway, 1975).

Bonus Points for Early Work

A system which encourages students to complete their work by starting them off early may be helpful in PSI courses. In Lloyd and Knutzen's (1969) and Sheppard and MacDermot's (1970) studies, students who ended the course with lower grades tended to start their work later than students who obtained higher grades.

Bitgood and Kuch (1973) and Lloyd (1971) used points toward final grades as reinforcers for the completion of early work in self-paced courses. Bitgood and Kuch examined the effects of a graduated point contingency with 34 students. The basic system used was that of Ferster, although the interview sessions were described by the authors as informal. Points for passing unit quizzes were reduced each two weeks. Although no comparison group was available, several important results were obtained. First, students began completing the course requirements by the second week. Second, about half of the students completed

the course by the fifth week. Third, all students who remained in the class received A grades. Finally, it should be noted that a high percentage of students withdrew (17%).

Lloyd attempted a more systematic study of a graduated point contingency. In one group, students received bonus points on a percentage basis depending upon the date of completion. Work completed during the first two weeks received 60% additional bonus points. Work completed during the second two weeks received an added 30% bonus. By week 10 no bonus points were available. Students in the control group received no bonus points. Results showed that the students in the bonus section began work early and continued at a constant pace for the remainder of the term. Students in the group without bonus points paused at the beginning of the course and some began working quite late in the term. Although there was no statistical difference between the grades obtained by the two groups, the bonus group definitely had an advantage over the no-bonus group by beginning early and finishing early.

Cheney and Powers (1971) awarded bonus points to students who turned in work early. The work assigned was writing abstracts from published literature. No comparison group was used in the study. It is important to note that the rate of papers turned in was relatively constant. As the authors of the study pointed out, we would usually expect students to show an acceleration of papers turned in during the last part of the term when bonus points are not used.

Powers, Edwards, and Hoehle (1973) divided an introductory psychology class into two sections. In one section, a graduated point

system was used to reinforce exams taken early. This contingency was in effect for the first five weeks of the course. The second section was identical in all respects except that bonus points were not received. Results showed that more students in the bonus group took more exams in the bonus period than students in the no-bonus group. Students in the bonus group tended to finish the course earlier than the students in the no-bonus group. The results demonstrated that bonus points can effect early responding for students in a self-paced course. Powers and Wald (1974) similarly studied a graduated system in an introductory psychology class and found results similar to those of Powers et al. (1973).

Withdrawals

Born and his associates (Born, 1971; Born & Herbert, 1972; Born & Whelan, 1973) have described some of the characteristics of student withdrawals from Keller-method PSI courses. Born and Whelan (1973) detailed the behavior of withdrawing students. When compared with a lecture course, three to five times more students withdrew from the PSI courses (14-25%) than from the lecture course (5%). More students with poor academic records withdrew from PSI classes while few with high academic records did so. Most students who withdrew had passed the last exam taken, 40% had not had to retake an exam in the self-paced course, and 33% had been asked to retake more than one exam. Top students, with respect to grade point averages (GPA's), tended to pass more tests early in the course. But, there was no difference between the high and low GPA students for the number of tests passed by the end

of the course. Many of the withdrawing students (96%) were behind normal progress at mid-term. Those who remained (of those behind at the mid-term) were usually behind until the last one-fourth of the term.

At the time of the reports of Born and his colleagues, no procedures had yet been located which would retain students in the PSI courses comparable to traditional lecture courses. Similar high rates of withdrawing have been reported by others (cf. Sherman, 1974). Sheppard and MacDermot (1970), in a comparison of Ferster's method with traditional lecture methods, similarly found about three times more withdrawals from the PSI course (17%) than from the comparison lecture class (6%). In another study using Ferster's method, Powers and Edwards (1974) found that students who delayed initial responding in the classwork were most likely to withdraw from the course. However, studies report conflicting results. For example, McMichael and Corey (1969) using Keller's method reported fewer withdrawals from the PSI class than from the control classes. Austin and Gilbert (1974) in a physics course also showed fewer drops from the class taught using Keller's method than from the class taught using standard procedures. The problem of withdrawals from PSI classes has not yet been resolved (Keller & Sherman, 1974).

Attitudes

One important consideration in studies of teaching methods in college courses is the student's reported attitudes. Morris and Kimbrell (1972) measured the effects of the Keller method on performance

and on attitude reports. Comparisons of student attitudes under two types of teaching methods were made. The PSI portion of the introductory psychology showed more favorable final exam distributions. PSI students also reported favorable ratings for their class more often than the control students.

Witters and Kent (1970) examined 13 college courses involving 850 undergraduates and 6 faculty members over an 18-month duration. Student attitude reports generally favored the self-paced methods as reported on the surveys. In a more recent paper, Witters and Kent (1972) reported use of the Keller method in cultural anthropology and general psychology courses. Students in the general psychology course rated their enjoyment of the Keller method course more highly than their controls in the lecture course.

Most studies in PSI have shown from student reports benefits from the course. It is not yet certain what factors produce this result. It may be that instructors have simply conveyed their own excitement to the students. Some other possibilities might include the way the questions are worded, the lack of anonymity for the student, and the timing of the survey. The questions may be worded in ways leading to the answers the instructors wish. Instructors may ask the students to identify themselves and thus bias the report. Even without directly implicating themselves, many students believe that researchers have ways to determine their identity. Survey timing may be set so that the students are aware of the final grade, and thus be delighted with the course as a consequence of the final grade alone.

Proctoring in PSI

Several recent studies have examined proctoring in PSI under a number of various conditions. In a study by Farmer, Lachter, Blaustein, and Cole (1972), students in one group were given scheduled exams scored in the absence of the student. The exams were returned to the student allowing him to pass to the next unit if the exam was satisfactorily completed or to retake the exam if it was unsatisfactory. In other groups, students were present and proctored Keller-style during the term. Students in the groups were proctored over 25%, 50%, 75%, or 100% of the unit exams. Results showed that the students proctored on most of the exams received scores no different from students proctored on fewer exams. Students proctored on no exams, however, received significantly lower final exam scores. Students with proctored exams also retook significantly fewer exams than students without proctored exams.

Whitehurst (1972) reported observing weekly test scores for students under three conditions. The first condition required students to hand in written answers to discussion questions which were returned to students on the next day. A second condition required the students to engage in a one-to-one tutorial lasting 10 minutes one day each week. The third condition involved a group discussion in which all students were required to participate. The results showed fewer errors for written exercises and tutorial conditions than for the discussion condition on weekly exams. There was no difference between the oral and written conditions on the weekly test scores. When students were asked

which type of exercise they found most helpful for test preparation, and which was most enjoyed, the group discussion was selected. The second most helpful and enjoyed was the tutorial method.

Johnston (1971) studied oral and written verbal responding to written fill-in questions presented by proctors. Performance was recorded and defined in terms of rate of responding in one condition and number of responses in another condition. Results were not significantly different for oral and written quizzes. In a replication by Johnston and Pennypacker (1971), similar results were reported. Students in the oral group performed less well than students in the written group on first test attempts. But students in the oral group improved on later attempts to nearly equal the performance of the written group.

In an attempt to equate oral and written conditions, Edwards and Gottula (1973) first obtained volunteers from a psychology of motivation class. Students were then divided into either a self-paced oral or written procedure based on self-selection. Except for the mode of presentation by the student, all other characteristics of the course were identical. Students were used from their respective groups to assist the experimenters in administering the quizzes. Results showed students in the oral group obtained a higher mean score on the final exam than students in the written group; however, the difference was not statistically significant. Students in the oral group showed a high positive correlation between the number of quizzes given and their final exam scores. Students in the written group showed a high negative correlation between the number of quizzes given and their final exam scores. Both correlations were significant.

A recent study by Calhoun (1974) examined the performance of students in three different semesters enrolled in a psychology of personality course. Proctors were carefully selected from prior terms. About 150 students were tested by either oral or written methods each term. Section performances were tested at the completion of the unit series. Post-test scores showed significant differences between the number of sections completed and the pre- and post-test scores. No difference was found between oral and written testing on the post-test scores. But each procedure showed significantly higher scores compared with performance on units from which students were excused.

Summary and Discussion

The foregoing review of PSI literature was undertaken primarily to report the research in Keller's PSI and variations of Keller's method relevant to the present research. The focus has been placed on three central methods of instruction: Keller's, Ferster's, and Michael's. Several other possible methods were omitted from the discussion because they would add little information which pertains to the present research.

Keller's method was emphasized as the starting point for PSI although it is certainly not the first method to use each of its separate characteristics. Ferster's method deviates only slightly from Keller's putting a stress on the oral interview instead. Michael's method differs from both Keller's and Ferster's methods through its stress on instructor-pacing. Students in Michael's method are allowed to repeat a test on which mastery was not demonstrated, but only within

the test week during its administration. Under the added teacher-paced constraint of Michael's method, examination of certain variables is possible which the self-paced approach prohibits.

Preparation for a personalized course is viewed as a complex arrangement which differs considerably from preparation for more traditional classroom procedures. Prior planning, careful attention to detail, and step-by-step programming are required for an effective class. PSI programmers stress the use of several months to prepare materials, and that instruction should be limited to relatively small classes. The instructor should have few responsibilities involving teaching and non-teaching activities which might conflict with the PSI course. If the course has been fully prepared, the programmer can attend to the progress of the class and make any necessary corrections while it is in progress. Few traditional methods of instruction require more preparation from the instructor than his past lecture notes, and many of these methods allow little teacher-student interaction.

Lectures and attendance in PSI courses were reviewed. Methods for controlling attendance were suggested from the data, such as bonus points toward final examinations or final grades. Once attendance has been gained, learning might be enhanced through the use of study guides and taped lectures without drastic lack of the self-paced feature. Lectures in PSI courses might be most useful if tape-recorded so that students may listen to them at their own leisure. Taped lectures may also be replayed whenever particular points are not clear.

Learning and retention both seem to be increased through PSI procedures. It seems to matter little whether the course is at a beginning level or more advanced. This phenomenon may be a function of self-pacing which allows for student heterogeneity. It may be a function of the unit mastery requirement. It may be a function of the quiz technique (oral or written). Or it may be a function of any of a number of interrelated events. Coupled with these possibilities is that individuals who have difficulty with the material as a consequence of insufficient background skills may withdraw from the class prior to learning and retention measures. These possibilities need further exploration.

Withdrawing from PSI courses seems to occur in greater proportions than in traditional courses of instruction employing the lecture method. This appears to be related to students with lower GPA's even though PSI seems to be ideal for students without histories of success. Bonus points for earlier test-taking may reduce this tendency, but it is not yet known why this trend persists. A possibility is that PSI programmers follow a program of informed consequences to the point that the material to be covered frightens the student out of the class early. Traditional classes often fail to inform the student of the grade requirements in any clear fashion until the final grades are issued. Perhaps the criteria for grades should not be stated if apprehensive students and early withdrawals are a result of announcing criteria.

In spite of the problems in PSI-taught courses, students have tended to report favorably about the methods and materials used. Since

most surveys are made after the course is over, students in some cases have received their grades. In some cases students may not have received their grades, but because of the grade criteria, are fully aware of the grade they will be getting in the course. Since few surveys are well controlled, the attitude reports are, at best, only suggestive. However, it is important that the course is geared for the student, and that its activities are in fact reinforcing to the student. The PSI programmers should not neglect this aspect of instructional method, and they appear to conduct attitude surveys in nearly all reports.

Proctoring has been found important to learning in PSI courses. Studies comparing the effects of proctoring oral and written performances, however, have varied. One study showed no differences using weekly test scores as the dependent variable, but students rated oral proctoring more helpful and more enjoyable than written material proctored. Some studies have shown no differences using rate of question-answering as a dependent variable. And, one study showed no differences between oral and written proctoring on post-course test scores. Another study showed a high positive correlation between the number of oral quizzes proctored and final exam scores. In the same study, a high but negative correlation was shown between the number of written quizzes proctored and final exam scores. Since proctoring has been shown to produce better learning than no proctoring, it is concluded that the procedure is of particular importance. But, better techniques of proctoring should not be ignored. Although studies have failed to indicate differences between oral and written proctoring procedures, they may exist.

Robinson (1970) states that:

> One of the most effective devices to retard forgetting is ... to practice it in the way it will later have to be done. Since students have to show their learning through recitation in class or on tests, the student should practice reciting beforehand. (p. 28)

Evidence for the value of self-recitation immediately following the source (in this case reading) has been presented by Spitzer (1939). Students who were given a review test immediately after reading showed 83% on a test taken seven days later. Students who were given a review test one day later showed 46% on a test taken two weeks later. Students in the third group in which no review test was given, showed only 20% on the test taken two weeks later. It was concluded that more is forgotten soon after contact when retention is unaided than when retention is aided by recall.

Robinson (1970) attempted to answer the question regarding the form self-recitation should take. In doing so, he noted no distinction between oral and written self-recitation. In fact, he stressed that the most effective method used in the past has been that of reading a heading and then "jotting down" a key phrase from memory "in the reader's own words." Thus, although Robinson (1970) cited much evidence for recitation leading to "retarded forgetting," no clear distinction between oral and written recitation was made. Such a distinction was attempted in the present research.

This review has had the primary purpose of initiating the reader to the research conducted in various aspects of contingency-managed instruction used by Ferster, Keller, and Michael. As seen, the theory

of personalized instruction taken from the experimental analysis of behavior is beginning to form, but many questions remain to be asked. It is hoped that the reader has seen the importance of the proctor (i.e., the instructor's student help) in the method. It is on this importance that the present research is focused in order to determine a solid base for a teaching method which involves a written or oral quiz technique conducted by peers as proctors. The next chapter presents the first experiment in which oral and written quizzes as teaching techniques were compared. Chapter V presents a second experiment in which differences between the two techniques as well as the effects of rating peer performances on test scores were examined.

CHAPTER III

Experiment I

Robinson (1970) stressed that self-recitation was a powerful method of learning. However, no distinction was made between oral and written recitation. Teachers in the past have used written (i.e., "drill") rather than oral recitation as a teaching procedure. This preference may have been due to the ease of monitoring written work in groups and the difficulty of monitoring oral work except on a one-to-one tutorial basis.

In the present experiment, oral and written quizzes as teaching techniques were examined in two ways. First, two sections of introductory psychology students and two sections of psychology of adjustment students were compared for test scores when taught by the two separate techniques. Second, the techniques for the sections were reversed for each and a similar comparison was made. In order to examine quizzes as teaching techniques, all other conditions were held constant as nearly as possible.

Methods

Subjects

Seventy students enrolled in two sections of a course in introductory psychology and 60 students enrolled in two sections of a course in psychology of adjustment at Minot State College served as subjects in this experiment. Nineteen (10.2%) of the students withdrew from the

courses, one student was granted an incomplete due to illness, and one student was withdrawn from the data analysis because of a failure to complete one of the tests. Seven of the withdrawals were from the introductory course and 12 were from the adjustment course. Of the students remaining, 47 were female and 23 were male in the introductory course; 41 were female and 19 were male in the adjustment course. (See Table 16, Appendix E.)

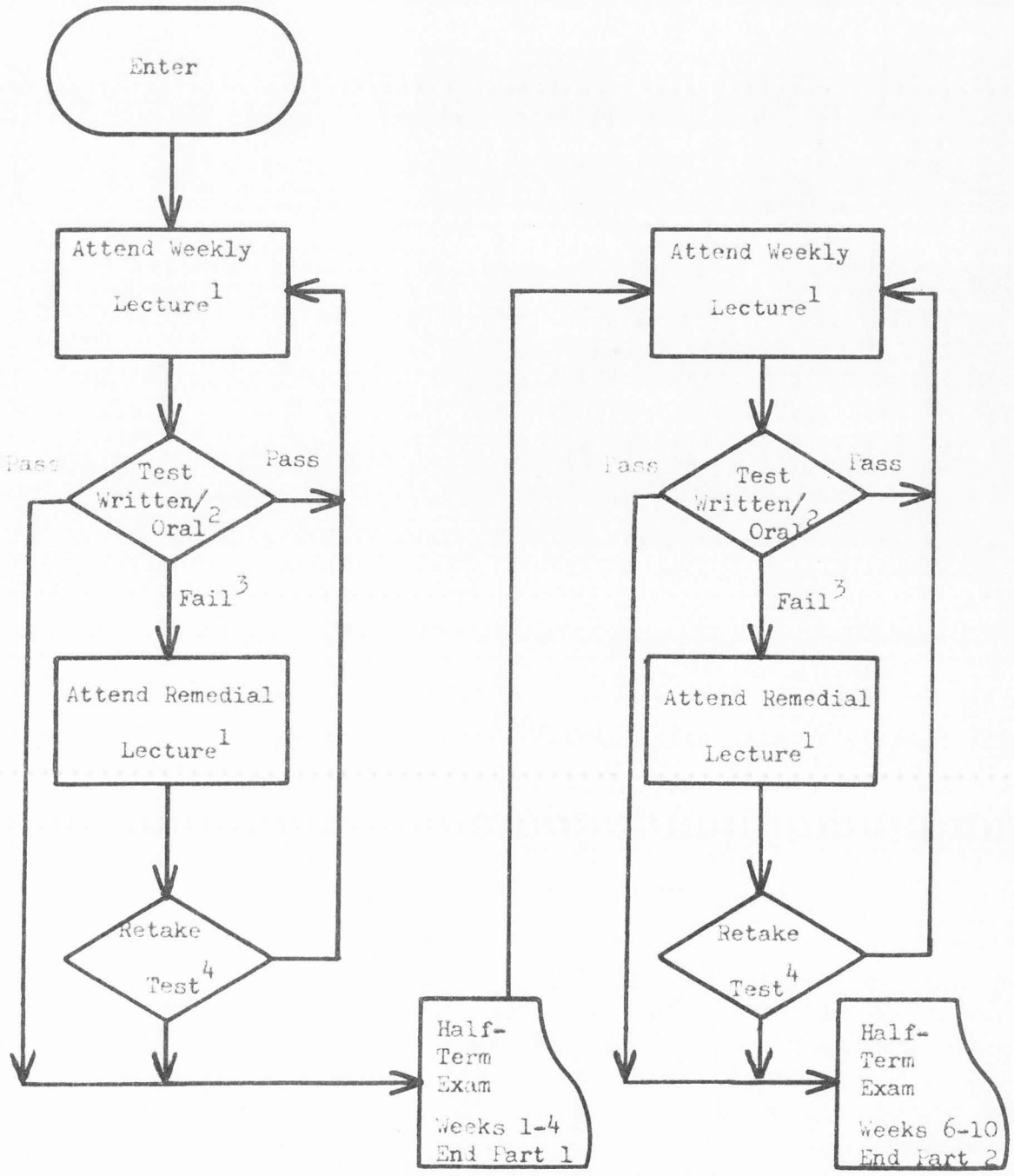
Apparatus and Setting

Texts used in the study were Psychology: An Introduction (Kagan & Havemann, 1972) in the introductory psychology course and Psychology and Effective Behavior (Coleman, 1969) in the psychology of adjustment course. Both courses offered 4 quarter-hours credit and both met four days weekly for one hour each day during the 10-week quarter. Quizzes and tests were held only at scheduled class times in the assigned rooms. In this experiment, quizzes were defined as oral or written performance measured by the ratings of assistants drawn from the class at large. Tests were defined as objective, instructor-graded performances at periodic intervals during the course without oral interaction.

Procedures

Assignment of the oral or written procedure to each section was decided by the toss of a coin. Figure 1 describes the course procedures in a flow-chart diagram. Each half of the course consisted of four units. A unit consisted of a set of study questions, a lecture, an oral or written quiz, a remedial lecture, and a quiz retake. The same

Figure 1. Flow-chart diagram of experimental course procedures.



1. Attending weekly lectures was worth points.
2. Each section took either an oral or a written quiz at different times.
3. There were three ways to fail.
4. Quiz values differed according to points obtained.
5. Final grade assignments may be found in the Appendix.

instructor gave all lectures and monitored all quizzes. Table 1 describes the course conditions for all sections in the two courses. Note that the first test was taken at the end of the first week of the course following lectures and study materials.

Table 1

Temporal Conditions for Each Section in Each Course

		Week				
Course	Section	1	2-4	5	6-9	10
101	A	Lecture & Test 1	Written Quizzes	Test 2	Oral Quizzes	Test 3
	B	Lecture & Test 1	Oral Quizzes	Test 2	Written Quizzes	Test 3
201	A	Lecture & Test 1	Written Quizzes	Test 2	Oral Quizzes	Test 3
	B	Lecture & Test 1	Oral Quizzes	Test 2	Written Quizzes	Test 3

Tests and surveys. Prior to test 1, students in their respective courses received identical lectures, study guides, and course suggestions. No quiz or test was given prior to test 1; thus, the sample of student performances was obtained under essentially normal classroom conditions. Following the first four units, during the fifth week of instruction, test 2 was given. Prior quiz material from weeks 1-4 was included on the test. Following the second four units, during the tenth week of classes, test 3 was given. Test 3 covered the material from the

last half of the course. All tests were multiple-choice and were identical for both sections of the same course. Although students were allowed to retake quizzes and test 1, no opportunities were available to retake tests 2 or 3. The tests used in the course varied in the number of questions asked and in the number of total points available. Test 1 consisted of 20 items, test 2 consisted of 50 items, and test 3 consisted of 100 items. Test 1 was worth 20 points, test 2 was worth 50 points and test 3 was worth 50 points. The first score obtained on test 1 was used in the data analysis. A questionnaire was handed out with test 3 asking for evaluations of certain aspects of the course and for some demographic data. At the completion of the first half of the term, the procedures were reversed for each section so that students were exposed to the opposite quiz technique. At the end of the second half of the term, students were asked to compare the oral and written quiz techniques.

Study questions and lectures. The first two days prior to the term were used to introduce the course procedures and explain the grading system. On the third day, study guides for the first unit test and the course syllabus were handed out. In all subsequent weeks of the term, a set of study questions for each unit was handed out on the quiz day prior to the next unit lecture. Study questions pointed to specific topic areas students were to focus upon for study, and all quiz items were taken from the study questions. Study question materials and quiz items were prepared by the instructor. Five points were granted for attendance at the weekly lectures. Five points were also

given to students attending the remedial lecture who had obtained less than 90% on the first weekly quiz of any given week.

Oral and written quizzes. All weekly quizzes were oral or written, short-answer, essay-type. The oral and written quiz materials were identical for both sections of the same course. If the student scored 90% or better on the quiz as rated by one of his peers, he was allowed to exempt himself from attending class for the remainder of the week.

Remedial lectures and quiz retakes. A remedial lecture and quiz retake were given to students who failed to demonstrate mastery (90%+) on a unit rated by his peers (student assistants) on the first opportunity. If the student retook the quiz, reduced points were given. The grade category (pass, questionable, rotten, fail) on a particular quiz was determined by the percent obtained for correct answers on the quiz (90%, 80%, 70%, or less than 60%) respectively. Points obtained on the first quiz of the week corresponded with the grade category at 100, 35, 20, and 0 points, respectively. On the quiz retake, points were adjusted downward at 60, 35, 15, and 0, respectively, with respect to grade category. In all cases, students were able to obtain maximum points available for the week's unit if a rotten or a fail was not obtained at either quiz session, and if a pass was obtained on the retake. Final course grades were determined by points accumulated at the end of the course (A = 884+, B = 808+, C = 733+, and D = 658+).

Student assistants. Students who obtained the top 10 scores on test 1 assisted in rating students on the first quiz. Students who assisted with rating performances on the second unit (first quiz) were

also exempted from the unit quiz and received full credit for the unit. Points were issued to assistants to equate total available points for all students. Five points were granted for attendance at the first weekly lecture, 50 points were granted for attendance at the quiz session, and 50 points were given for attendance at the retake session. It should be noted that quiz answer keys were provided the assistants to aid in rating quizzes. Guidelines for rating quizzes accompanied the keys. Assistants were not required to attend the remedial lectures. No student was permitted to serve as an assistant for two consecutive weeks.

Selection of assistants for a particular quiz during the first half of the term was made by choosing those students with the top 10 scores on the previous unit quiz. If more than 10 scores were tied for the top, selection was made by assigning each of those students a number and selecting 10 students by using a table of random numbers. In the second half of the term, all students were considered equally eligible to serve as assistants. Selection was made on the quiz day at the beginning of each session by assigning numbers to each student and randomly selecting 10 assistants by using a table of random numbers.

Assistants rating oral interviews asked a question, allowed the student to respond in his own words, and continued asking questions until the set was completed or until the student terminated the interaction. Similarly, the assistant rating written material allowed students to write answers to a dittoed form of the questions, one at a time. The assistant rated the answer while the student worked on the

next question. Subsequently, the sheet was turned back to the student who handed over the next answer, etc. Assistants and students were allowed to discuss the material after the quiz was completed. Each assistant rated the performance of two students in most sessions. At the end of the session, assistants turned in the rating forms with the student's grade recorded. (See Appendix C for the evaluation form used.) Both the student and the assistant were asked to agree on the rating granted before leaving the classroom.

Data analysis. The primary dependent variable in this study was the test score for each student on the multiple-choice test administered after each treatment. Differences between section test means were analyzed using a t test for independent means (Ferguson, 1971) to examine initial group differences following lectures in test 1; to examine differential treatment effects in test 2; and to examine the effects of a reversal of the differential treatment in test 3. It was assumed that the members of each section were drawn from different parent populations but that the populations were essentially similar. No attempt was made to examine inter-rater reliability. An additional dependent variable was the ratings on the questionnaires for attitudes and interests.

Selection of the two groups was initially determined by enrollment in two different sections, and assignment was not governed in any way different from previous terms. Analysis of the means of two groups in each course was made by an analysis of covariance (Winer, 1962). Lee (1975) has suggested that the analysis of covariance is inappropriate

to use in within-subjects comparisons. Therefore, separate analyses were made for each test taken during the term by comparison with the American College Test (ACT) scores as the covariate. ACT scores were available for 65 students in introductory psychology and 48 in psychology of adjustment. These scores are commonly used as college predictors of success.

Test scores were analyzed for those students who completed all three tests. The three conditions included contact with lecture only, oral quizzes, and written quizzes. Chi-square comparisons (Siegel, 1956) were made for each question asked on the questionnaires accompanying test 3.

Attitude and interest data from the survey following test 3 were assessed on five factors: learning, studying, interest, informativeness, and new acquaintances. Students were asked to compare the course taken with other courses taken during the same term with respect to those five factors. Analysis was made using the one-sample chi-square comparison procedure (Siegel, 1956). Students were also asked how they would recommend the course. Students who would not recommend the course were combined with the students who "condemned" the course for the analysis. Analysis was made by using the one-sample chi-square comparison (Siegel, 1956). Finally, a one-sample chi-square analysis of oral versus written procedural preferences (Siegel, 1956) was made for classes, sections, and combined courses. (See Appendix D for the course evaluation form.)

Results

Learning

Mean percent test scores for each section in the two courses are shown in Figure 2. (See Appendix A for raw test scores.) In the first panel showing scores for the introductory psychology class, little differences between means can be seen for sections. Note for test 1 that the test scores following lectures were virtually identical. In the second panel, the mean of the scores on tests following oral quizzes were higher in test 2 than the mean test scores after written quizzes. Since the main effect examined is the difference between means for individual tests, the drop in scores which occurred for both classes is not of particular importance in this study. The reason for the reduction in scores may have been due to increased test difficulty but it is not clear what may have produced it.

The number of students completing the course, mean test scores, and test standard deviations for each section in the two courses are shown in Table 2. Using a t-test for independent groups to compare mean differences for each test, no differences were found between sections of the introductory psychology course on test 1 ($t = 0.1198$, $df = 69$, n.s.), test 2 ($t = 0.1849$, $df = 69$, n.s.), or test 3 ($t = 1.1870$, $df = 69$, n.s.). Differences were found between sections of the adjustment class indicating higher mean scores for section B on test 1 ($t = -2.3857$, $df = 59$, $p < .025$) and on test 2 ($t = -3.4000$, $df = 59$, $p < .005$). No differences were found on test 3 ($t = -1.1201$, $df = 59$, n.s.).

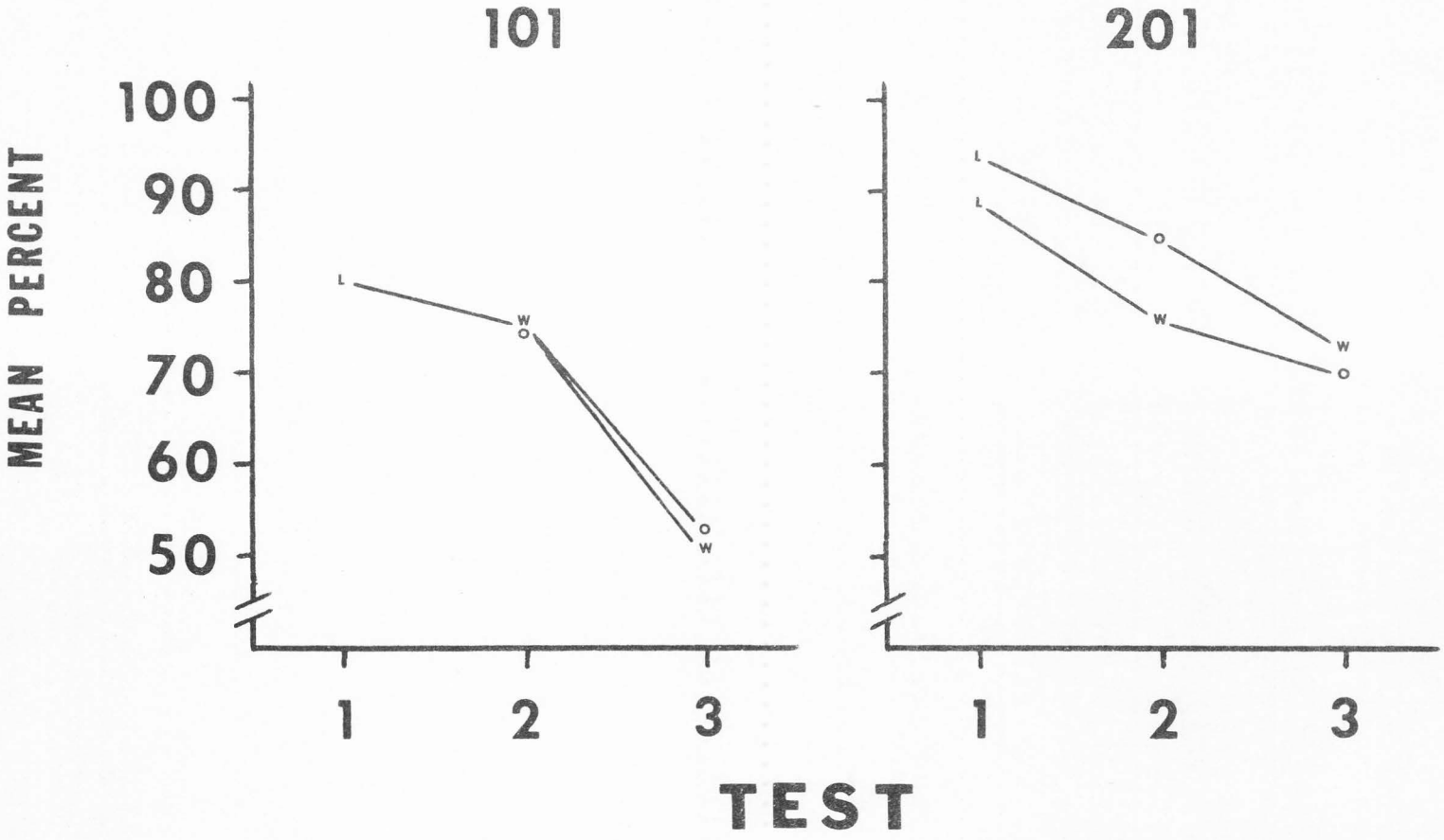

101
201
MEAN PERCENT
100
90
80
70
60
50
L
W
O
1
2
3
TEST

Table 2

Means and Standard Deviations of Test Scores
On Test 1, Test 2, and Test 3 for Each Class Section in Each Course[a]

			$\bar{X}$ Test Score			σ Test Score		
Class	Section	N	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3
Introductory	101A	33	80.61L	75.00W	52.55O	11.91L	7.59W	9.09O
	101B	37	80.81L	75.14O	51.03W	11.82L	9.19O	9.79W
Adjustment	201A	32	89.06L	76.09W	70.25O	7.98L	11.65W	9.92O
	201B	28	93.93L	84.50O	72.82W	6.85L	6.35O	7.57W

[a]Events preceding each test are indicated by L (lecture), W (written quizzes), or O (oral quizzes) just to the right side of the test score entry.

An analysis of covariance for the introductory psychology course was conducted to determine whether the lack of differences in the means between sections could be attributed to initial differences. Using ACT scores as the covariate for each separate test, no differences were found between means of test 1 ($F < 1$), test 2 ($F < 1$), or test 3 ($F < 1$). An analysis of covariance was also conducted for the psychology of adjustment class to determine whether the significant differences between means on test 1 were attributable to initial group differences as measured by ACT test scores as the covariate. With the alpha level set at .01, differences between means in test 1 were not significant ($F = 4.77$, $p > .01$), differences between means in test 2 were significant ($F = 7.44$, $p < .01$), and differences between means in test 3 were not significant ($F < 1$). These results, in combination

with the earlier t-test results, suggest that the differences between means on test 1 was a result of initial group differences, and that the treatment effects of oral quizzes on test 2 was a real effect. (See Appendix B for the tabled analyses of covariance.)

Attitudes and Interest

Figure 3 summarizes five measures of attitudes and interest for the combined sections of the introductory psychology class. (See Appendix E for tabled course evaluations not described here.) Students were asked to rate the class as "more," "same," or "less" on some characteristic compared with other classes taken in the same term. Only the number of students reporting "more" or "less" were included in a one-sample chi-square analysis (Siegel, 1956). The students reporting "same" were not of interest. The important comparison was the proportion of students indicating a preference in favor of, or opposed to, the characteristic in question. Assuming that an equal proportion of students would report "more," "same," or "less," differences on the "same" measure could have produced significant differences when "more" or "less" measures were not, in fact, different. For example, about 21% of the 57 students in Figure 3 who responded to the question on informativeness reported that the class was "more" informative than other classes taken in the same term while about 20% reported that the class was "less" informative. About 60% of the students reported that the class was as informative as others taken in the same term. If "same" were included, using a chi-square analysis, the three measures would probably yield significance. Since the "more" and "less"

Figure 3. Distribution of student's reported comparisons of the introductory psychology class with other classes taken in the same term. Except where noted, N = 57.

PERCENT
80
60
40
20
0
Learning
Studying
N = 59
Interest
Informativeness
New
Acquaintances
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS

measures are not different, the one-sample chi-square analysis would yield no differences if the "same" measure was excluded.

Proceeding across the figure from left to right, a greater proportion of students reported that they had learned "more" than those reporting "less" (chi-square = 6.1250, df = 1, $p < .02$). A significantly greater proportion of students also reported "more" studying in the course than those reporting "less" studying (chi-square = 21.5641, df = 1, $p < .01$). There were no differences in the proportion of students indicating interest in the course (chi-square = 2.2857, df = 1, $p > .05$) or informativeness (chi-square = 0.1818, df = 1, $p > .90$). Finally, a greater proportion of students reported that "more" new acquaintances were made in the course than those who reported "less" (chi-square = 5.7619, df = 1, $p < .02$).

Figure 4 summarizes the five measures of attitudes and interest for the psychology of adjustment course. Proceeding across the figure from left to right, a greater proportion of students reported that they had learned "more" in the course than students who reported they had learned "less" (chi-square = 9.9655, df = 1, $p < .01$). A significantly greater proportion of students reported "more" studying in the course (chi-square = 14.2353, df = 1, $p < .01$). A significantly greater proportion of students reported that the course was "less" interesting (chi-square = 7.2000, df = 1, $p < .01$) and "less" informative (chi-square = 4.2632, df = 1, $p < .05$). Finally, there were no differences in the number of new acquaintances made in the course in comparison with other courses (chi-square = 1.8000, df = 1, $p > .10$).

Figure 4. Distribution of student's reported comparisons of the psychology of adjustment class with other classes taken in the same term. Except where noted, N = 49.

PERCENT
80
60
40
20
0
Learning
N = 48
Studying
Interest
N = 48
Informativeness
New
Acquaintances
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS

Given a choice between oral or written quizzes as teaching techniques, a greater proportion of students in the introductory psychology class preferred the oral procedure (chi-square = 29.0000, df = 1, $p < .01$). Students in the psychology of adjustment course showed no preference for a technique (chi-square = 2.5745, df = 1, $p > .10$). A significantly greater proportion of students in the introductory psychology class (chi-square = 15.5106, $p < .01$) and in the psychology of adjustment class (chi-square = 4.2353, df = 1, $p < .05$) reported that they were willing to recommend the class to their friends.

Summary of Results

Mean scores on tests taken following oral quizzes were significantly higher than scores on tests taken following written quizzes for students at the midpoint of a psychology of adjustment course. Later, at the end of the quarter when the two sections had been placed under opposite conditions, no significant differences were seen. Mean scores on tests taken by introductory psychology students following oral quizzes were not different from those test scores following written quizzes.

Attitudes and interests on five measures were compared within classes on the basis of student's self-reports from questionnaires at the end of the term. A greater proportion of students in both courses reported learning and studying "more" in the class than students indicating learning and studying "less" than in other classes taken in the same term. Students in the introductory psychology class reported interest and informativeness in the class equivalent to other classes.

Significantly more students in the psychology of adjustment class, however, reported that the class was "less" interesting and "less" informative than students reporting "more." These comparisons were between "more" and "less" categories only. The "same" category was excluded.

In terms of several social questions, students in the psychology of adjustment course indicated no more new acquaintances than in other classes taken, but a greater proportion of students in the introductory psychology course indicated they had made more new acquaintances in the class. Students in both courses reported a willingness to recommend the course to their friends. Other data from the survey indicated a significant preference for the oral procedure in the introductory psychology course.

Discussion

With respect to learning, no differences were seen in the introductory psychology class between section test means. With respect to attitudes, a greater proportion of students in the introductory psychology class indicated a preference for the oral technique rather than the written technique. In the psychology of adjustment course, higher mean test scores were obtained when the oral quizzes preceded the tests than when written quizzes preceded tests on test 2. No difference was seen when the procedures were reversed. Noting this, an order effect is suggested. In essence, when students become accustomed to one technique, the effects may be irreversible; or, as Born, Gledhill,

and Davis (1972) suggested, changing procedures in the middle of a term may be disruptive.

Campbell and Stanley (1963) cite a study similar to the present one in which initial differences between the means of two groups were found on a pre-test, but on the post-test the differences were eliminated. This was observed on the psychology of adjustment comparisons using the t-test for independent means for each separate test during the course. An analysis of covariance was applied to determine the effects of initial differences. The analysis of covariance was similarly applied to that course to determine whether some initial difference may have been responsible for the "no difference" finding. It was concluded that the initial differences did not affect the lack of significance in the introductory psychology course. It was also concluded that initial differences in the psychology of adjustment course were not responsible for the mean differences in test 2 or the lack of differences on test 3. Thus, the interpretation that the procedures were responsible for the test results is supported.

Analysis of student's self-reports for both the introductory psychology class and the psychology of adjustment class showed concurrence with reports of other studies using Michael's method (Cooper & Greiner, 1971). In the adjustment and introductory courses, a greater proportion of students reported "more" learning and "more" studying than those reporting "less" when compared with other classes taken in the same term. However, on "interest" and "informativeness," the psychology of adjustment class did not fare well--students rated

that course significantly "less" in both categories. Although it was expected in both courses, only the introductory psychology class indicated more new acquaintances were made when compared with other courses taken in the same term. Since the students in the adjustment class were more advanced, perhaps there were fewer new acquaintances to be made on the small campus.

Although no significant differences were found in terms of learning in the introductory psychology course, it is apparent that the students preferred the oral procedures when it was taken first. It has been well demonstrated that PSI techniques are effective in introductory psychology classes (Born & Herbert, 1971; Corey & McMichael, 1974; McMichael & Corey, 1969). Possibly beginning psychology students may not be affected by subtle differences in those techniques such as oral and written quizzes, at such an early date in their college careers. Johnston and Pennypacker (1971) showed little differences between oral and written procedures using rate of response to fill-in questions as a measure of learning. Since students indicated that they were willing to recommend the course to their friends and, in general, the oral procedures were preferred, perhaps more experience in college life is necessary before the techniques affect learning.

Several questions remain to be answered. Since class members served as either students or proctors in different class sessions, the effect of proctoring and being proctored was confounded. These two variables need to be separated and analyzed further. There is also a question of design. Lee (1975) stated that the application

of the analysis of covariance to within-subjects designs is inappropriate. Thus, the analysis was used separately for each test in the experiment. A recent report indicated several possible designs which are appropriate to use with an analysis of covariance (Maxwell & Cramer, 1975).

In conclusion, the design difficulties in the present study render an unequivocal interpretation difficult. The design difficulties are as follows: 1) students in the two sections of each course were not randomly selected, 2) the three tests were not equated for difficulty, 3) because of the lack of random selection, the use of non-equated tests, and the within-subjects design, an analysis of covariance may have been inappropriate, 4) scores on tests were progressively lower, thus the level of difficulty may have increased on tests until a "floor" effect was obtained. By a "floor" effect, it is meant that the means could not have been lowered. A second experiment sought to correct for these difficulties.

CHAPTER IV

Experiment II

Evidence of order effects in the first experiment suggested the need for an additional study using a randomized treatment design. In the present experiment, effects of oral and written quizzes on final examination performances in a teacher-paced course were compared. The first experiment also failed to separate the confounding effects of students serving as proctors. Thus, the second experiment was performed so that the effects of rating student's performances on oral and written quizzes could be investigated.

The rationale underlying this research was that there may be differences between taking oral and written quizzes, between rating oral and written quizzes, and between any other pair of these exposures to coursework on test scores. The exposure to quizzes may be considered as both teaching and studying techniques (see Robinson, 1970).

Methods

Subjects

Seventy students were enrolled in two sections of a course in psychology of adjustment (Psy 201) at Minot State College. The two sections met at different times of the day. Thirty-eight students were enrolled in section A and 32 were enrolled in section B who completed the requirements for the study. Seven additional students (8.6%) withdrew before the end of the study and four students were excluded from

the learning data due to incomplete information. All 74 students were included in the attitude and interest analysis. Data from 70 students were plotted for pre-course and post-course test result comparison. Data from 50 students were included in a nonparametric Friedman rank sums test for multiple comparisons (Hollander & Wolfe, 1973).

Materials

A pre-course test and an identical post-course test, each consisting of 140 multiple-choice instructor-made test items, were developed from instructor-made pre-recorded lectures and from Watson and Tharp's Self-Directed Behavior (1972). A similar form of the tests was previously administered to students at the end of an extension course in psychology of adjustment. The extension students had no prior contact with the text or the tape-recorded lectures, but the students were exposed to similar material through the course. Most of the questions for the tests used in the present study were taken from the extension course exam. The criterion for test item selection was that 25% to 75% of the students taking the test answered the item correctly. Additional questions not included on the original test were developed and added to complete the experimental requirement of 10 questions for each of the 14 units.

Quiz questions were related to the pre- and post-course tests. From each unit consisting of 10 test questions, multiple-choice type, four to seven quiz questions were constructed. Each quiz question consisted of one to five parts. As nearly as was possible, elements

from all major concepts covered in the course were tested over on both the quizzes and the tests. (See Appendix G for a sample quiz.)

Course Procedures

The design consisted of two experimental conditions, oral or written quizzes following lectures, and a control condition in which no quiz followed the lecture. One section of the class was run under a series of randomly determined experimental and control conditions. The other section was run under a similar, but separately generated, set of conditions. Table 3 shows the order of presentation for each section. For example, section A and B students were given the pre-course test on the first class day. On the second class day, a written quiz was given to section A students and an oral quiz was given to section B students following the unit 1 lecture. On the third class day, no quiz was given to section A students and section B students were given a written quiz after the unit 2 lecture. On the fourth class day, an oral quiz was given to section A students and no quiz was given to section B students after the lecture on unit 3. On the fifth class day, section A and B students were given an exam over the three units and additional questions were asked from future units. Questions on the future units were not counted in the exam scores transmitted to students. The fourth unit quiz was given on class day six, and so on. Note that section B students were absent on class day 19 and were not exposed to the unit 14 lecture. This absence was due to unavoidable flood conditions in the community and was not related to the course. The data obtained on the pre- and post-course

Table 3

Course Procedures and Conditions for Each Section of the Psychology of Adjustment Class[a]

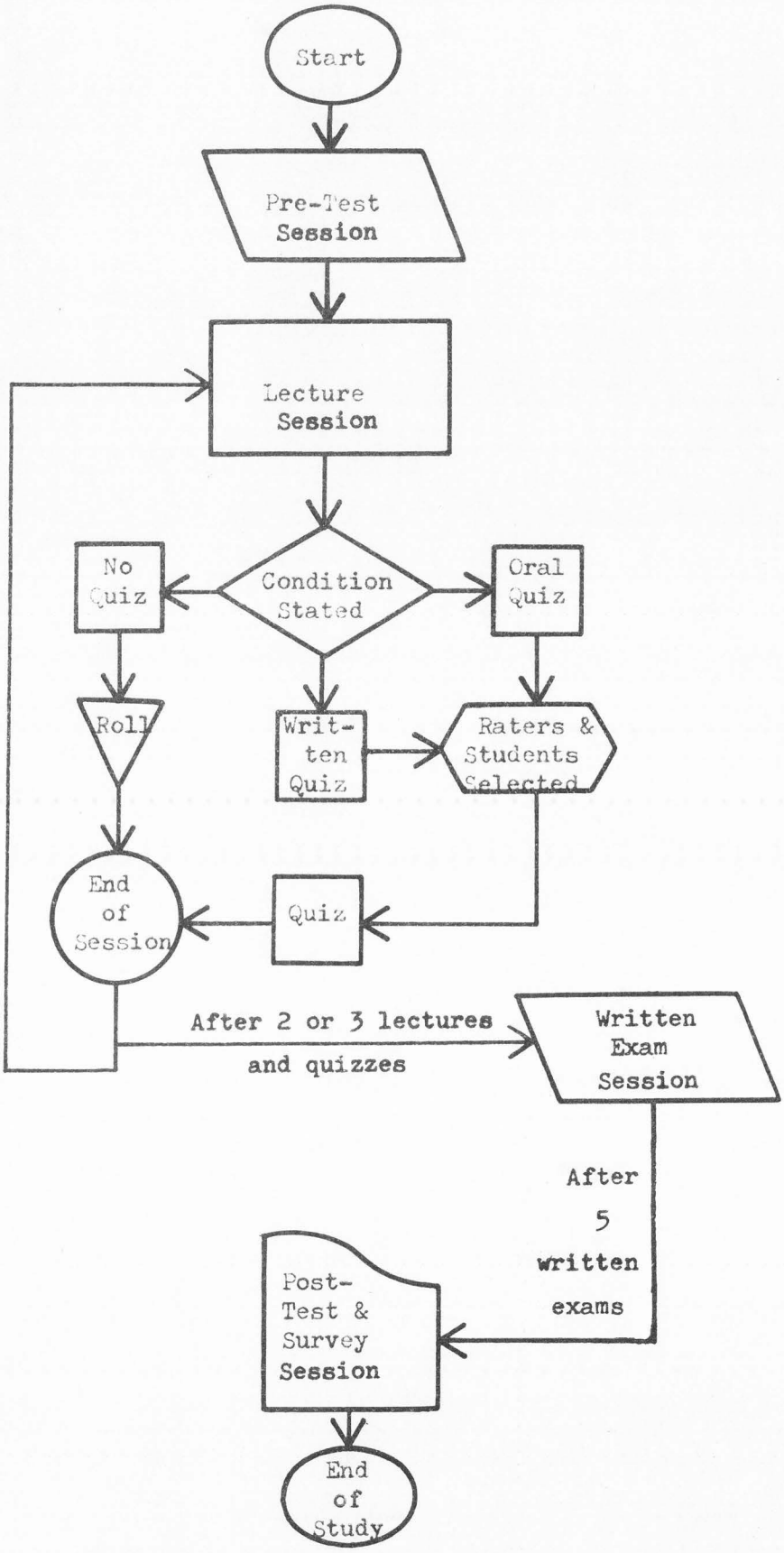
	Class Day																				
Section	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	Pre	W	N	O	E	W	O	N	E	O	W	W	E	N	O	N	E	W	O	E	P
B	Pre	O	W	N	E	N	W	O	E	N	O	N	E	W	O	W	E	O	A	E	P

[a]Symbols: Pre=pre-course test; O=oral quiz; W=written quiz; N=no quiz; A=all students absent; E=weekly exam; P=post-course test and survey of attitudes and interests.

tests for unit 14 for the section were disregarded in the critical test using the Friedman rank sums test for multiple comparisons as were all other quiz absences for individual students.

Two additional experimental conditions included the rating of oral and written quizzes. As students were selected to take quizzes, so were their peers selected to rate quiz performances. Under the rubric of oral and written quiz conditions, then, are quiz taking and quiz rating. These techniques of exposing the students to the study of unit material were expected to yield differential scores on the post-course test given on class day 21. The major dependent variable for learning was the difference scores obtained by subtracting the pre-course unit test score from the post-course unit test score. Figure 5 shows a flow-chart diagram of the course procedures described briefly above and in greater detail below.

Figure 5. Flow-chart diagram of instructor-paced procedures in the psychology of adjustment course.


Start
Pre-Test Session
Lecture Session
No Quiz
Condition Stated
Oral Quiz
Roll
Writ-ten Quiz
Raters & Students Selected
End of Session
Quiz
After 2 or 3 lectures and quizzes
Written Exam Session
After 5 written exams
Post-Test & Survey Session
End of Study

Lectures and exams. Lectures were taped on cassettes and lasted 15 to 30 minutes. The lectures were played during class days and were immediately followed by an experimental or control condition. The taped lectures insured that both sections of the course received identical information from the lectures. After each third lecture, students were given a written exam. Experimental conditions varied with control conditions so that each condition occurred twice weekly except in the last week when the written conditions occurred only once. Sessions in which students were absent were excluded from the data analysis. Students absent from sessions frequently enough to exclude any whole condition were excluded from the analysis of learning data. In order to maximize learning under all conditions, exams were preparatory and cumulative. For example, an exam for students in section A following lectures 4, 5, and 6 on class days 6, 7, and 8 consisted of six items from lectures 1, 2, and 3; five items from the written quiz on unit 4; five items from the oral quiz on unit 5; five items from the no-quiz unit 6; and two items from each of the units to be taken, units 7-14. A total of 37 items were included in the exam. The weekly quiz items were worded differently, and the choices were ordered in a different way, than the pre- and post-course test items, to avoid teaching right or wrong answers on the basis of familiarity.

Instructions to the students. Students were told of several conditions at the beginning of the lecture series: (a) The daily quizzes were to be used as a measure of proficiency. (b) The scores on the quizzes partially affected the final grades. (In fact, the

scores did not affect the final grade.) (c) Results of quiz performances would be made available to individuals interested in their own scores when the 14 lectures in the series of taped lectures was completed. (d) A large part of the grade was dependent upon attendance at lectures. (e) Completion of only the lecture series and the post-course test would earn the student enough points for a D grade in the course. (f) Additional opportunities for higher grades would follow the lecture series. (g) Any student who did not do well on the quizzes would be given an opportunity to make up the deficiency after the lecture series was completed.

Quizzes. In oral quizzes, students answered short-answer essay questions aloud to one or two peer raters. Once a question was terminated by beginning another question, the rater(s) marked the points obtained on a form handed out after the lecture and before the quiz. There were no more than seven questions on a quiz and no questions contained more than five parts on which a student was rated. On each part of the quiz, completely clear and correct answers were rated two points, muddled answers which contained key terms without logical order were rated one point, and unclear or absent answers were rated no points. The student's score was totaled, the sum was divided by the total possible points for the particular quiz, and the result was multiplied by 100 to obtain the quiz percent score. Raters were instructed to avoid indicating ratings to students, but students were allowed to obtain the rating from the instructor if requested. This procedure was used to reduce peer pressure from students who were taking the quizzes.

Written quizzes were similarly conducted except that students wrote answers to dittoed questions. Students were instructed to write a single answer, hand it to a rater, and write another answer while the rater(s) read the first. This procedure allowed raters to evaluate answers while students were answering other questions. If a second rater was used, the unmarked answer sheets were turned over to the second rater as soon as the first rater's evaluation of the questions was completed. As with oral ratings, students were instructed to request ratings from the instructor if they were desired.

Students in the two experimental conditions were permitted to use notes, books, or other materials to "jog their memory" but they were not permitted to substitute reading for a recitation. The instructor monitored all quizzes by moving around the room as unobtrusively as possible. All materials were collected as soon as each student and rater completed the quiz. These materials included all written answers to quiz questions and handouts but did not include lecture notes.

Under the no-quiz control condition, students were excused from class immediately after the lecture and after they had signed the roll sheet.

Rating of performances. Raters were selected from students immediately after each lecture by a random procedure. Following the taped lecture, the experimental condition of oral or written quiz was announced as taken from a previously scheduled randomized order, or the control condition of no quiz was announced. If the condition was

oral or written quizzes, students were randomly selected to take the quiz or rate the performances of their peers. A form specifying the experimental conditions and instructions for interviewing was randomly distributed among the students. (See Appendix I for rating of quiz instructions.) Those who received quizzes were required to report answers to the raters. Raters were those who received copies of the instructions, a score sheet on which to record performance scores, and keys with which student's answers were judged. Students and raters were matched by pairing numbers written on the quiz sheets and the instruction sheets. Two to six raters were selected at each quiz in addition to those students and raters already paired. These additional raters were used to obtain inter-rater reliability information through random pairing. (See Appendix J for student evaluation forms.)

Learning data analysis. The main effects under study in this experiment were the difference scores obtained from unit pre-course and post-course test scores after taking oral quizzes, taking written quizzes, rating oral quizzes, rating written quizzes, or taking no quizzes. A Friedman rank sums test for multiple comparisons (Hollander & Wolfe, 1973) was used to determine whether the sums of the ranked scores for students in each treatment differed significantly as measured on difference scores. Since the two sections received the conditions in different order, the data from both sections were examined separately as well as combined. Note that all students received all conditions where the Friedman rank sums test for multiple comparisons was used. Practice effects were examined by comparing the results

from the first half of the course with those of the second half using a sign test (Siegel, 1956).

Attitude and interest data analysis. Student's ratings of the course were obtained immediately following the lecture series along with the administration of the post-course test. Student's opinions of the different class procedures of oral and written quizzes were requested. Other questions were asked related to learning, informativeness, studying, interest, and new acquaintances in the class compared with other classes taken in the same term. In addition, students were asked whether they were willing to recommend the course to their friends. A one-sample chi-square comparison (Siegel, 1956) was used to determine differences between "more" and "less" while answers of "same" were omitted from the analysis for reasons given in Experiment I. (See the course evaluation form in Appendix K.)

Reliability data analysis. Inter-rater reliability was determined in three ways: (a) The number of rater's agreements and disagreements were counted and the agreements were divided by the sum of the total observations. (b) The total agreements in one condition was summed and the total observations in the condition was divided into the number of agreements for the condition. (c) The number of disagreements in ratings of no points, one point, and two points were summed; the number of disagreements ranging by one point were summed; and the number of disagreements ranging two points were summed. The number of disagreements ranging one point was divided by the total number of disagreements and multiplied by 100 for the percent of one-point disagreements. Finally, the number of disagreements ranging two points was

divided by the total number of disagreements and multiplied by 100 for the percent of two-point disagreements. Percent disagreements were also found for oral and written conditions in each section. (See Appendix H for the reliability data with individual students.)

Validity data analysis. Validity of the rater's reports was analyzed by comparing judgments on individual quizzes with difference scores for related unit test items. First, the mean of a pair of ratings for one student's quiz performance on one unit was found. The student's score was then ranked along with those scores of all other students with paired raters for that quiz. Next, the student's difference score for the unit was obtained by subtracting the pre-course test score from the post-course test score for the unit. The difference score for the student was ranked along with those of other students who had been rated by two raters. Finally, the Spearman rank correlation coefficient with correction for ties (Siegel, 1956) was used to obtain the correlations. Since judgments by the raters were expected to be positively correlated and significant at the .05 level, the one-tail test was used to evaluate the null hypothesis stating p less than or equal to .05. If a high and positive correlation was obtained, the rater's judgment of the student's quiz performances could be considered valid. Otherwise, quiz performances as rated by the raters would not be a valid measure of learning as measured by difference scores.

Results

Reliability and Validity

Inter-rater reliability checks which were made for quizzes by an item-by-item agreement divided by the total possible observations ranged from 56% to 100%. Out of a total of 1680 items observed, agreement was seen on 1274 items. Reliability on this measure averaged 76%. For oral quizzes, ratings averaged 77% reliability. For written quizzes, ratings averaged 74% reliability. Item-by-item agreements were consistent for sections and conditions.

Examinations of all disagreements with 1- or 2-point spreads between scores showed 77% 1-point differences and 23% 2-point disagreements. Section B showed 74% 1-point differences and 26% 2-point spreads. Comparison of quiz methods for sections A and B showed similar results with one exception: in the written condition for section A, only 17% 2-point disagreements were shown.

Table 4 shows the correlation of ratings with difference scores in an analysis of the rating validity. No significant correlation was found between rating and test scores in section A; but in section B, a high and positive correlation was found in one instance. Only in this one instance was it possible to reject the null hypothesis ($p < .05$). Since the probability of one out of 19 correlations appearing as significant is high, it can be concluded that its appearance was due to chance. In one written quiz and one oral quiz, high but negative correlations were shown. These distributions of correlations indicate a lack of validity in ratings of quizzes compared with test performances at the end of the term.

Table 4

Correlations of Quiz Rating With Difference Scores on Pre-Course and Post-Course Tests

Section	Type of Quiz	Lecture 1	2	3	4	5	6	7	8	9	10	11	12	13	14
201 A	Oral	-	-	-.049	-	-.101	-	+.044	-	-	-	-.279	-	-	+.322
	Written	+.007	-	-	+.093	-	-	-	-.076	+.206	-	-	-	+.207	-
N		16		16	16	15		15	15	15		13		12	14
201 B	Oral	-.192	+.445	-	-	-	-.703	-	+.603[a]	-	-	+.487	-	-.128	-
	Written	-	-	-	-	-.863	-	-	-	-	+.117	-	+.310	-	-
N		15	14			11	12		10		12	9	8	9	

[a] $p < .05$; one-tail test.

Learning

Figure 6 shows the means of students' pre-course test scores before oral, written, and no-quiz conditions, and post-course test scores following exposure to these conditions. (See Appendix F for the individual means.) Recall that 10 questions were devoted to each unit of the pre- and post-course tests. Note also that the data for students taking the quizzes and students rating the quizzes are combined in the figure. The mean number of correct items in each unit quiz were consistently higher after treatments for each section. Table 5 shows the pre- and post-course test means, standard deviations, and variances for each section of the psychology of adjustment class. Note that these data are based on 140 possible points. Thus, in percentages, the pre-course test means for sections A and B are 44% and 47%, respectively. The post-course test means for sections A and B are 59% and 62%, respectively.

Difference scores of pre-course and post-course test scores for the first seven units were compared with the difference scores for the second seven units. It was possible that the performance of the last half of the course was somehow related to performance in the first half of the course in a number of ways. For example, if scores on the second half of the test were higher, this might imply a practice or recency effect in the sense that the material learned in the latter half was learned later in the course and closer to the post-course test. On the other hand, the cumulative effect of asking questions from earlier units on each exam could produce higher scores on the

Figure 6. Mean correct answers for each unit on the pre-course and post-course tests in two sections of a psychology of adjustment course. Sections are represented by the letters A and B on the pre-course test; on the post-course test, the quiz techniques used are represented by O (oral), W (written), and N (no quiz). (Scores from 70 students were plotted in each of the tests: pre-course and post-course.)

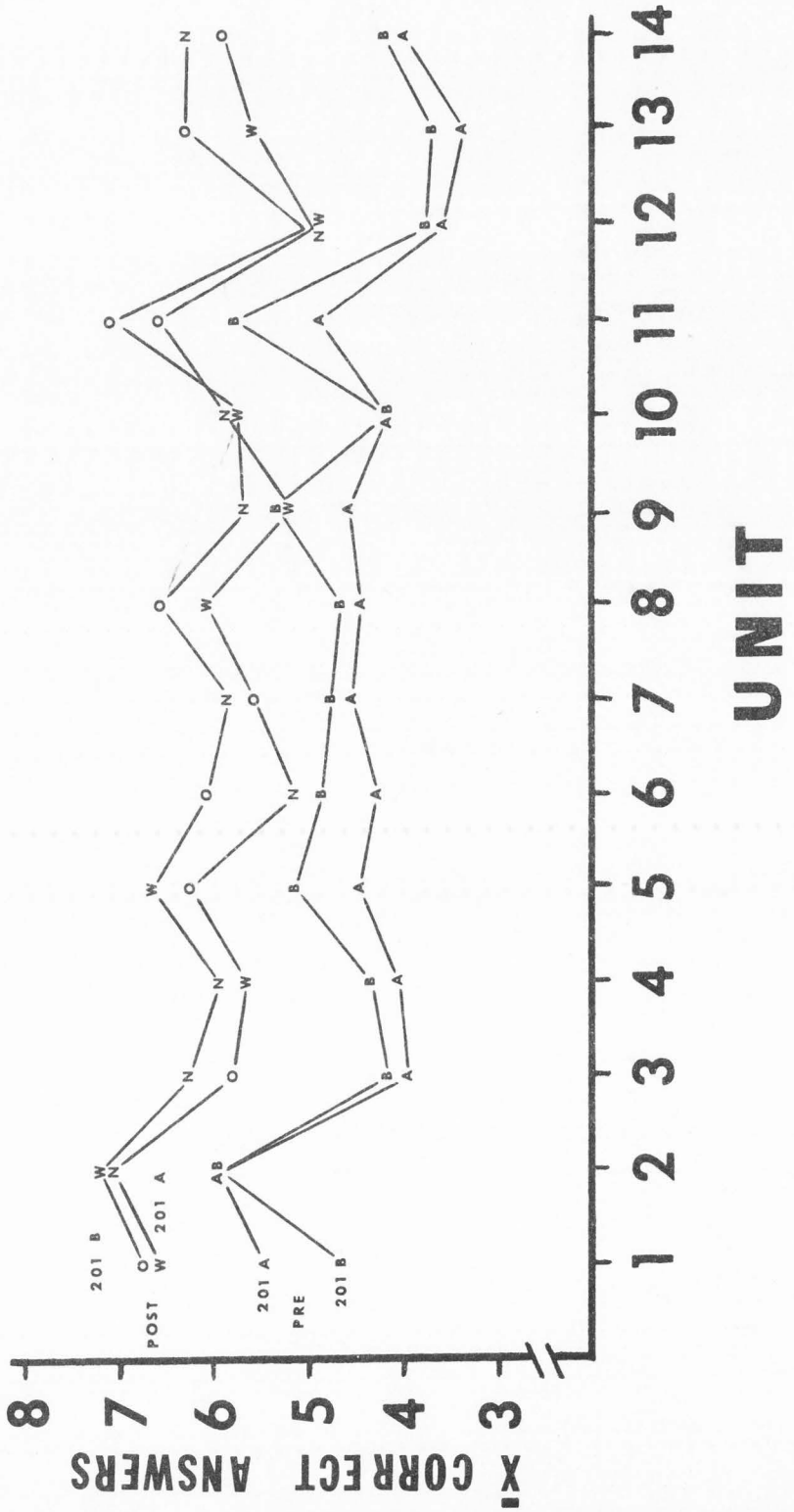

X̄ CORRECT ANSWERS
8
7
6
5
4
3
1 2 3 4 5 6 7 8 9 10 11 12 13 14
UNIT
POST
PRE
201 A
201 B
201 B
201 A

Table 5

Pre-Course and Post-Course Test Means, Standard Deviations, and Variances for the Psychology of Adjustment Class

Section	N	Pre-Course Test			Post-Course Test		
		$\bar{X}$	SD	V	$\bar{X}$	SD	V
A	38	62.24	10.30	106.58	82.58	16.14	260.47
B	32	65.41	8.97	80.38	87.28	15.60	243.50

first half of the test. Scores for each student's post-course test from units 1-7 and 8-14 were each summed and the difference obtained. The student was then assigned a plus, minus, or tie. Sums of signs were obtained and the sign test (Siegel, 1956) was applied. The hypothesis of no difference was confirmed ($Z = -1.375$, n.s.).

In Figure 7, the mean difference scores for each section under oral, written, and no-quiz conditions are shown. It is clear that there were no significant differences for conditions. The scores for students taking and rating oral and written quizzes are combined in the data. These differences were obtained by subtracting the pre-course test score from the post-course test score for each student and summing the columns. Data from all 70 students were used.

Figure 8 shows the distribution of summed ranks for each condition based on differences between pre- and post-course test scores. Distributions of summed ranks shows the highest ranks were obtained when students rated other student's oral performances (OR) and the lowest ranks were

Figure 7. Mean difference scores for students in both sections: pre-course test scores were subtracted from post-course test scores. Students in sections A and B are represented in their labeled histograms above the conditions designated by O (oral), W (written), and NQ(no quiz), (N = 70).

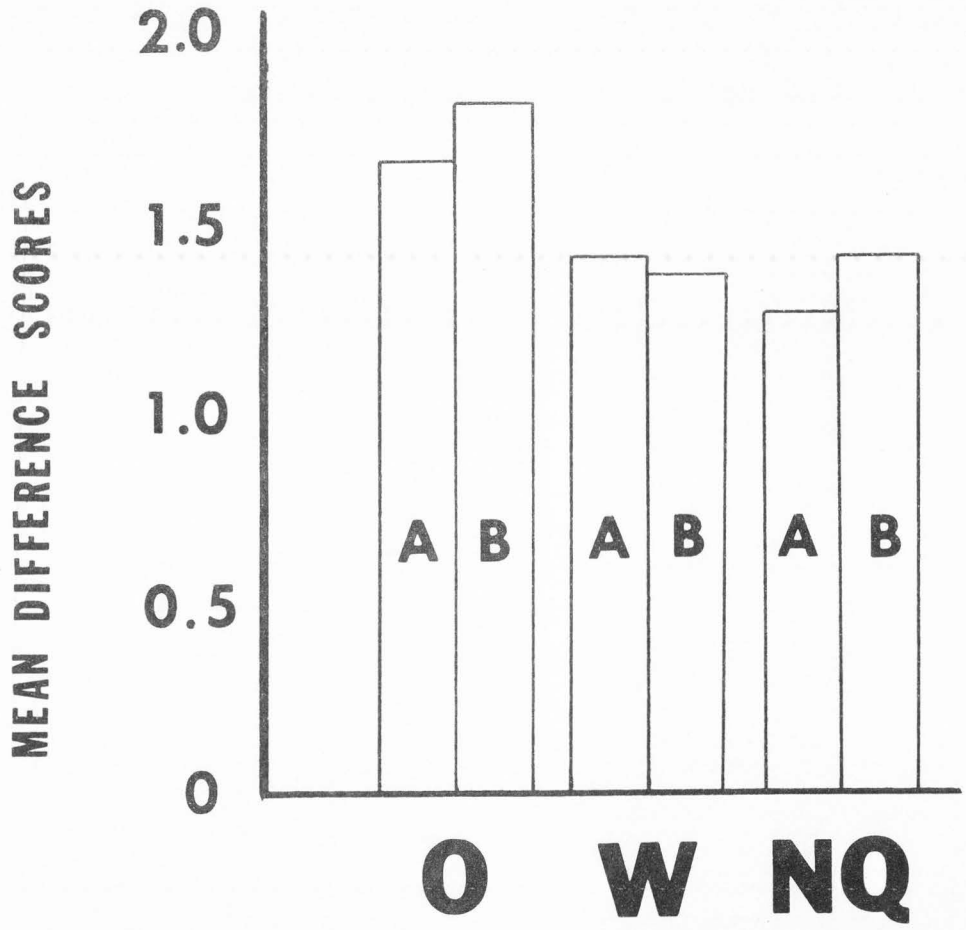
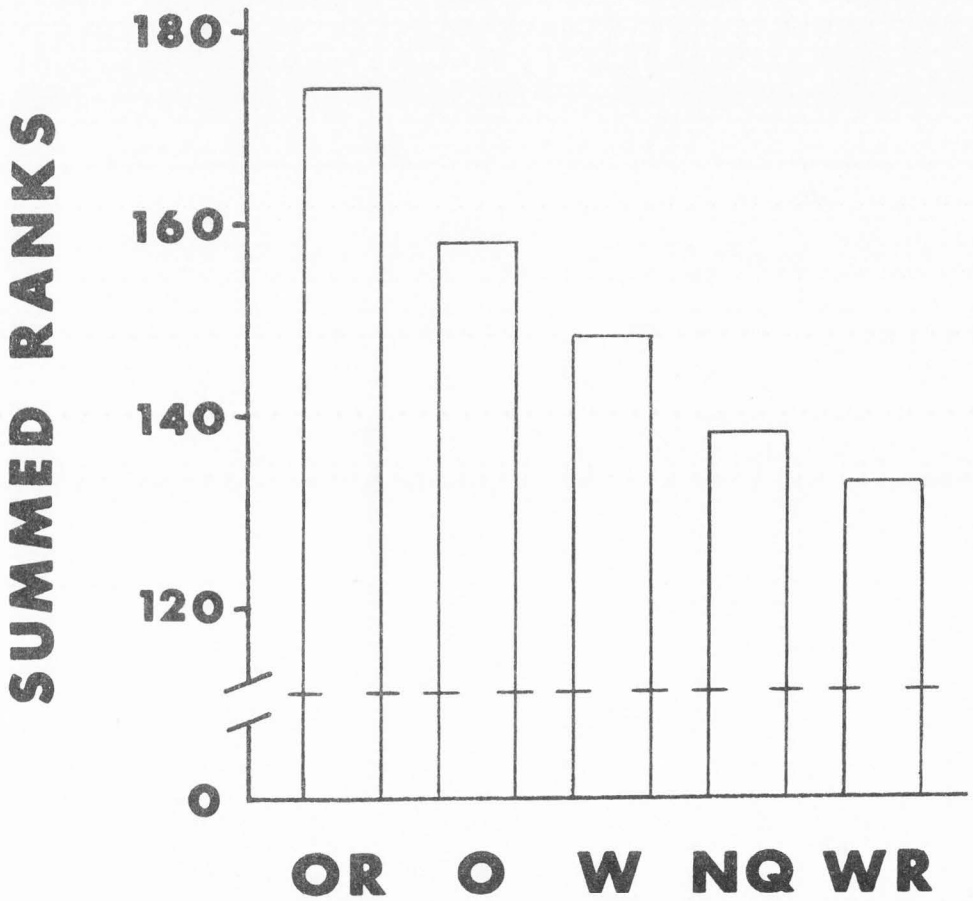

MEAN DIFFERENCE SCORES
2.0
1.5
1.0
0.5
0
A
B
A
B
A
B
O
W
NQ

Figure 8. Distributions of summed ranks based on difference scores between pre-course and post-course tests and the condition to which students were exposed: Oral quizzes (O), Written quizzes (W), or No Quizzes (NQ). Scores for student raters are represented as: WR (written rating) or OR (oral rating) (N = 50).


SUMMED RANKS
180
160
140
120
0
OR
O
W
NQ
WR

obtained after students rated other student's written quiz performances (WR). Application of the Friedman rank sums test for multiple comparisons showed no significant differences between any two conditions based on difference scores. It should be noted that the difference scores showed that oral rating and written rating difference approached significance ($p < .10$). Data analyzed included only that of the 50 students whose quiz activities involved at least one score from each of all five conditions.

Attitudes and Interest

Figure 9 summarizes five measures of attitudes and interest for the students in the class. Of particular interest is the number of students stating "more" or "less" favorable characteristics of the class compared with other classes taken in the same term. With respect to the five variables in the figure, learning, studying, interest, informativeness, and new acquaintances, there were no significant differences for each. Finally, students did not differ as to the number who would or would not recommend the course to their friends, nor did the students differ with respect to those who favored the oral or written procedure.

Summary of Results

Reliability of raters' item-by-item agreements was indicated at an overall average of 76%. This was relatively consistent for all sections and conditions. Of all disagreements, 77% differed by one point while 23% of the disagreements were spread by two points for section A and 74% were one point apart with 26% two points apart for section B. An

Figure 9. Distribution of student's reported comparisons of the psychology of adjustment class with other classes taken in the same term.

PERCENT
80
60
40
20
0
Learning
N : 72
Studying
N : 73
Interest
N : 72
Informativeness
N : 72
New acquaintances
N = 70
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS
MORE SAME LESS

analysis of rating validity indicated little correlation between raters judgments and difference scores with the exception of one unit in one section.

No differences between comparisons of the scores during the first half of the course with the second half of the course was found. Thus, a practice effect was not noted. A Friedman rank sum test for multiple comparisons showed no significant differences between each pair of five conditions using difference score data. Graphically, the highest scores were obtained on tests following oral rating, then oral quizzes, written quizzes, no quizzes, and finally, written rating.

On five measures of attitudes and interests comparing the class with other classes taken in the same term, it was also found that there were no differences on learning, studying, interest, informativeness, or new acquaintances. It was also indicated that students were no more inclined to recommend the course to their friends than to not recommend the course. Preference for the oral or written quiz technique was not found.

Discussion

Neither oral nor written quizzes were followed by higher test scores in two sections of a psychology of adjustment course. The scores for those who rated oral performances were not significantly different from scores for those who rated written performances or took no quiz at all. Analysis of difference scores as indicated by individual pre-course and post-course test scores failed to show that rating oral

performances provides any greater learning when statistically compared with several other forms of quiz taking. Graphically, rating written performances showed the least adequate form of preparing for final exams while rating oral performances appeared to be the most adequate form. However, these methods of preparation did not significantly differ. These results ($p < .10$) were suggestive for further research, though, as were the results from a similar study by Edwards and Gottula (1973). In the Edwards and Gottula study, oral quizzes were also suggested as preferable to written quizzes according to differential final exam scores between the two groups ($p < .10$). The Edwards and Gottula study also showed a negative correlation for rating written performances and final exam scores, and a positive correlation between rating oral performances and final exam scores.

The within-subjects design used in the present experiment rendered it difficult to use the traditional analysis of variance statistical test. Even such tests as those suggested by Shine and Bower (1971) and by Keselman and Leventhal (1974) for single subject analyses were inappropriate due to the unequal trials in the study. The only appropriate tests suggested in the literature were nonparametric (Hollander & Wolfe, 1973), particularly the Friedman rank sums test for multiple comparisons. It should be noted additionally that a great deal of sensitivity is lost using the nonparametric tests. Since the present study produced a close approximation to significance, it is suggested that a further study use a design allowing a more sensitive statistical test. It should be further noted that few individual students typified

of directly involving the student enrolled in the class needs to be considered before using outside help, though. It is notable that the validity of the ratings was virtually non-existent. This would indicate that the untrained rater's reports are not sufficient grounds for grading students taking quizzes. The data do suggest that the oral interaction may have some positive effects on learning; but, because of the lack of validity, this interpretation is at best tenuous.

CHAPTER V

General Discussion

Two experiments were performed to examine the effects of oral and written quizzes as teaching techniques on test performance. The purpose was to determine whether differences on examination performance follow instruction by oral or written quizzes. Additionally, it was sought to determine the effects of proctoring (i.e., rating) oral and written quizzes on the test performances of the proctors (i.e., raters). In the two experiments, the data were suggestive, but not conclusive. The results of one test indicated significantly higher scores on tests following oral quizzes compared with test scores following written quizzes. From the two comparisons of treatment effects made in the first experiment, then, only one showed significance. As in an earlier study (Edwards & Gottula, 1973), the data suggested that oral quizzes were superior to written quizzes.

There are several possible explanations for the failure to show a reversal in the first experiment. First, students under the control of oral quiz techniques prior to written quiz techniques may have performed poorly as a function of earlier procedures. Second, the students under the written quiz technique as a first condition may have been insensitive to an oral technique which followed it. Third, it may be that shifting course requirements in the middle of the term was disruptive to test-taking performance (Born, Gledhill, & Davis, 1972). Fourth, the third test may have been too difficult to show differences.

This was hinted at by a continued reduction of absolute test scores on each test as the term passed. Fifth, and perhaps most important, since all students served as proctors, scores for students taking quizzes and proctoring quizzes were confounded. Most studies have used experienced proctors, and Edwards and Gottula (1973) showed that proctoring written quizzes may have an effect producing poorer scores on tests.

Experiment II also provided suggestive evidence that rating oral performances may produce higher test scores than rating written performances on quizzes. Rating written quizzes was followed by lower ranked test scores than taking no quizzes. When scores were grouped, differences were not significant. In addition, no differences were found between oral and written quizzes compared with no quizzes. This lack of differences may have been due to the procedures involving no feedback in the second experiment. That is, perhaps the motivation provided by knowledge of outcome would have produced higher scores (Hilgard & Bower, 1975). This is, of course, subject to future experimentation.

Studies comparing oral and written quiz performances have consistently produced evidence that the oral component of quizzes are more important than the written component. Farmer et al. (1972), in a comparison of varying amounts of proctoring, showed that no proctoring produced the lowest final exam scores and the highest number of quiz retakes. Thus, the removal of the oral component produced poorer performance on both the quizzes and on the final exam performance. Whitehurst (1972) showed no differences between oral (i.e., tutorial)

and written quizzes on weekly test performances, but students rated the oral quizzes more helpful for test preparation than written quizzes. Edwards and Gottula (1973) showed higher final exam scores for students taking oral quizzes than for students taking written quizzes. Hursh, Wildgen, Minkin, Minkin, Sherman, and Wolf (1975) examined discussions following written quizzes. Students who received discussions following the quizzes retook fewer quizzes than students who did not receive discussions. In addition, students scored higher on general questions when discussions were used. Students nearly always elected discussions when given an option. As in the Farmer et al. (1972) study, the Hursh et al. (1975) study simply removed the oral component from a proctoring condition. It was found in both studies that the written component alone was less effective than both components operating together.

In the present research, Experiments I and II do not support the findings of the above investigations. In spite of a lack of consistent significant differences, most studies reviewed have indicated superiority for the oral quizzes over written quizzes. None of these studies, however, have determined the effects of rating quiz performances by students enrolled in the class.

Three major topics for general discussion will be considered below. The first topic of interest is that of raters and rating. The second topic deals with self-paced instruction compared with the teacher-paced procedures used in the present research. The final topic of concern is the implications of the present research.

On Rating and Raters

At first glance it seems contradictory that rating oral quizzes might be followed by higher (albeit nonsignificantly higher) test scores than rating written quizzes or taking no quizzes. Rating oral quizzes seems to be a "passive" learning approach. Since raters were responsible for judging student's performances on quizzes, there may have been a good deal of "internal" or covert responding. Since there may be a difference between rating oral and written quizzes, it is suggested that reading may be harmful to learning. By this, it is meant that reading without recitation may produce little in the way of testable changes in performance (see Fox, 1962). This hypothesis is not intended to suggest that listening to lectures may produce "better" learning. Rather, it is suggested that listening followed by immediate consequences might produce measurable changes in performance. The immediate consequences of rating one's peers may be sufficient to produce learning for the rater. However, the rating itself may be invalid with respect to the rated student's test performance as shown in the present research.

Proctoring seems to be the central characteristic of PSI. Without proctors, there can be no "personalized" system of instruction. Proctors can advise students on matters in which professors are unable, and they can advise the instructor on matters with which students find difficulty. The present research suggests that it may be the proctor of oral performance who is gaining the greatest amount of learning in the personalized courses. We should not, then, lose sight of the fact

that it is the student enrolled in the course who should be gaining the most out of the class; if we can find ways to include the student as a proctor, then we should develop those methods. The major concern in such a procedure is that the student taking the quiz may suffer while the proctor is learning more than is otherwise possible.

Several investigators have not thought it necessary for aids (i.e., assistants) to have prior experience in the class (Sherman, 1974; Edwards & Gottula, 1973). They simply selected students as proctors from class enrollees who proceeded at the most rapid rate. Proctors have also been drawn at random from enrollees (Gaynor & Wolking, 1974). From an a priori standpoint, it may seem most rational to choose students with the highest grade point averages or the highest class standing. Several problems are attached to those criteria for selection, however. For example, seniors may be little interested in performing any more work in the class than is necessary for acquisition of their degree, and college freshmen of high quality may be overlooked due to a lack of any measured history. ACT scores have been used as an apparently useful predictor (Wood & Wylie, 1975). Pre-tests have been used in one of my classes as a screening device for selection of student aids. However, close monitoring of aids was necessary. Any excessively absent or unprepared student aids were replaced immediately. Note that the above methods overlook the important aspect of proctoring in PSI: the student who is proctored may not be learning as much as the student who is doing the proctoring.

The reliability data in Experiment II indicated that the reports for untrained students as raters were fairly consistent and relatively

high. Gaynor and Wolking (1974), in a recent replication of Johnston and Pennypacker (1971), indicated that training is not a necessary condition for proctoring. Indeed, as the authors stated, the use of proctors drawn from within the enrollment may serve as a useful alternative to proctoring by previously trained students. Sherman (1974) suggested use of proctors from within the class as have others (Edwards & Gottula, 1973). Validity data in the present study indicated that rater's reports should not be the basis for a grade assignment. Thus, it is suggested that the ratings serve as adjuncts to grading procedures, but ratings should not supplant the final examination or other grading techniques administered by the instructor.

On Self-Paced Instruction

Future studies are needed to examine the relationship of self-paced instruction to learning and attitudes. The generality of the current findings would then be extended to further PSI instruction. Comparisons could be made for possible differences under oral and written quizzes in self-pacing, teacher-pacing, and teaching with deadline contingencies under more careful control than was used in the present study. It was found in this study that students in the totally teacher-paced class did not find the course more interesting or informative than other courses taken in the same term. In addition, no more students were willing to recommend the course than were willing to give the class a poor recommendation. Students were also relatively indifferent to the method of instruction (i.e., oral or written). In the first

experiment, when the class was taught using Michael's teacher-paced method with weekly deadlines, a greater proportion of students reported "more" learning and "more" studying than "less," but they indicated "less" informativeness. A self-paced summer class using similar materials, however, showed favorable attitudes and interest on nearly all measures.

It should be noted additionally here that 24% of the students enrolled in the self-paced course withdrew. By comparison, 10% of the students in Experiment I (Michael's method) and 8% of the students in Experiment II (instructor-paced) withdrew. These data are consistent with the withdrawal rates reported in other studies of PSI (e.g., Born & Whelan, 1973; Keller, 1968; Powers & Edwards, 1974).

From the attitude and interest data, the totally teacher-paced method of instruction is not recommended for a "personalized" course. In fact, the self-pacing used in a summer term using similar materials appears to be the most favored part of the entire course, except that the withdrawals were greater in the self-paced course.

Implications for Further Research

Follow-up studies of the foregoing experiments may produce additional information concerning the differential retention of material learned under oral quizzes, written quizzes, rating oral quizzes, rating written quizzes, and no quizzes. No prior study has examined retention of material taught under those methods with the exception of a follow-up of students in Edwards and Gottula's (1973a) study (Edwards & Gottula, 1973b).

In order to avoid the loss of degrees of freedom noted in Experiment II, a follow-up study should be considered. One such study could include random selection of students in each session to perform in one of the five conditions. An analysis of covariance for repeated measures could be used to examine the data of the 14 units. The pre-course test scores for each unit could be used as the covariate. Statistical treatment of such a methodological design should provide a more sensitive test than the Friedman rank sums test for multiple comparisons.

Nevertheless, several additional points have emerged from these experiments which demand further study. One, taking oral quizzes did not produce significantly different results from taking written quizzes. Two, rating oral quiz performances did not produce significantly higher test scores than rating written quizzes. Three, if students are taking quizzes, they cannot be concurrently rating quizzes in the way the two experiments were conducted. Four, if students take written quizzes, someone must rate them.

From what we know, students who are rating quizzes are in essence listening to a lecture, but they are required to respond immediately. They are required to note errors (even if invalidly) as raters. Rating oral performances somehow differs from rating written performances in the response required from the rater. We also know from the operant studies in vigilance that "error" responses can be examined on schedules of error presentation (Holland, 1966). It seems possible, then, to further demonstrate the effects of listening (i.e., rating) by presenting

"lecture" material on individual tapes and specifying schedules of intervals and ratios with varying and fixed errors. Thus, the student in a self-paced course may be examined on schedules of errors presented on tapes while he rates the "lecture." There are no "students" as quiz takers; therefore, all students serve only as raters of oral performances. An experiment designed to examine such behavior must, of course, provide the necessary controls and the appropriate design. Such a series of studies are planned.

There is a long way to go in PSI research. The present state of the art is such that the field is wide open for studies like those proposed above. The present research has indicated some useful information and related studies should be explored.

CHAPTER VI

Summary and Conclusions

In some earlier studies, weak support for oral quizzes has been provided when compared with written quizzes. However, the methodology and measurement designs used by the researchers provided no unequivocal evidence for either technique. Evidence from several studies have indicated that contingency-managed teaching techniques produce more effective student test performance and study behaviors than more traditional teaching techniques such as lecture or discussion. Further examination was needed to determine whether any actual differences exist in oral and written quiz teaching techniques.

This study was conducted to determine whether any differences existed between the effectiveness of oral and written quizzes as teaching techniques. In the first experiment, taught by Michael's method, quizzes followed lectures by a one day lapse with opportunities to retake a failed quiz. Tests were given at the middle of the term and at the end of the term to ascertain the effects of oral and written quizzes on each of the two sections of two courses. In the second experiment, taught using taped lectures with quizzes immediately following, students were randomly assigned to rate or take oral or written quizzes or take no quiz. Tests were given before and after the course and differences between the two tests were used to determine the effects of the treatment.

Summary

In the first of two experiments, one section under the oral quiz procedure showed significantly higher test scores than the section under the written quiz procedure on the second of three tests in a psychology of adjustment course. Sections of the introductory psychology course did not differ significantly on any of the three tests. Students in both courses reported that the class required "more" studying and that they had learned "more" than in other classes taken in the same term. Students in the introductory psychology class reported that they had made "more" new acquaintances in the class. Students in both classes reported that they were likely to recommend the course to their friends. The students in the introductory psychology class indicated a preference for the oral procedure. Students in the psychology of adjustment class reported that the class was "less" interesting than other courses taken in the same term.

A second experiment was conducted to remove many of the methodological problems encountered in Experiment I. Examination of inter-rater reliability showed consistent and reliable reports from paired raters. Examination of validity indicated little correlation between rater's judgments and student's test scores. Thus, the rating validity was low. A Friedman rank sums test for multiple comparisons showed no significant differences between any pair of the five conditions tested: oral quiz rating, oral quiz taking, written quiz rating, written quiz taking, and no quizzes. Examination of the oral quiz rating and written quiz rating rank sums showed higher rank sums for the oral quiz

rating approaching significance ($p < .10$). Students' attitudes and interests with respect to learning, studying, interest, informativeness, and new acquaintances showed no difference from other courses taken in the term. Neither the oral nor written procedure was preferred, and students were no more likely to recommend the course as to not recommend it.

Discussion and Conclusions

The first experiment lacked information concerning several variables. First, raters of the quizzes took the same tests as the students who were quizzed and the data were not separated. Second, one class under investigation showed no differences between means when the section data were compared while the other class showed significant differences in one test but not in others. Third, attitudes toward the oral procedure was significant only for the introductory psychology class. Fourth, a possibility of practice effects was noted. Fifth, inter-rater reliability was not investigated.

In the second experiment, it was noted that students rating other students did so reliably, but the ratings were not valid. It should be noted that the study employed a nonparametric statistical test. Although the differences between oral rating test scores and written rating test scores approached significance, no significance was shown between any two of the five conditions, oral quiz rating, oral quiz taking, written quiz rating, written quiz taking, and no quizzes. The data thus indicated the need for still further study of differences between oral and written quizzes.

The present experiments failed to unequivocally demonstrate differences between oral and written quizzes as teaching techniques. However, it is suggested that those students who rate the oral quiz performances of other students may benefit more than students involved in any other type of quiz-taking. It is further suggested that the students who rate written quiz performances may in fact be distracted from learning the material. Reliability checks indicated that students may rate each other reliably, but examination of validity indicated that student's evaluations were not useful indexes of student's test performances. An instructor may fare better by concluding the course with a fair final examination than by using rater's evaluations of quiz performances as his sole evaluative tool.

In Experiment I, students reported learning more and studying more, but they also found the course less interesting than other courses taken in the same term. In Experiment II, no differences in the number of students reporting "more" or "less" on any attitude and interest measure were found.

Further studies should examine retention and self-pacing with respect to the five conditions examined in the present research. In addition, any further studies should examine the variables using both group statistical methods and single subject designs. The group methods could include repeated random assignment of students to such roles as rating oral and written quizzes, taking oral and written quizzes, and taking no quizzes in a single session for several sessions. With such a design, a powerful parametric statistical test such as the

analysis of covariance for repeated measures could be used to examine group data. In addition, the single subject data can be used in a separate experiment using a multiple-baseline technique in which several students are progressively assigned to different roles.

Finally, since the studies suggested that rating oral quizzes without proctoring may be beneficial to learning, it may be the case that students would benefit more from rating cassette-recorded oral "quiz" performances with programmed errors on specific schedules of errors. In such a program, it would be possible to examine the effects of schedules of errors on the learning of material rated by the student. The student would not suffer from the lesser role of being rated nor would the student suffer from rating written quiz performances.

References

Austin, S. M., and Gilbert, K. E. Student performance in a Keller-plan course in introductory electiricity and magnetism. In J. G. Sherman (Ed.) Personalized system of instruction: 41 germinal papers. Menlo Park, Calif.: W. A. Benjamin, 1974, 82-87.

Bitgood, S. C., and Kuch, D. O. A contingency managed introductory psychology course with a graduated point system. Unpublished manuscript, McMaster University, 1973.

Born, D. G. Proctor manual: initial draft. Salt Lake City: Center to Improve Learning and Instruction, University of Utah, 1970. (a)

Born, D. G. Instructor manual for development of a personalized instruction course. Salt Lake City: Center to Improve Learning and Instruction, University of Utah, 1970. (b)

Born, D. G. Student withdrawals in personalized instruction courses and in lecture courses. Paper presented at the meeting of the Rocky Mountain Psychological Association, Denver, Colorado, May 1971.

Born, D. G., and Herbert, E. W. A further study of Keller's personalized system of instruction. Journal of Experimental Education, 1971, 40, 6-11.

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

Born, D. G., Gledhill, S. M., and Davis, M. L. Examination performance in lecture-discussion and personalized instruction courses. Journal of Applied Behavior Analysis, 1972, 5, 33-43.

Calhoun, J. F. Measurement of the effectiveness of an undergraduate course taught by the "Keller" method. Unpublished manuscript, State University of New York, Stony Brook, New York, 1974.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally, 1971.

Cheney, C. D., and Powers, R. B. A programmed approach to teaching in the social sciences. Improving College and University Teaching, 1971, 19, 164-166.

Coleman, J. C. Psychology and effective behavior. Glenview, Illinois: Scott, Foresman and Company, 1969.

Cooper, J. L. Learning theory and effective instruction. Journal of Higher Education, 1973, 1, 217-234.

Cooper, J. L., and Greiner, J. M. Contingency management in an introductory psychology course produces better retention. The Psychological Record, 1971, 21, 391-400.

Corey, J. R., and McMichael, J. S. Using personalized instruction in college courses. New York: Appleton-Century-Crofts, 1970.

Corey, J. R., and McMichael, J. S. Retention in a PSI introductory psychology course. In J. G. Sherman (Ed.) Personalized system of instruction: 41 germinal papers. Menlo Park, Calif.: W. A. Benjamin, 1974, 17-19.

Corey, J. R., McMichael, J. S., and Tremont, P. J. Long term effects of personalized instruction in an introductory psychology course. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, New Jersey, April 1970.

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

Edwards, K. A. Self-managed studying in college courses. Unpublished manuscript, Minot State College, Minot, North Dakota, 1974.

Edwards, K. A. Attendance at lectures and films in self-paced courses. Unpublished manuscript, Minot State College, Minot, North Dakota, 1975.

Edwards, K. A., and Gottula, P. Proctoring oral and written performance in a motivation course. Proceedings, 81st Annual Convention, APA, 1973, 8, 911-912. (Summary) (a)

Edwards, K. A., and Gottula, P. Retention of material taught by oral and written quizzes. Unpublished data, Utah State University, 1973. (b)

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

Ferguson, G. A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1971.

Ferster, C. B. Individualized instruction in a large introductory psychology college course. The Psychological Record, 1968, 18, 521-532.

Ferster, C. B., and Perrott, M. C. Behavior principles. New York: Appleton-Century-Crofts, 1968.

Flammer, G. H. Learning as the constant and time as the variable. Engineering Education, 1971, 61, 511-514.

Fox, L. Effecting the use of efficient study habits. Journal of Mathetics, 1962, 1, 75-86.

Gage, N. L. Can science contribute to the art of teaching? Phi Delta Kappan, 1968, 49, 399-403. (a)

Gage, N. L. An analytical approach to research on instructional methods. Phi Delta Kappan, 1968, 49, 601-606. (b)

Gallup, H. F. The introductory psychology course at Lafayette College, Easton, Pennsylvania: a descriptive and tentative evaluation. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1969.

Gaynor, J. F., and Wolking, W. D. The effectiveness of currently enrolled student proctors in an undergraduate special education course. Journal of Applied Behavior Analysis, 1974, 7, 263-269.

Green, B. A. Workshop on the personalized system of instruction ("PSI by PSI"). Washington, D.C.: The Center for Personalized Instruction, 1974 (not for sale).

Hager, P. E. Nineteenth century experiments with monitorial teaching. Phi Delta Kappan, 1959, 40, 164-167.

Hergenhahn, B. R. Program learning vs. the lecture: a compromise. Teaching of Psychology Newsletter, March, 1972, 5-6.

Hess, J. H. Keller plan instruction: implementation problems. In J. G. Sherman (Ed.) Personalized instruction: 41 germinal papers. Menlo Park, Calif.: W. A. Benjamin, 1974, 136-147.

Hilgard, E. R., and Bower, G. H. Theories of learning. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Holland, J. G. Human vigilance. In R. Ulrich, T. Stachnik, and J. Mabry (Eds.) Control of human behavior. Glenview, Illinois: Scott, Foresman and Company, 1966, 53-64.

Hollander, M., and Wolfe, D. A. Nonparametric statistical methods. New York: John Wiley & Sons, 1973.

Hursh, D. E., Wildgen, J., Minkin, B., Minkin, N., Sherman, J. A., and Wolf, M. M. Proctors' discussions of students' quiz performance with students. In J. M. Johnston (Ed.) Behavior research and technology in higher education. Springfield, Illinois: Charles C. Thomas, 1975, 375-384.

Johnston, J. M. An examination of critical variables in a behavioral approach to college teaching. Dissertation Abstracts International, 1971, 32, 541-B.

Johnston, J. M. (Ed.) Behavior research and technology in higher education. Springfield, Illinois: Charles C. Thomas, 1975.

Johnston, J. M., and Pennypacker, H. S. A behavioral approach to college teaching. American Psychologist, 1971, 26, 219-244

Kagan, J., and Havemann, E. Psychology: an introduction. New York: Harcourt, Brace, Jovanovich, 1972.

Keller, F. S. A personal course in psychology. Paper presented at the meeting of the American Psychological Association, Philadelphia, August 1963. Reprinted in R. Ulrich, T. Stachnik, and J. Mabry (Eds.) Control of human behavior. Glenview, Ill.: Scott, Foresman and Co., 1966, 91-93.

Keller, F. S. Neglected rewards in the educational process. Paper presented at the meeting of the American Conference of Academic Deans, Los Angeles, January 1967. Reprinted by Independent Learning Systems, Inc.

Keller, F. S. "Good-bye, teacher ..." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Keller, F. S. A programmed system of instruction. In W. C. Becker (Ed.) An empirical basis for change in education: selections on behavioral psychology for teachers. Chicago: Science Research Associates, 1971, 506-522.

Keller, F. S., and Sherman, J. G. The Keller plan handbook. Menlo Park, Calif.: W. A. Benjamin, 1974.

Keselman, H. J., and Leventhal, L. Concerning the statistical procedures enumerated by Gentile et al.: another perspective. Journal of Applied Behavior Analysis, 1974, 7, 643-645.

Kulik, J. A., Kulik, C., and Carmichael, K. The Keller plan in science teaching. Science, 1974, 183, 379-383.

Lee, W. Experimental analysis and design. San Francisco: W. H. Freeman, 1975.

Lloyd, K. E. Contingency management in university courses. Educational Technology, 1971, 11, 18-23.

Lloyd, K. E., Garlington, W. K., Lowry, D., Burgess, H., Euler, H. A., and Knowlton, W. R. A note on some reinforcing properties of university lectures. Journal of Applied Behavior Analysis, 1972, 5, 151-155.

Lloyd, K. E., and Knutzen, N. J. A self-paced programmed undergraduate course in the experimental analysis of behavior. Journal of Applied Behavior Analysis, 1969, 2, 125-133.

Malott, R. W., and Svinicki, J. G. Contingency management in an introductory psychology course for one thousand students. The Psychological Record, 1969, 19, 545-556.

Maxwell, S., and Cramer, E. M. A note on analysis of covariance. Psychological Bulletin, 1975, 82, 187-190.

McKeachie, W. J. Research on teaching at the college and university level. In N. L. Gage (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963, 1118-1172.

McKeachie, W. J. Instructional psychology. Annual Review of Psychology, 1974, 25, 161-193.

McMichael, J. S., and Corey, J. R. Contingency management in an introductory psychology course produces better learning. Journal of Applied Behavior Analysis, 1969, 2, 79-83.

Meyer, A. E. An educational history of the American people. New York: McGraw-Hill, 1967.

Meyer, A. E. An educational history of the western world. New York: McGraw-Hill, 1972.

Michael, J. A behavioral analysis of behavioral college teaching systems. Paper presented at the meeting of the American Psychological Association, Washington, D.C., September 1971.

Morris, C. J., and Kimbrell, G. M. Performance and attitudinal effects of the Keller method in an introductory psychology course. The Psychological Record, 1972, 22, 523-530.

Postlethwait, S. N., Novak, J., and Murray, H. T. The audio-tutorial approach to learning: through independent study. Minneapolis: Burgess, 1969.

Powers, R. B., and Edwards, K. A. Personalized introductory psychology at Utah State University: a progress report, 1970-1971. Proceedings of the Utah Academy of Sciences, Arts, and Letters, 1971, 48, 33-38.

Powers, R. B., and Edwards, K. A. Performance in a self-paced course. The Journal of Experimental Education, 1974, 42, 60-64.

Powers, R. B., Edwards, K. A., and Hoehle, W. F. Bonus points in a self-paced course facilitates exam-taking. The Psychological Record, 1973, 23, 533-538.

Powers, R. B., and Wald, B. Some effects of a graduated bonus system on a PSI course. Paper presented at the National Conference of Personalized Instruction in Higher Education, Washington, D.C., April 1974.

Robinson, F. P. Effective study (4th ed.). New York: Harper & Row, 1970.

Ryan, B. A. Keller's personalized system of instruction: an appraisal. Washington, D.C.: American Psychological Association, 1974.

Sheppard, W. C., and MacDermot, H. G. Design and evaluation of a programmed course in introductory psychology. Journal of Applied Behavior Analysis, 1970, 3, 5-11.

Sherman, J. G. (Ed.) Personalized system of instruction: 41 germinal papers. Menlo Park, Calif.: W. A. Benjamin, 1974.

Shine, L. C., and Bower, S. M. A one-way analysis of variance for single-subject designs. Educational and Psychological Measurement, 1971, 31, 105-113.

Siegel, S. Non-parametric statistics. New York: McGraw-Hill, 1956.

Skinner, B. F. The behavior of organisms: an experimental analysis. New York: Appleton-Century Crofts, 1938.

Skinner, B. F. Teaching machines. Science, 1958, 28, 969-977.

Spitzer, H. F. Studies in retention. Journal of Educational Psychology, 1939, 30, 641-656.

Stachnik, T. J., and Ulrich, R. E. A laboratory section for educational psychology. The Psychological Record, 1969, 19, 129-132.

Sutterer, J. R., and Holloway, R. E. An analysis of student behavior with and without limiting contingencies. In J. M. Johnston (Ed.) Behavior research and technology in higher education. Springfield, Illinois: William C. Thomas, 1975, 303-329.

Washburn, C. W., and Marland, S. P. Winnetka: the history and significance of an educational experiment. Monterey, Calif.: Brooks/Cole, 1972.

Watson, D. L., and Tharp, R. G. Self-directed behavior: self-modification for personal adjustment. Belmont, Calif.: Wadsworth, 1972.

Whitehurst, G. J. Academic responses and attitudes engendered by a programmed course in child development. Journal of Applied Behavior Analysis, 1972, 5, 283-291.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Witters, D. R., and Kent, G. W. An experimental evaluation of programming student study behavior in undergraduate courses. (HEW 9-C-040) Bridgewater, Virginia: Bridgewater College, 1970.

Witters, D. R., and Kent, G. W. Teaching without lecturing: evidence in the case for individualized instruction. The Psychological Record, 1972, 22, 169-175.

APPENDIXES

Appendix A

Experiment I: Raw Test Scores

Table 6

Raw Scores on Tests During the Term

Student	Course and Section	Pre-test	Mid-term	Final
AK	Introductory (101A)	20	40	29.5
BL		18	34.5	28
CP		13	38	23
DK		15	41.5	23
DR		18	40	28
DG		16	42.5	26
DJ		17	41.5	28
FR		14	31	18.5
GJ		11	33.5	22
HJ		15	34.5	16.5
HN		20	37	28.5
HJa		16	37	33.5
HT		19	39	28.5
JD		12	33	18.5
JC		12	36	24
JG		17	36	28
KB		15	35.5	27
KJ		17	38	26
KD		18	35.5	20
MB		18	40	28
MM		18	42.5	31
MA		16	36	20.5
MC		18	44.5	32
PP		16	36	25
PH		19	34	26
SC		16	39	27
SB		15	41.5	33
SN		15	46	32
SM		16	43	27
SG		17	34.5	20.5
SS		16	40	29.5
SJ		11	34	32.5
TS		18	44	31.5

Table 6 (Continued)

Student	Course and Section	Pre-test	Mid-term	Final
AJ	Introductory (101B)	11	33	23
BD		17	37	26.5
BDe		16	41	26.5
BJ		16	35.5	21
BK		12	33	23
BM		16	34.5	22.5
BS		16	34	23.5
CW		18	44.5	32.5
DS		18	37.5	28
DSh		13	40	21.5
EE		16	40	26.5
GR		16	38	21
HS		17	38	21.5
HJ		17	37.5	27
KJ		10	27	15
KM		14	44	28.5
LK		15	36	16.5
MK		17	40	24.5
ML		19	41.5	31
MN		13	39	21.5
MJ		18	41.5	27
ME		12	37.5	0
NB		19	41.5	33.5
OA		18	41	35
OM		18	42.5	29
PP		15	37	26.5
RT		18	43	35
RM		19	45.5	29.5
RK		16	44.5	33
SC		19	39	26.5
SR		15	37.5	23
SL		12	30	17.5
VS		19	32.5	24.5
WJ		18	31.5	29.5
WC		17	42.5	20.5
WM		18	41	26.5
WT		17	40	23
WM		15	37.5	23.5
AV	Adjustment (201A)	18	39.5	42.5
AR		20	39.5	33.5
BK		20	39.5	34.5
BJ		16	37	29.5
CM		19	42.5	38.5

Table 6 (Continued)

Student	Course and Section	Pre-test	Mid-term	Final
DA	Adjustment (201A)	16	31	27
DT		17	44	36.5
DAu		17	32.5	34
EF		18	44	41.5
EE		18	34	39.5
EM		18	25	25
FK		14	44	39.5
GD		18	45.5	42.5
GR		19	44	37
GT		15	35.5	28.5
GTo		17	34	30.5
JT		19	44	40.5
KS		18	32.5	34
LL		17	29.5	28
MM		17	44	40.5
MC		18	39.5	34
NL		17	47	36
PM		20	44	37.5
PMi		20	39.5	38
SV		17	44	38
SD		20	31	36.5
TR		20	34	28.5
VD		20	41	42
WJ		18	32.5	35
WD		16	28	27.5
WC		17	37	34.5
WK		18	38.5	33.5
AJ	Adjustment (201B)	19	39.5	35
BJ		19	42.5	29
DK		16	38.5	38
FJ		19	42.5	31.5
GB		20	38.5	35
GS		18	47	44
HK		16	38.5	32.5
HB		16	41	37
HE		20	44	39
HBa		20	44	34.5
JW		20	41	31
KV		20	44	38
KN		18	44	38.5
KD		19	44	38
LD		20	39.5	38.5
MC		20	39.5	36.5

Table 6 (Continued)

Student	Course and Section	Pre-test	Mid-term	Final
OJ	Adjustment (201B)	20	42.5	33.5
PK		20	44	38.5
PH		18	45.5	38.5
RJ		19	44	39.5
RB		18	45.5	37.5
SL		18	48.5	37
SK		20	32.5	34
TP		19	42.5	41.5
TR		17	41	27
TP		20	42.5	41
WS		20	42.5	37
WC		17	44	38.5

Table 7

Percent Test Scores, Quiz Conditions, and Percent Differences for Individual Students in Two Courses

	Student	Pre-test (%)	Test 1 (%)	Test 2 dif.	Test 2 (%)	Test 2 dif.
Introductory Psychology Section 101A (W/O)	1	100	87	-13	84	-16
	2	90	75	-15	80	-10
	3	65	83	+18	66	+ 1
	4	75	90	+15	66	- 9
	5	90	87	- 3	80	-10
	6	80	92	+12	74	- 6
	7	85	90	+ 5	80	- 5
	8	70	67	- 3	53	-17
	9	55	73	+18	63	+ 8
	10	75	75	0	47	-28
	11	100	80	-20	81	-19
	12	80	80	0	96	+16
	13	95	85	-10	81	-14
	14	60	72	+12	53	- 7
	15	60	78	+18	69	+ 9
	16	85	78	- 7	80	- 5
	17	75	77	+ 2	77	+ 2
	18	85	83	- 2	74	- 9
	19	90	77	-13	57	-33
	20	90	87	- 3	80	-10
	21	90	92	+ 2	89	- 1
	22	80	78	- 2	59	-21
	23	90	97	+ 7	91	+ 1
	24	80	78	- 2	71	- 9
	25	95	74	-21	74	-21
	26	80	85	+ 5	77	- 3
	27	75	90	+15	94	+19
	28	75	100	+25	91	+16
	29	80	93	+13	77	- 3
	30	85	75	-10	59	-26
	31	80	87	+ 7	84	+ 4
	32	55	74	+19	93	+38
	33	80	96	+16	81	+ 1
		2650	2735	+85	2481	-167
	$\bar{X}$ =	80.30	82.88	+2.58	75.18	-5.06

Table 7 (Continued)

	Student	Pre-test (%)	Test 1 (%)	Test 1 dif.	Test 2 (%)	Test 2 dif.
Introductory Psychology Section 101B (O/W)	1	55	72	+17	66	+11
	2	85	80	- 5	76	- 9
	3	80	89	+ 9	76	- 4
	4	80	77	- 3	60	-20
	5	60	72	+12	66	+ 6
	6	80	75	- 5	64	-16
	7	80	74	- 6	67	-13
	8	90	97	+ 7	93	+ 3
	9	90	82	- 8	80	-10
	10	65	87	+12	61	- 4
	11	80	87	+ 7	76	- 4
	12	80	83	+ 3	60	-20
	13	85	83	- 2	61	-24
	14	85	82	- 3	77	- 8
	15	50	59	+ 9	43	- 7
	16	70	96	+26	81	+11
	17	75	78	+ 3	47	-28
	18	85	87	+ 2	70	-15
	19	95	90	- 5	89	- 6
	20	65	85	+20	61	- 4
	21	90	90	0	77	-13
	22	60	82	+22	0	-60
	23	95	90	- 5	96	+ 1
	24	90	89	- 1	100	+10
	25	90	92	+ 2	83	- 7
	26	75	80	+ 5	76	+ 1
	27	90	93	+ 3	100	+10
	28	95	99	+ 4	84	-11
	29	80	97	+17	94	+14
	30	95	85	-10	76	-19
	31	75	82	+ 7	66	- 9
	32	60	65	+ 5	50	-10
	33	95	71	-24	70	-25
	34	90	68	-22	84	- 6
	35	85	92	+ 7	59	-26
	36	90	89	- 1	76	-14
	37	85	87	+ 2	66	-19
	38	75	82	+ 7	67	- 8
		3050	3168	+108	2698	-352
	$\overline{X}$ =	80.26	83.37	+2.84	71.00	-9.26

Table 7 (Continued)

	Student	Pre-test (%)	Test 1 (%)	Test 1 dif.	Test 2 (%)	Test 2 dif.
Psychology of Adjustment: Section 201A (W/O)	1	90	81	- 9	97	+ 7
	2	100	81	-19	76	-24
	3	100	81	-19	78	-22
	4	80	76	- 4	67	-13
	5	95	88	- 7	88	- 7
	6	80	64	-16	61	-19
	7	85	91	+ 6	83	- 2
	8	85	67	-18	77	- 8
	9	90	91	+ 1	94	+ 4
	10	90	70	-20	90	0
	11	90	52	-48	57	-33
	12	70	91	+21	90	+20
	13	90	94	+ 4	97	+ 7
	14	95	91	- 4	81	-11
	15	75	73	- 2	65	-10
	16	85	70	-15	69	-16
	17	95	91	- 4	92	- 3
	18	90	67	-23	77	-13
	19	85	61	-24	64	-21
	20	85	91	+ 6	92	+ 7
	21	90	81	- 9	77	-13
	22	85	97	+12	82	- 3
	23	100	91	- 9	85	-15
	24	100	81	-19	86	-14
	25	85	91	+ 6	86	+ 1
	26	100	64	-36	83	-17
	27	100	70	-30	65	-35
	28	100	85	-15	95	- 5
	29	90	67	-23	80	-10
	30	80	58	-22	63	-17
	31	85	76	- 9	78	- 7
	32	90	79	-11	76	-14
		2860	2511	-359	2551	-306
	$\bar{X}$ =	89.40	78.47	-11.22	79.72	-9.56

Table 7 (Continued)

	Student	Pre-test (%)	Test 1 (%)	Test 1 dif.	Test 2 (%)	Test 2 dif.
Psychology of Adjustment: Section 201B (O/W)	1	95	81	-14	80	-15
	2	95	88	-13	66	-29
	3	80	79	- 1	86	+ 6
	4	95	88	- 7	72	-23
	5	100	79	-21	80	-20
	6	90	97	+ 7	100	+10
	7	80	79	- 1	74	- 6
	8	80	85	+ 5	84	+ 4
	9	100	91	- 9	89	-11
	10	100	91	- 9	78	-22
	11	100	85	-15	70	-30
	12	100	91	- 9	86	-14
	13	90	91	+ 1	88	- 2
	14	95	91	- 4	86	- 9
	15	100	81	-19	88	-12
	16	100	81	-19	83	-17
	17	100	88	-12	76	-24
	18	100	91	- 9	88	-12
	19	90	94	+ 4	88	- 2
	20	95	91	- 4	90	- 5
	21	90	94	+ 4	85	- 5
	22	90	100	+10	84	- 6
	23	100	67	-33	77	-23
	24	95	88	- 7	94	- 1
	25	85	91	+ 6	61	-24
	26	100	88	-12	93	- 7
	27	100	88	-12	84	-16
	28	85	91	+ 6	88	+ 3
		2630	2449	-187	2318	-312
	$\bar{X}$ =	93.93	87.46	-6.68	82.79	-11.14

Appendix B

Experiment I: Tabled Analysis of Covariance

Table 8

Analysis of Covariance: Introductory Psychology Course (Test 1)

Sample	f	Corrected Values Σx^2	Σxy	Σy^2	Regression Coefficient	f'	Deviation from Regression Σd^2	Mean Square
101A	28	17375.79	4018.97	2643.28	+0.1521	27	3616.86	133.96
101B	35	22859.00	4818.75	5347.50	+0.2339	34	3567.79	104.93
Within						61	7184.65	117.78
Regression Coefficient						1	66.07	66.07
Common	63	40234.79	8837.72	7990.78	+0.1986	62	7250.72	116.95
Adjusted Means						1	16.13	16.13
Total	64	41321.14	8843.85	8072.39		63	7266.85	

$F < 1$; 0.1379

	N	Σx	Σx^2	Σy	Σy^2	Σxy
101A	29	1616	107426	2350	194450	133595
101B	36	1710	104084	2895	237625	142860
Total	65	3326	211510	5245	432075	276455

Table 9

Analysis of Covariance: Introductory Psychology Course (Test 2)

Sample	f	Corrected Values Σx^2	Σxy	Σy^2	Regression Coefficient	Deviation from Regression f'	Σd^2	Mean Square
101A	28	17375.79	1748.55	1571.86	+0.1006	27	1395.90	51.70
101B	35	22859.00	4434.50	3034.75	+0.1940	34	2174.49	63.96
Within						61	3570.39	58.53
Regression Coefficient						1	86.05	86.05
Common	63	40234.79	6183.05	4606.61	+0.1537	62	3656.43	58.97
Adjusted Means						1	25.55	25.55
Total	64	41321.14	6181.16	4606.62		63	3681.99	

$F < 1$; 0.4333

	N	Σx	Σx^2	Σy	Σy^2	Σxy
101A	29	1616	107426	2177	164997	123060
101B	36	1710	104084	2703	205985	132827
	65	3326	211510	4880	370982	255887

Table 10

Analysis of Covariance: Introductory Psychology Course (Test 3)

Sample	f	Corrected Values Σx^2	Σxy	Σy^2	Regression Coefficient	Deviation from Regression f'	Σd^2	Mean Square
101A	28	17375.79	3539.69	1927.03	+0.2037	27	1205.95	44.66
101B	35	22859.00	4754.50	3446.97	+0.2080	34	2458.07	72.30
Within						61	3664.02	60.07
Regression Coefficient						1	0.18	0.18
Common	63	40234.79	8294.19	5374.00	+0.2061	62	3664.20	59.10
Adjusted Means						1	0.11	0.11
Total	64	41321.14	8507.39	5415.85		63	3664.31	

$F < 1$; 0.0019

	N	Σx	Σx^2	Σy	Σy^2	Σxy
101A	29	1,616	107,426	1,525	82,121	88,519
101B	36	1,710	104,084	1,835	96,981	91,917
Total	65	3,326	211,510	3,360	179,102	180,436

Table 11

Analysis of Covariance: Psychology of Adjustment Course (Test 1)

Sample	f	Corrected Values Σx^2	Σxy	Σy^2	Regression Coefficient	Deviation from Regression f'	Σd^2	Mean Square
201A	21	10629.82	137.27	1496.59	+0.0129	20	1494.82	74.74
201B	25	18142.62	-16.92	1215.38	-0.0009	24	1215.36	50.64
Within						44	2710.18	61.60
Regression Coefficient						1	1.29	1.29
Common	46	28772.44	120.35	2711.97	+0.0042	45	2711.47	60.25
Adjusted Means						1	287.23	287.23
Total	47	29303.92	516.88	3007.82		46	2998.70	

F = 4.7669; .01 < p < .05

	N	Σx	Σx^2	Σy	Σy^2	Σxy
201A	22	1,212	77,400	1,955	175,225	107,840
201B	26	1,606	117,344	2,440	230,200	150,700
	48	2,818	194,744	4,395	405,425	258,540

Table 12

Analysis of Covariance: Psychology of Adjustment Course (Test 2)

Sample	f	Corrected Values Σx^2	Corrected Values Σxy	Corrected Values Σy^2	Regression Coefficient	Deviation from Regression f'	Deviation from Regression Σd^2	Deviation from Regression Mean Square
201A	21	10629.82	3963.82	2903.32	+0.3729	20	1425.23	71.26
201B	25	18142.62	1963.69	994.65	+0.1082	24	782.11	32.59
Within						44	2207.33	50.17
Regression Coefficient						1	469.49	469.49
Common	46	28772.44	5927.51	3897.97	+0.2060	45	2676.82	59.48
Adjusted Means						1	442.35	442.35
Total	47	29303.92	6526.34	4572.67		46	3119.17	

$F = 7.4363$; $p < .01$

	N	Σx	Σx^2	Σy	Σy^2	Σxy
201A	22	1,212	77,400	1,685	131,959	96,792
201B	26	1,606	117,344	2,187	184,955	137,053
Total	48	2,818	194,744	3,872	316,914	233,845

Table 13

Analysis of Covariance: Psychology of Adjustment Course (Test 3)

Sample	f	Corrected Values Σx^2	Σxy	Σy^2	Regression Coefficient	f'	Deviation from Regression Σd^2	Mean Square
201A	21	10629.82	2766.36	2202.77	+0.2602	20	1482.84	74.14
201B	25	18142.62	1854.31	1208.65	+0.1022	24	1019.13	42.46
Within						44	2501.96	56.86
Regression Coefficient						1	167.41	167.41
Common	46	28772.44	4620.67	3411.42	+0.1606	45	2669.37	59.32
Adjusted Means						1	26.12	26.12
Total	47	29303.92	4824.92	3489.92		46	2695.49	

$F < 1$; 0.4403

	N	Σx	Σx^2	Σy	Σy^2	Σxy
201A	22	1,212	77,400	1,525	107,913	86,780
201B	26	1,606	117,344	1,869	135,561	117,301
	48	2,818	194,744	3,394	243,474	204,081

Appendix C

Experiment I: Student Evaluation Form

Speaker: _____ Date: _____

Listener: _____ Unit Name: _____

Time Started: _____ Time Ended: _____ No. prior interviews: _____

* * * * * * * * * * * * * * * *

Check list before interview:	YES	NO
Is the speaker prepared? (If no, stop)		
Is the listener prepared? (If no, stop)		
Has the speaker his notes ready to consult?		
Has the listener his notes ready to consult?		
Is there enough time left for the interview? (If not, stop)		

Check list during interview:	XXX	XXX
Is the information logically sequenced?		
Are critical terms stated correctly?		
Are critical terms defined when used the first time?		
Can someone who has not studied this material understand it?		

Notes during interview:

Commission Errors (mistakes in reporting names, dates, figures, etc.)-

Omission Errors (important information not mentioned) -

(Note: Give the speaker an opportunity to clarify the errors after he has stated that he is finished.)

Rank of the speaker's performance:

90 - 100% _____
80 - 89% _____
70 - 79% _____
less than 70% _____

Recommendation: Pass (90% or better): _____ Re-interview: _____

Appendix D

Experiment I: Course Evaluation Form

COURSE EVALUATION

Course: _____ Date: _____

Section: _____ SS#: _____

____ 1. I am a (a) psychology major, (b) psychology minor, (c) not in psychology.

____ 2. My overall grade point is (a) below 2.0, (b) between 2.0 and 2.5, (c) between 2.5 and 3.0, (d) between 3.0 and 3.5, (e) above 3.5.

____ 3. I plan to receive in this class, a(n) (a) A, (b) B, (c) C, (d) D, (e) F.

____ 4. For this class, I study on the average (a) less than 1 hour each week, (b) 1-2 hours each week, (c) 2-3 hours each week, (d) 3-4 hours each week, (e) 4-5 hours each week, (f) 5-6 hours each week, (g) more than 6 hours each week.

____ 5. Compared to courses with similar material, I spent (a) more time studying, (b) about the same amount of time studying, (c) less time studying.

____ 6. I feel that I learned _____ in this class than in other classes. (a) more, (b) about the same, (c) less.

____ 7. In courses with two or three tests each quarter, I usually have (a) studied every week for about the same amount of time, (b) studied a little every week but studied more in the week before a test, (c) studied nearly all the material to be covered on the test within a few days before the test.

____ 8. I usually read (a) all of the assigned material, (b) only the parts specified on the study guides, (c) most of the material, (d) a little more than the parts specified on the study guides.

____ 9. As far as the study guides are concerned, I usually (a) read the material before trying to answer the study guide questions, (b) tried to answer the study guide questions while reading the material, (c) did not answer the study guide questions until after the first weekly test and before the make-up test, (d) copied the answers from a friend.

____10. The course was _____ interesting than any other class taken this term. (a) more, (b) equally as, (c) less.

____11. I would be willing to _____ this class to my friends. (a) recommend, (b) not recommend, (c) condemn, (d) say nothing about.

____12. The course was _____ informative than any other classes taken this term. (a) more, (b) equally as, (c) less.

____13. The course could be improved by making the following changes:

____14. I would rate the oral interviewing procedure: (a) highly, (b) so-so, (c) low

____15. I would rate the written exam procedure: (a) highly, (b) so-so, (c) low

____16. I preferred the _____ method. (a) oral, (b) written

____17. In terms of new acquaintances, I believe I made _____ in other classes I have taken this quarter. (a) more than, (b) the same as, (c) less than

____18. I would estimate the number of people I know in this class well enough to talk to outside of this class at about: (a) none, (b) 1, (c) 2, (d) 3, (e) 4, (f) 5, (g) 6 or more

____19. I would estimate the number of people I know in other classes I am taking this term at an average of about: (a) none, (b) 1, (c) 2, (d) 3, (e) 4, (f) 5, (g) 6 or more

Appendix E

Experiment I: Tabled Survey Data

Table 14

Majors, Minors, and Students Not in Psychology

Exam	Class	Course: Introductory A	Course: Introductory B	Course: Adjustment A	Course: Adjustment B
Mid-term	Major	0	3	6	4
	Minor	0	0	5	9
	Not in Psych	25	30	14	11
Final	Major	1	3	7	5
	Minor	0	1	4	8
	Not in Psych	24	29	14	11

Table 15

Sex of Students Enrolled

Sex	Course: Introductory A	Course: Introductory B	Course: Adjustment A	Course: Adjustment B
Males	11	15	11	8
Females	22	23	21	20

Table 16

Attrition from Each Class at the Term's End

Sex	Course							
	Introductory				Adjustment			
	A		B		A		B	
	Start	Drop	Start	Drop	Start	Drop	Start	Drop
Males	15	4	17	3	14	4	11	3
Females	23	0	24	0	22	2	23	3
Total	38	4	41	3	36	6	34	6

Table 17

Ratings of Oral and Written Procedures and Resultant Preferences Collected at Test 3

Rating	Course			
	Introductory		Adjustment	
	A (W/O)[a]	B (O/W)[b]	A (W/O)	B (O/W)
(Oral)				
High	12	16	15	5
Middle	10	14	5	18
Low	3	3	2	1
(Written)				
High	9	8	8	5
Middle	13	19	15	15
Low	3	6	0	4
(Preference)				
Oral	16	27	16	13
Written	9	5	7	11

[a]Written during first half, oral during second half.

[b]Oral during first half, written during second half.

Table 18

Number of Students and Hours Spent Studying
In Each Course by Section

	Course			
	Introductory		Adjustment	
Hours	A	B	A	B
-1	0	2	1	0
1-2	2	3	1	4
2-3	4	9	8	4
3-4	8	10	7	4
4-5	3	3	3	5
5-6	2	3	2	5
6+	5	2	0	0

Table 19

Number of Students Reporting a Particular Method of Reading Assigned Material

	Course			
	Introductory		Adjustment	
Reading Method	A	B	A	B
Read all material	9	7	6	4
Read study guide material only	2	10	2	4
Read most of the material	9	5	9	8
Read study guide material plus extra	5	10	7	8

Table 20

Number of Students Within a Self-Estimated Grade-Point-Average Interval

	Course			
	Introductory		Adjustment	
Grade Point Average	A	B	A	B
0-2.0	0	1	4	0
2.0-2.5	6	9	6	4
2.5-3.0	8	10	8	9
3.0-3.5	8	8	4	7
3.5-4.0	1	4	2	4

Table 21

Number of Students Expecting a Particular Grade From the Course

	Course			
	Introductory		Adjustment	
Grade Expected	A	B	A	B
A	14	18	18	20
B	9	13	4	2
C	2	1	2	0
D	0	0	0	0
F	0	0	0	0

Table 22

Final Grade Distributions for Introductory Psychology and Psychology of Adjustment Courses

Final Grade	Course	
	Introductory	Adjustment
A	38	40
B	22	9
C	5	6
D	5	5
F	0	0

Table 23

Final Grade Distributions for the Oral/Written and Written/Oral Sections

Final Grade	Class	
	Oral/Written	Written/Oral
A	45	33
B	13	18
C	4	7
D	3	7
F	2	0

Appendix F

Experiment II: Means of Pre-Course Test, Post-Course Test, and Difference Scores

Table 24

Raw Score Means for Pre-Course, Post-Course, and Difference Scores From Pre-Course to Post-Course Tests[a]

	Pre-Course					Post-Course					Difference				
S	O	W	OR	WR	NQ	O	W	OR	WR	NQ	O	W	OR	WR	NQ
AA[b]	3.33	3.33	2.50	3.00	3.75	8.33	6.00	4.50	5.50	5.50	5.00	2.67	2.00	2.50	1.75
BA	6.50	5.00	5.33	6.75	5.25	6.00	6.00	7.00	7.75	7.00	-0.50	1.00	1.67	1.00	1.75
CM	2.00	3.00	4.67	4.00	2.33	3.50	3.50	4.67	3.67	4.00	1.50	0.50	0	-0.33	1.67
CC	3.25	5.00	5.00	3.00	4.50	6.75	6.33	2.00	4.00	4.75	3.50	1.33	-3.00	1.00	0.25
CL	4.50	5.25	6.67	6.00	4.75	9.00	7.50	8.67	8.00	8.50	4.50	2.25	2.00	2.00	3.75
DR	3.33	3.00	5.50	3.00	3.25	5.33	5.67	6.00	4.50	5.50	2.00	2.67	0.50	1.50	2.25
FL	3.50	3.50	4.00	3.00	4.50	6.00	5.00	5.67	7.00	4.75	2.50	1.50	1.67	4.00	0.25
FR	5.33	2.00	4.00	5.00	4.33	4.67	4.00	4.00	5.75	6.00	-0.66	2.00	0	0.75	1.67
GL	4.75	4.00	7.00	6.50	5.25	6.25	4.00	7.00	5.50	5.25	1.50	0	0	-1.00	0
HC	4.00	3.50	3.67	3.00	5.00	6.50	6.50	6.00	5.67	5.00	2.50	3.00	2.33	2.67	0
HD	3.50	5.00	2.00	4.50	4.00	6.50	7.50	8.00	7.50	6.67	3.00	2.50	6.00	3.00	2.67
HT	4.50	4.33	3.67	4.50	5.00	6.00	7.00	7.33	7.50	6.75	1.50	2.67	3.66	3.00	1.75
HJi	5.00	4.50	5.00	5.00	6.33	7.00	6.50	8.00	7.67	5.00	2.00	2.00	3.00	2.67	-1.33
HJo	4.50	4.67	3.00	2.50	3.00	5.00	5.67	5.50	4.00	4.50	0.50	1.00	2.50	1.50	1.50
JR	4.00	3.00	5.00	3.75	3.75	5.50	6.00	7.00	3.50	5.25	1.50	3.00	2.00	-0.25	1.50
JC	4.50	6.00	5.00	5.50	4.67	4.50	5.50	5.00	6.50	6.33	0	-0.50	0	1.00	1.66
MK	4.00	6.50	5.50	3.33	5.00	6.00	7.00	7.50	7.33	5.75	2.00	0.50	2.00	4.00	0.75
MJ	4.00	4.00	3.33	6.00	4.25	7.00	5.67	6.00	6.00	6.00	3.00	1.67	2.67	0	1.75
MD	4.00	4.00	3.00	3.33	5.75	9.00	9.00	7.50	6.67	6.25	5.00	5.00	4.50	3.34	0.50
OT	4.25	7.00	5.00	4.00	3.67	3.50	8.00	6.00	5.00	5.67	-0.75	1.00	1.00	1.00	2.00

Table 24(continued)

	Pre-Course					Post-Course					Difference				
S	O	W	OR	WR	NQ	O	W	OR	WR	NQ	O	W	OR	WR	NQ
PL	3.67	4.67	4.50	2.50	4.50	5.00	5.00	6.00	6.00	5.25	1.33	0.33	1.50	3.50	0.75
RMa	6.00	4.25	3.33	5.00	4.25	8.00	6.25	7.00	6.00	6.50	2.00	2.00	3.67	1.00	2.25
RMe	4.33	6.00	4.50	5.67	5.25	6.00	8.50	7.00	7.33	7.25	1.67	2.50	2.50	1.66	2.00
SB	5.33	5.00	3.33	4.00	5.50	5.33	6.67	6.00	3.50	6.50	0	1.67	2.67	-0.50	1.00
SJ	4.00	4.00	3.67	3.00	4.00	8.00	8.00	7.00	6.00	6.25	4.00	4.00	3.33	3.00	2.25
SSt	4.00	3.50	4.67	6.00	3.33	3.00	3.50	4.00	2.00	4.33	-1.00	0	-0.67	-4.00	1.00
UP	6.00	6.00	5.00	4.33	4.75	4.00	9.00	5.33	6.00	5.50	-2.00	3.00	0.33	1.67	0.75
VD	4.00	5.50	5.00	5.00	5.25	9.00	6.50	9.00	6.00	7.00	5.00	1.00	4.00	1.00	1.75
WaC	2.00	2.50	4.50	3.50	4.25	4.33	4.00	2.50	4.00	4.25	2.33	1.50	-2.00	0.50	0
WP	3.00	4.50	4.33	5.33	4.75	6.00	6.00	6.33	6.33	6.75	3.00	1.50	2.00	1.00	2.00
WiC	7.00	7.00	5.00	4.50	5.50	6.00	7.50	8.50	7.50	5.75	-1.00	-0.50	3.50	3.00	0.25
BC[c]	4.50	4.00	4.00	5.00	3.25	5.75	6.00	8.00	4.00	3.75	1.25	2.00	4.00	-1.00	0.50
BL	4.33	5.50	3.00	2.00	4.67	6.33	6.00	5.50	6.00	6.00	2.00	0.50	2.50	4.00	1.33
CW	4.50	6.00	5.33	4.67	6.25	6.50	7.00	6.00	5.00	6.00	2.00	1.00	0.67	0.37	-0.25
FD	4.75	5.00	4.00	5.00	4.50	7.25	8.00	8.00	7.50	5.75	2.50	3.00	4.00	2.50	1.25
GJ	4.33	5.50	5.00	5.00	3.67	7.67	5.00	7.00	7.50	7.00	3.34	-0.50	2.00	2.50	3.33
GD	4.33	6.00	4.50	5.33	4.00	6.67	4.00	6.50	5.33	6.50	2.34	-2.00	2.00	0	2.50
JP	4.67	4.00	1.00	5.50	5.00	6.00	4.00	6.00	5.00	1.00	1.33	0	5.00	-0.50	-4.00
MJ	5.00	5.00	6.00	3.00	4.25	6.50	5.00	7.00	6.50	6.00	1.50	0	1.00	3.50	1.75
ND	3.50	3.00	3.00	4.00	3.75	6.00	3.00	3.00	3.33	3.50	2.50	0	0	-0.67	-0.25
NJ	6.50	6.50	4.00	6.00	5.50	7.50	6.50	6.00	8.00	7.25	1.00	0	2.00	2.00	1.75
PP	4.50	4.00	3.50	7.00	5.25	6.50	8.00	7.50	8.00	7.50	2.00	4.00	4.00	1.00	2.25
RC	8.00	5.00	5.00	5.50	5.00	9.00	9.00	7.75	6.00	7.25	1.00	4.00	2.75	0.50	2.25

Table 24(continued)

	Pre-Course					Post-Course					Difference				
S	O	W	OR	WR	NQ	O	W	OR	WR	NQ	O	W	OR	WR	NQ
RL	2.50	6.00	5.00	4.50	4.50	6.50	7.00	9.00	5.50	4.50	4.00	1.00	4.00	1.00	0
SSh	3.00	5.00	3.67	6.00	4.50	2.00	2.00	6.00	7.00	2.75	-1.00	-3.00	2.33	1.00	-1.75
SP	7.00	6.50	7.00	3.00	6.00	8.50	8.00	8.00	8.00	7.00	1.50	1.50	1.00	5.00	1.00
SC	2.00	5.00	4.50	4.00	3.33	7.00	8.00	8.25	7.00	8.00	5.00	3.00	3.75	3.00	4.67
TB	7.00	5.00	5.00	4.50	8.00	9.00	8.00	7.00	6.50	6.00	2.00	3.00	2.00	2.00	-2.00
TT	6.50	6.67	4.33	5.00	6.00	7.50	8.00	7.67	8.00	8.33	1.00	1.33	3.34	3.00	2.33
ZG	6.00	6.00	5.00	7.00	5.00	8.00	9.00	9.50	9.00	8.25	2.00	3.00	4.50	2.00	3.25

[a]Only data from those students with representative scores from each of the five conditions are included (N=50). Means were computed from one or more performances in a given condition.

[b]Section 201A data follow.

[c]Section 201B data follow.

Appendix G

Experiment II: Sample Quiz

Behavior-Environment Relationships:
Some Basic Principles

Quiz 3

1. Define learning, state what is excluded from the definition, and describe its relationship to inheritance. (4 parts)

2. Describe each part of a "situation" and how its parts are interrelated. (5 parts)

3. Describe respondent conditioning. (5 parts)

4. Why are operants called operants and respondents, respondents? How is each learned? (4 parts)

5. How might a behavioral consequence strengthen behavior? State the name of the principle and describe it. (4 parts)

6. Describe the type of punishment in which aversive stimuli are presented in terms of its two parts. (2 parts)

7. Describe the second type of punishment. (4 parts)

Appendix H

Experiment II: Reliability Data

Lecture # (Written)	Item Reliability Ag	Dis	Rel
1	20	11	65%
1	29	2	94%
1	24	7	77%
1	21	10	68%
1	19	12	61%
(Oral)			
3	23	5	82%
3	18	10	64%
3	18	10	64%
3	16	12	57%
(Written)			
4	13	3	81%
4	11	5	69%
4	12	4	75%
4	12	4	75%
(Oral)			
5	17	2	89%
5	14	5	74%
5	14	5	74%
5	11	8	58%
5	17	2	89%
5	15	4	79%
(Oral)			
7	16	6	73%
7	15	7	68%
7	17	5	77%
7	19	3	86%
(Written)			
8	12	8	60%
8	18	2	90%
8	11	9	55%
(Written)			
9	16	0	100%
9	11	5	69%
9	11	5	69%
9	15	1	94%

Lecture # (Oral)	Item Reliability Ag	Dis	Rel
1	18	13	58%
1	23	8	74%
(Written)			
2	15	12	56%
2	21	6	78%
(Written)			
5	11	8	57%
5	14	5	74%
5	10	9	53%
(Oral)			
6	17	4	81%
6	18	3	86%
(Oral)			
8	17	3	85%
8	14	6	70%
8	16	4	80%
8	19	1	95%
(Written)			
10	18	5	78%
10	20	3	87%
10	21	2	91%
10	19	4	83%
(Oral)			
11	20	4	83%
11	21	3	88%
11	21	3	88%
11	21	3	88%

Lecture #	Item	Reliability	
(Oral)			
11	18	6	75%
11	22	2	92%
11	17	7	71%
11	22	2	92%
11	22	2	92%
(Written)			
13	24	9	73%
13	23	10	70%
13	28	5	85%
13	28	5	85%
(Oral)			
14	17	4	81%
14	14	7	76%
14	18	3	86%
14	19	2	90%
14	14	7	67%

Lecture #	Item	Reliability	
(Written)			
12	13	4	76%
12	9	8	53%
12	13	4	76%
(Oral)			
13	26	7	79%
13	25	8	76%
13	24	9	73%
13	19	14	58%

Total Agreement = 1274

Total Disagreement = 406

1274/1680 = .7583

Reliability average = 76%

Oral 201A: Ag = 413; Dis = 126; Reliability = 77%

Written 201A: Ag = 358; Dis = 117; Reliability = 75%

Oral 201B: Ag = 319; Dis = 93; Reliability = 77%

Written 201B: Ag = 184; Dis = 70; Reliability = 72%

Disagreements

201A

Lecture #	2-pt Dis	1-pt Dis	Total Dis
(Written)			
1	1	10	
1	1	1	
1	0	7	
1	3	7	
1	2	10	42
(Oral)			
3	1	4	
3	4	6	
3	3	7	
3	2	10	37
(Written)			
4	0	3	
4	0	5	
4	0	4	
4	1	3	16
(Oral)			
5	1	1	
5	1	4	
5	1	4	
5	1	7	
5	0	2	
5	0	4	26
(Oral)			
7	4	2	
7	1	6	
7	3	2	
7	0	3	21
(Written)			
8	3	5	
8	1	1	
8	2	7	
			19
(Written)			
9	0	0	
9	2	3	
9	0	5	
9	0	1	11
(Oral)			
11	1	5	
11	2	0	
11	4	3	
11	0	2	
11	0	2	19

201B

Lecture #	2-pt Dis	1-pt Dis	Total Dis
(Oral)			
1	4	9	
1	0	8	
			17
(Written)			
2	7	5	
2	3	3	
			8
(Written)			
5	2	6	
5	1	4	
5	1	8	18
(Oral)			
6	1	3	
6	1	2	
			5
(Oral)			
8	1	2	
8	4	2	
8	1	3	
8	0	1	8
(Written)			
10	1	4	
10	1	2	
10	0	2	
10	0	4	12
(Oral)			
11	1	3	
11	3	0	
11	2	1	
11	0	3	7
(Written)			
12	0	4	
12	1	7	
12	0	4	
			15

201A				201B			
Lecture #	2-pt Dis	1-pt Dis	Total Dis	Lecture #	2-pt Dis	1-pt Dis	Total Dis
(Written)				(Oral)			
13	1	8		13	0	7	
13	3	7		13	1	7	
13	0	5		13	5	4	
13	0	5	29	13	2	12	30
(Oral)							
14	2	2					
14	3	4					
14	0	3					
14	0	2					
14	3	4	23				
Total	57	186	243		43	120	163

Total Disagreements = 406
Total 2-point Disagreements = 100
Total 1-point Disagreements = 306
Total % 2-point Disagreements = 25%
Total % 1-point Disagreements = 75%

Oral 201A: 2-point = 37; 1-point = 89
%2 point = 29%; %1-point = 71%

Written 201A: 2-point = 20; 1-point = 97
%2-point = 17%; %1-point = 83%

Oral 201B: 2-point = 26; 1-point = 67
% 2-point = 28%; %1-point = 72%

Written 201B: 2-point = 17; 1 point = 53
%2-point = 25%; % 1-point = 75%

Appendix I

Experiment II: Rating Quiz Instructions to Raters

Read before beginning

Proctoring

Each test proctor has been selected by a random procedure for this class session. Your task is to read each answer written by your assigned student as they are completed and rate the answer on the rating form which accompanies these instructions. Use the quiz answer keys to aid in your rating.

The student's answers need not be identical to the keyed answers, but in order to receive a "2" his answers must be quite close to the key. Rate the answer a "1" when most of the important terms and concepts are present but the total answer is muddled. Rate the answer "0" when it is unquestionably wrong or no answer has been attempted.

There are no more than 7 questions on each quiz and each question contains no more than 5 parts. Be sure to rate each part of each question.

On occassion, a second proctor will be asked to read the answers of your student. The student's answers should then be given to that proctor as soon as you have completed your scoring.

The student will use his or her own paper to answer each question. As soon as one question has been answered, that question should be rated; the student will not be allowed to make corrections once the answer has been completed.

Do not talk to the student until the quiz has been completed. Do not answer any questions the student might ask. Simply state: "I cannot answer any questions until the quiz has been completed." Do not allow the student to see his or her rating. If the student is interested in seeing the rating, send him or her to see the instructor.

If there are no questions, you may proceed. You will have about 20 minutes to complete the quiz. Be sure all materials are handed in before leaving. Be sure all forms are completed. Thank you.

Read before beginning

Interviewing

Each quiz interviewer has been selected by a random procedure for this class session. Your task is to listen to your assigned student answer each question orally and as each question has been completed to rate the answer on the rating form which accompanies these instructions. Use the quiz answer keys to aid in your rating.

The student's answers need not be identical to the keyed answers, but in order to receive a "2" his answers must be quite close to the key. Rate the answer a "1" when most of the important terms and concepts are present but the total answer is muddled. Rate the answer "0" when it is unquestionably wrong or no answer has been attempted.

There are no more than 7 questions on each quiz and each question contains no more than 5 parts. Be sure to rate each part of each question.

On occasion, a second interviewer will be asked to listen to the same student. Do not talk to the other interviewer under any

circumstances until the interview has been completed.

The student is to read each question aloud as soon as he or she is ready and has completed the previous question. As soon as the student begins the next question, rate his or her performance on the last question. Do not allow the student to return to any previous question once an answer has been indicated as finished.

Do not otherwise talk to the student until the quiz has been completed. Do not answer any questions the student might ask. Simply state: "I cannot answer any questions until the quiz has been completed." Do not allow the student to see his or her rating. If the student is interested in seeing the rating, send him or her to see the instructor.

If there are no questions, you may proceed. You will have about 20 minutes to complete the quiz. Be sure all materials are handed in before leaving. Be sure all forms are completed. Thank you.

Appendix J

Experiment II: Student Performance Evaluation Form

Rater's Evaluation Form

Rater: Date:

Student:

Course: Section:

This form is structured to provide you, the rater, with an objective method for evaluating the student's quiz performance. Be sure to accurately mark the student's responses on this form. Use the quiz keys which accompany this form to guide your evaluation. Each quiz key will be marked with the appropriate question number and its corresponding part. Be sure to complete all of the information at the top of the form before proceeding. Be sure all of the information on the form has been completed before handing it in. Thank you.

Mark answers which are clear 2 points, answers which are muddled 1 point, and either wrong or no answer 0 points.

Question	Part	Question	Part	Question	Part
1	A___	4	A___	7	A___
	B___		B___		B___
	C___		C___		C___
	D___		D___		D___
	E___		E___		E___
2	A___	5	A___	Total 2:	_____
	B___		B___	Total 1:	_____

Question	Part	Question	Part
2	C___	5	C___
	D___		D___
	E___		E___
3	A___	6	A___
	B___		B___
	C___		C___
	D___		D___
	E___		E___

Total 0: _____

Total Points: _____

% $(\frac{\text{Total obtained}}{\text{Total possible}})$: ___

Appendix K

Experiment II: Course Evaluation Form

COURSE EVALUATION

Course: _____ Date: _____

Section: _____ SS#: _____

____ 1. I am a (a) psychology major, (b) psychology minor, (c) not in psychology.

____ 2. My overall grade point is (a) below 2.0, (b) between 2.0 and 2.5, (c) between 2.5 and 3.0, (d) between 3.0 and 3.5, (e) above 3.5.

____ 3. I plan to receive in this class, a(n) (a) A, (b) B, (c) C, (d) D, (e) F.

____ 4. For this class, I study on the average (a) less than 1 hour each week, (b) 1-2 hours each week, (c) 2-3 hours each week, (d) 3-4 hours each week, (e) 4-5 hours each week, (f) 5-6 hours each week, (g) more than 6 hours each week.

____ 5. Compared to courses with similar material, I spent (a) more time studying, (b) about the same amount of time studying, (c) less time studying.

____ 6. I feel that I learned _____ in this class than in other classes. (a) more, (b) about the same, (c) less.

____ 7. In courses with two or three tests each quarter, I usually have (a) studied every week for about the same amount of time, (b) studied a little every week but studied more in the week before a test, (c) studied nearly all the material to be covered on the test within a few days before the test.

____ 8. I usually read (a) all of the assigned material, (b) only the parts specified on the study guides, (c) most of the material, (d) a little more than the parts specified on the study guides.

____ 9. As far as the study guides are concerned, I usually (a) read the material before trying to answer the study guide questions, (b) tried to answer the study guide questions while reading the material, (c) did not answer the study guide questions until after the first weekly test and before the make-up test, (d) copied the answers from a friend.

____10. The course was _____ interesting than any other class taken this term. (a) more, (b) equally as, (c) less.

____11. I would be willing to _____ this class to my friends. (a) recommend, (b) not recommend, (c) condemn, (d) say nothing about.

____12. The course was _____ informative than any other classes taken this term. (a) more, (b) equally as, (c) less.

____13. The course could be improved by making the following changes:

____14. I would rate the oral interviewing procedure: (a) highly, (b) so-so, (c) low.

____15. I would rate the written exam procedure: (a) highly, (b) so-so, (c) low.

____16. I preferred the _____ method. (a) oral, (b) written

____17. In terms of new acquaintances, I believe I made _____ in other classes I have taken this quarter. (a) more than, (b) the same as, (c) less than

____18. I would estimate the number of people I know in this class well enough to talk to outside of this class at about: (a) none, (b) 1, (c) 2, (d) 3, (e) 4, (f) 5, (g) 6 or more

____19. I would estimate the number of people I know in other classes I am taking this term at an average of about: (a) none, (b) 1, (c) 2, (d) 3, (e) 4, (f) 5, (g) 6 or more

20. Please rank the course materials from the most liked (1) to the least liked (5).

_____ Films & tapes
_____ Lectures
_____ Interviews
_____ Exams
_____ Projects

VITA

K. Anthony Edwards

Candidate for the Degree of

Doctor of Philosophy

Dissertation: An Analysis of Oral and Written Quizzes as Teaching Techniques

Major Field: Psychology

Biographical Information:

Personal Data: Born at Spokane, Washington, March 2, 1939.

Education: Received Associate of Arts degree from North Idaho Junior College, with a major in engineering, in 1965; received Bachelor of Arts degree from Eastern Washington State College, with a major in psychology, in 1967; received Master of Science degree from Eastern Washington State College, with a major in psychology, in 1968. Master's Thesis: Trifluoporazine (Stelazine): Effects on DRL Schedule Transition.

Professional Experience: 1975 to present, assistant professor in psychology, Minot State College, Minot, North Dakota; 1973 to 1975, instructor in psychology, Minot State College, Minot, North Dakota; 1969 to 1970, Psychologist I, State Hospital North, Orofino, Idaho; 1968 to 1969, instructor in psychology, College of St. Teresa, Winona, Minnesota.

Other College and Professional Experience: 1969 to present, book reviewer for AAAS Science Books; 1974 to present, book reviewer for Science Activities; 1975 to present, book reviewer for Library Journal; 1970 to 1973, graduate teaching assistant, Utah State University; 1970, extension instructor, University of Idaho, Moscow, Idaho; 1967 to 1968, animal colony manager, Eastern Washington State College; 1964 to 1965, sophomore class vice-president, North Idaho Junior College.

Professional Organizations: American Psychological Association, Associate, Divisions 2, 22, 25; Psi Chi National Honor Society; National Science Teachers Association; American Association of University Professors; North Dakota Psychological Association; Midwestern Psychological Association.

Committee Assignments: Library Committee, Minot State College, 1974 to present; Undergraduate committee for general psychology, Utah State University, 1971 to 1973.

Honors: Psi Chi Chapter Research Award, Eastern Washington State College, May 1968.

Books: (Ed.) Self-control: readings in "adjustment." Lexington, Mass.: Xerox, 1975.

Publications:

The effects of trifluoperazine on DRL schedule transition. Proceedings of the Utah Academy of Sciences, Arts, and Letters, 1971, 48(2), 3-9. (with C. D. Cheney and R. B. Powers)

Personalized introductory psychology at Utah State University. Proceedings of the Utah Academy of sciences, Arts, and Letters, 1971, 48(2), 33-38. (with R. B. Powers)

The student as a teacher. Research in Education, 1972, 7(12), 42. (abstract) (ERIC ED 066 032)

Testing in a personalized system of instruction. Research in Education, 1973, 8(2), 137. (abstract) (ERIC 068 536) (with E. Sides)

Self-paced instruction at USU: introductory psychology. Instructional Development, MLLRP, 1973, 7, 4. (with R. Powers)

Self-pacing in a personalized system of instruction: work patterns and course completion. Research in Education, 1973, 8(5), 78. (abstract) (ERIC ED 071 451) (with R. B. Powers)

Proctoring oral and written performance in a motivation course. Proceedings, 81st Annual Convention, APA, 1973, 8 911-912. (with P. Gottula)

Bonus points in a self-paced course facilitates exam-taking. The Psychological Record, 1973, 23, 533-538. (with R. B. Powers and W. F. Hoehle) Abstracted in Human Behavior, 1974, 3(5), 55.

A reliable and expensive method for selection of graduate students: maintaining professorial standards. The Worm Runner's Digest, 1973, 15, 69-71. (with M. B. Lafferty and D. J. Burns)

The reduction of lies: an exercise in self-management. The Worm Runner's Digest, 1973, 15(2), 56-57. (with C. J. Gravis)

Performance in a self-paced course. The Journal of Experimental Education, 1974, 42(4), 60-64. (with R. B. Powers)

Proctoring oral and written performance in a motivation course. Research in Education, 1974, 9(9), 147. (abstract) (ERIC ED 091 403) (with P. Gottula)

Physical restraint as time-out in therapy. The Psychological Record, 1974, 24, 393-397.

A personalized "psychology of adjustment" laboratory course. In J. B. Maas and D. A. Kleiber (Eds.) Directory of teaching innovations in psychology. Washington, D.C.: Division of the Teaching of Psychology, American Psychological Association, 1975, 331-332.

Interdisciplinary physiological psychology with a laboratory and personalized instruction. In J. B. Maas and D. A. Kleiber (Eds.) Directory of teaching innovations in psychology. Washington, D.C.: Division of the Teaching of Psychology, American Psychological Association, 1975, 383-384.

Extension teaching with taped lectures and personalized instruction. In J. B. Maas and D. A. Kleiber (Eds.) Directory of teaching innovations in psychology. Washington, D.C.: Division of the Teaching of Psychology, American Psychological Association, 1975, 469-470.

Stimulus control, small steps, and self-pacing in college instruction. New Directions in Teaching, in press. (with P. Gottula)

Self-management in a beginning psychology course as a laboratory exercise. The New Scholar, in press. (with R. B. Powers)

Book Reviews:

G. A. Theodorson and A. G. Theodorson A modern dictionary of sociology. New York: Crowell, 1969. Science Books: A Quarterly Review, 1970, 6(1), 4.

H. Borland, et al. (Eds.) The crisis of survival. New York: William Morrow, 1970. Science Books: A Quarterly Review, 1970, 6(4), 310.

I. L. Horowitz (Ed.) The use and abuse of social science. New Brunswick, N.J.: Transaction Books, 1971. Science Books: A Quarterly Review, 1971, 7(3), 189-190.

R. Schlegel Inquiry into science: its domains and limits. New York: Doubleday, 1972. Science Books: A Quarterly Review, 1972, 8(2), 112.

A. A. Harrison Psychology as a social science. Belmont, Calif.: Wadsworth, 1972. AAAS Science Books, 1973, 9(1), 5.

Scientific American The nature and nurture of behavior: developmental psychobiology. San Francisco: W. H. Freeman, 1973. AAAS Science Books, 1973, 9(3), 193.

A Birenbaum and E. Sagarin (Eds.) People in places: the sociology of the familiar. New York: Praeger, 1973. AAAS Science Books, 1974, 10(2), 97.

B. B. Brown New mind, new body: bio-feedback; new directions for the mind. New York: Harper and Row, 1974. Science Activities, 1974, 11(5).

A. Jacobs and W. W. Spradlin (Eds.) The group as agent of change: treatment, prevention, personal growth in the family, the school, the mental hospital and the community. New York: Behavioral Publications, 1974. AAAS Science Books, 1975, 10(4), 275.

D. H. Walterman People and drugs: a manual on drug abuse. Canogo Park, Calif.: International Council for Prevention of Drug Abuse, 1974. AAAS Science Books & Films, 1975, 11(1), 25.

T. Gordon T.E.T.: teacher effectiveness training. New York: Wyden, 1974. AAAS Science Books & Films, in press.

J. A. Howard The flesh-colored cage: the impact of man's essential aloneness on his attitudes and behavior. New York: Hawthorn, 1975. AAAS Science Books & Films, in press.

E. Fantino and G. Reynolds Introduction to contemporary psychology. San Francisco: W. H. Freeman, 1975. AAAS Science Books & Films, in press.

J. Narciso and D. Burkett Declare yourself: discovering the me in relationships. Englewood Cliffs, N.J.: Prentice-Hall, 1975. Library Journal, in press.

Cook, T. D., Appleton, H., Conner, R. F., Shaffer, A., Tamkin, G., and Weber, S. J. "Sesame Street" revisited. New York: Russell Sage Foundation, 1975. Library Journal, in press.

Papers:

Personalized courses. Paper presented at the meeting of the Minnesota Psychological Association, Minneapolis, April 1969.

Personalized instruction in a small liberal arts college. In G. Mertens (Chm.) What is happening in individualized education. Informal symposium presented at the meeting of the Midwestern Psychological Association, Chicago, May 1969.

The ward staff and management of contingencies: a preliminary report. Paper presented at the meeting of the Joint Conference of the Idaho Psychological Association and the Idaho Speech and Hearing Association, Moscow, Idaho, May 1970. (with L. Abel)

Disruption of "scalloped" responding in a fixed interval schedule by delayed reinforcement. Paper presented at the meeting of the Rocky Mountain Psychological Association, Denver, Colorado, May 1971. (with R. B. Powers)

A case study of a time-out procedure: physical restraints in behavior therapy. Paper presented at the meeting of the Rocky Mountain Psychological Association, Denver, Colorado, May 1971.

Self-management in a beginning psychology course as a laboratory exercise. Paper presented at an informal meeting of Division 25 at the American Psychological Association, Washington, D.C., September 1971. (with R. B. Powers)

The effects of trifluoperazine on DRL schedule transition. Paper presented at the meeting of the Utah Academy of Sciences, Arts, and Letters, Logan, Utah, September 1971. (with C. D. Cheney and R. B. Powers)

Personalized introductory psychology at Utah State University: a progress report, 1970-1971 in summary. Paper presented at the meeting of the Utah Academy of Sciences, Arts, and Letters, Logan, Utah, September 1971. (with R. B. Powers)

Personalized instruction from the viewpoint of a proctor. Paper presented at the meeting of the Utah Academy of Sciences, Arts, and Letters, Logan, Utah, September 1971. (with J. Ensign and R. B. Powers)

(Chm.) The personalized system of instruction as viewed by students. Symposium presented at the meeting of the Rocky Mountain Psychological Association, Albuquerque, New Mexico, May 1972.

The student as a teacher. Paper presented at the meeting of the Rocky Mountain Psychological Association, Albuquerque, New Mexico, May 1972.

Testing in a personalized system of instruction. Paper presented at the meeting of the Rocky Mountain Psychological Association, Albuquerque, New Mexico, May 1972. (with E. Sides)

Self-pacing in a personalized system of instruction: work patterns and course completion. Paper presented at the meeting of the Association for Educational Communications and Technology, Las Vegas, Nevada, April 1973. (with R. B. Powers)

An adjusting reinforcement schedule facilitates exam-taking in a self-paced course. Paper presented at the meeting of the Rocky Mountain Psychological Association, Las Vegas, Nevada, May 1973. (with R. B. Powers and W. Hoehle)

Proctoring oral and written performance in a motivation class. Paper presented at the meeting of the American Psychological Association, Montreal, Quebec, Canada, August 1973. (with P. Gottula)

Students want teachers - not professors. Paper and slide show presented at the meeting of the North Dakota Psychological Association, Bismarck, North Dakota, October 1973. (with P. Gottula)

Measurement techniques and problems in behavior management in the classroom. Paper presented to the Quarterly Pre-service Student Teacher's Meeting, Minot, North Dakota, May 1974 (with R. Charley)

Behavior modification for the classroom teacher: the challenge of the rotten kids. Paper presented at the Workshop in the Open Classroom, Minot, North Dakota, July 1974.

Essentials of classroom management. Paper presented at an In-service Teacher Training Program, White Shield, North Dakota, October 1974.

Behavior modification for the classroom teacher. Paper presented at the Quarterly Pre-service Student Teacher's Meeting, Minot, North Dakota, October 1974.

Behavior modification in the classroom. Paper presented to a meeting of Student Teacher Supervisors, Minot, North Dakota, March 1975.

Back to the one-room school: a step forward. Paper presented to the Quarterly Pre-service Student Teacher's Meeting, Minot, North Dakota, April 1975.

Manipulation of "tipping" behavior in a cocktail lounge. Paper presented at the meeting of the North Dakota Psychological Association, Fargo, North Dakota, October 1975. (with P. Gottula)

Manipulation of "tipping" behavior in a taxicab. Paper presented at the meeting of the North Dakota Psychological Association, Fargo, North Dakota, October 1975. (with P. Gottula and J. Hartung)

First impressions on tests may be better. Paper presented at the meeting of the North Dakota Psychological Association, Fargo, North Dakota, October 1975. (with C. Marshall)