

THE EFFECTS OF INSTRUCTION, COMMENTS, AND PRACTICE UPON TEACHING STUDENTS HOW TO WRITE
BETTER ANSWERS TO ESSAY QUESTIONS

By

DAVID MURRAY DEFRAIN

Bachelor of Arts

Central Missouri State College

Warrensburg, Missouri

1965

Submitted to the faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
MASTER OF SCIENCE
May, 1967

JAN 10 1968

THE EFFECTS OF INSTRUCTION, COMMENTS, AND PRAC-
TICE UPON TEACHING STUDENTS HOW TO WRITE
BETTER ANSWERS TO ESSAY QUESTIONS

Thesis Approved:

Roy Gledstone

Thesis Adviser

Norman E Wilson

Donald J. Smith

N. N. Durbin

Dean of the Graduate College

358668

PREFACE

The college student of today is often given essay examinations as a measure of his knowledge of a subject. In many instances, though, the student is handicapped in the exhibition of his knowledge due to the fact that he does not know how to present this information in an advantageous form. It seemed important, then, to try to minimize the distortion in the assessment of the student's learning. The following study was designed to explore the possibilities of teaching students how to write better answers to essay questions.

Sincere gratitude is extended to Dr. Roy Gladstone for his valued guidance in the execution of this study. To Dr. Donald Tyrell I express thanks for aid in the fulfillment of the mechanics of the study and to Dr. Norman Wilson for his help with the theoretical and conceptual aspects of experimentation. Further thanks is appropriate for Dr. David Bee and Dr. David Shoemaker for their help with the statistical aspects of the study.

Much help was given by Andrew Martin both in allowing me to use his students and in grading answers. Further thanks are extended to Donna DeFrain and Sherri Gragg for their conscientious grading.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Statement of Problem.	2
Review of Related Literature.	3
II. METHODOLOGY AND DESIGN	8
Sample.	8
Procedure	8
Phase I.	10
Phase II	11
Phase III.	13
III. RESULTS.	16
IV. DISCUSSION	21
General Results	21
Future Research	22
V. SUMMARY AND CONCLUSIONS.	24
SELECTED BIBLIOGRAPHY	26
APPENDIX.	27

LIST OF TABLES

Table	Page
I. Criteria for Evaluating Essay Answers	2
II. Grade Placement and Sex Breakdown by Treatment.	9
III. Analysis of Variance for Phase III.	16
IV. Results from Duncan's New Multiple Range Test	17
V. Observed and Projected Phase I Scores	17
VI. Analysis of Variance for Phase I.	18
VII. Test of Homogeneity for Phase III	19

LIST OF FIGURES

Figure	Page
1. Progress of the Experimental Groups Through the Experiment. Phases I and III Represent an Average Over Two Days of Testing, While Phase II Represents an Average Over Four Days of Testing	20

CHAPTER I

INTRODUCTION

The essay examination is often used by college and university teachers. These teachers generally feel the essay test measures some qualities of the student that remain untested with more objective varieties of tests. They feel the student is given a greater opportunity to express himself and structure his knowledge. For these reasons essay questions are used in spite of the difficulty of creating valid tests to cover course content widely and the reported unreliability of scoring essay answers (Stalnaker, 1937; Findlayson, 1951).

Many college students, especially those who have taken few essay examinations, may be faced with an unreasonable task when asked to express their knowledge in essay form. One can often overhear comments made by teachers about the incoherent quality of expression exhibited by many students on essay examinations. This leads to the feeling that students may often be penalized because of their inability to write essay answers, in spite of their mastery of the subject matter. Therefore, in cases where essay writing ability per se is not an objective it would be desirable to minimize the distortion in measurement introduced by variations in essay writing ability so that tests will measure the student's knowledge more accurately.

Statement of the Problem

The objective of this study was to compare the efficacy of four methods of teaching students to write better essay answers to "compare and contrast" questions in psychology. The dependent variable of the study was the ability to write essay answers to "compare and contrast" questions. This ability was analyzed into a number of specific categories. These categories appear in Table I. Categories were chosen for their usefulness in writing essay answers. The number of categories was limited so they could all be adequately covered in a single hour of instruction. The variables of spelling, grammar and syntax were ignored.

TABLE I
CRITERIA FOR EVALUATING ESSAY ANSWERS

-
- A. Does the subject compare and contrast rather than describe each concept separately?
 - B. Does the subject conceptualize comparison categories?
 - C. Does the subject give examples?
 - D. Does the subject define terms?
 - E. Does the subject exclude irrelevant material?
 - F. Does the subject write in sentence form?
-

The independent variables consisted of instruction in answering essay questions, comments written on the papers concerning the adequacy

of the answers, and practice in the form of rewriting answers after they had been commented upon. The three experimental groups each received one hour of instruction on how to write better answers to essay questions. Two of these groups received comments on their papers after taking each examination. One of these two groups was asked to rewrite answers, specifically trying to improve on the categories where they received comments. A control group was included which only took the tests and received a grade.

The study was concerned with how well the independent variables of instruction, comments, and practice aided subjects in learning the dependent variable of writing better answers to "compare and contrast" essay questions, as measured by the six criteria in Table I.

Review of Related Literature

A review of the literature concerning essay-type examinations indicated that little research has been published along the lines of teaching students better essay test writing. Most of the essay test studies have been concerned with the lack of validity in terms of covering course content using essay-type examinations and the lack of reliability in scoring essay test answers.

Some of the literature on essay testing has offered suggestions for improving reliability and validity when using essay-type examinations. Wood (1960) presented ideas on both creating and scoring essay questions. In creating essay questions, she felt short-answer questions would be the most useful. These questions should be so written as to limit the range of the question, in order that a great diversification of answers would not occur. The students should be given adequate time

in which to answer each question. The amount of time allowed for each question should depend upon the depth of the question and the length of answer needed to cover the question adequately.

Suggestions for grading the answers were also made by Wood. She proposed preplanning a scoring guide. This guide should then be used to grade a number of papers, checking to see if the guide would be sufficient for use in grading all the papers. As papers are graded, a few should be occasionally regraded to insure that variations in grading criterion are not occurring. Finally, for best grading results, Wood suggested using the average grade of a number of graders. This would tend to eliminate any biases of a single grader.

Travers (1950) made some suggestions which paralleled those of Wood and went into greater detail in some cases. When limiting the range of a question, Travers felt a number of techniques would be useful. The question should be stated in detail, dividing the question into a series of elements or factors. The examiner might also present an example which could aid the student in confining his answer. The range might be limited by asking for a specific number of concepts, facts or examples to be presented. The student could also be asked, Travers feels, to present his answer in an outline rather than in prose.

In grading the answers, Travers recommended using several judges. The judges should first read a number of answers, analyze the components of a good answer, then decide upon various acceptable answers for each question, keeping in mind certain basic ideas each answer should present.

With respect to the variable of lecture-type instruction, no studies could be found that indicated any use of lectures to teach

students better essay test writing ability. Only one study, using multiple choice questions, was found (Hashimoto, 1956) which gave any indication of the effect of practice upon learning from returned test papers. The multiple choice test was given to four groups of students. After grading the papers Group A did not have their papers returned, while Group B received their papers back with only a grade on them. Group C received comments on the test papers returned to them. Group D was the practice group. When their papers were returned, they were to recheck the multiple choice answers.

The results of a second test showed that Groups A and B made very little improvement. Group C showed the greatest improvement, while Group D showed some improvement, but not as much as Group C. It was also found that the lower performance students showed the greatest improvement between the first and second tests.

A number of other studies besides Hashimoto (1956) have indicated the effects of various types of comments on learning and later test performance. Bryan, Rigney, and Van Horn (1957) used multiple choice questions on three groups of subjects. After the test had been completed, the first group was given the right answer to each question. The second group was told only why the alternative they had chosen on each question was right or wrong. The third group was told what the results might be on each question if the alternative they had chosen would be followed up and acted upon. Although there were no significant differences between the three groups on a new test, all of the groups showed significant gains over the first test.

Page (1958) was interested in the effects of teachers' comments on students' future test performance. Two thousand high school and junior

high school students were first given objective tests over classroom material presented in their normal class situation. Test papers were returned in three different manners. One group received a grade only. Another group received standard written comments corresponding to the grade they made on their paper. These comments were all assumed to be of a positive nature. For the third group papers were returned with general comments given by the teacher whenever it was felt comments were in order. No special types of comments were used with this group.

The first test was used to equate the groups. A second test was used to measure the effects of the three treatments. Results showed both comment groups scored higher than the grade-only group. Further statistical tests showed the treatment effects to be equal to or more effective in the upper grades than in the lower grades. The experimenter suggested these results might extrapolate to at least the lower college years.

Stone (1955) used multiple choice questions in an attempt to show that tests can be used to help students learn. Special interest was shown to what type of verbal comments best aided this learning. Subjects were Air Force personnel attending special training. Five types of verbal comments were employed. Group A was given their total score only. Group B had all questions reread that they had missed, along with the incorrect answer that had been previously chosen. Missed questions were reread for Group C, and the subjects were told why their answers were not correct. They were not told the correct answers. The right answers were told to Group D for all questions they had missed. No mention was made of the other alternatives. The last group, Group E, was told why their answers were wrong, and were told the right answer.

The five groups were each divided in half, with the first half having their retest after twenty-four hours and the second half having their retest after thirty days. For the twenty-four hour condition the critique variation was significant. Group A was significantly different from Groups C, D, and E. The difference between Groups A and B approached significance. For the thirty-day groups the critique variation was not significant, although there was a significant difference between Groups A and E.

Such a result indicated to the experimenter that negative information alone or positive information alone were not enough for long term gains, even though they were significant over a short period of time.

Conclusions

All of the studies which were reviewed found a significant effect from comments. No differences were found between the various types of comments used in the Page (1958) and the Bryan, Rigney, and Van Horn (1957) studies. Stone (1955) found that over an extended period of time only the comments which contained both positive and negative information induced a significant difference over no comments at all.

The study by Hashimoto (1956) gave some indication of the effects of practice upon learning from returned test papers, although his study used multiple choice questions. His practice group showed evidence of some learning, but not as much as did his comments group. No studies were found that were concerned with the effects of verbal instruction.

CHAPTER II

METHODOLOGY AND DESIGN

Sample

Students in Introductory Psychology in the Fall Semester of 1966-67 at Oklahoma State University were used as subjects. The sample consisted specifically of students from four discussion sections, each of which met for an hour, one day a week. All four sections met on the same day and all were taught by the same instructor. There were 137 students enrolled in the four sections. Only those who attended all sessions and were within a specified range on the pretest were used in the final sample. The remaining students were then sampled-back to create an equal number of subjects for each treatment group. Forty-eight subjects, twelve in each group, were used in the final analysis. Table II presents the breakdown of sex and grade placement by treatment groups.

Procedure

All subjects were given a series of eight essay examinations of the "compare and contrast" variety. Each subject received one question a week for eight weeks. There were two to four questions created each week and these were divided randomly among the students. This was done to prevent the sharing of answers between students, both within and between groups.

TABLE II
GRADE PLACEMENT AND SEX BREAKDOWN BY TREATMENT

	Sex		Grade Placement		
	Male	Female	Freshman	Sophomore	Junior
Group A	4	8	9	2	1
Group B	3	9	10	1	1
Group C	6	6	9	2	1
Group D	4	8	10	2	0
Total	17	31	38	7	3

The questions for each week were drawn from the lecture and assigned readings of the previous discussion section. Both the suggestions by Wood (1960) and those of Travers (1950) were used in creating the questions. Every effort was made to design the questions for each week so they would be of approximately equal difficulty and yet would validly cover the appropriate course content. All questions used are reported in Appendix A.

The tests were used for both the student's class grade and the scores for the study, although a different grading system was used to determine the class grades and the scores for the study. Scores for the class grade were based upon the amount and quality of subject matter presented. No concern was made in the class grading with respect to the variables in this study.

The experimental procedure was divided into three phases.

Phase I

In Phase I the scores on the first two questions for each individual were used as a measure of the student's ability to answer "compare and contrast" essay questions. All scores were then arranged in a frequency distribution. Subjects who had scores which fell at the extremes were eliminated from the study. Appendix B gives the frequency distribution and cutoff points. It was felt that subjects who had extremely low scores might not have the ability needed to perform the tasks required of them. Those with extremely high scores might have reached a ceiling before the study was completed. The elimination also served to make the remaining subjects a more homogeneous group, thereby helping to remove a certain amount of ambiguity in the final results.

In all phases the criteria given in Table I were used as the basis for grading, but grading in Phase I was performed somewhat differently than in the other phases due to a difference in the presentation of the questions. The questions used in the first phase contained not only the main question but specific areas of the question as well. For example, in Question 1 the main question was "Compare and contrast the median and the mean." Two of the specific areas of Question 1 were "What are they measures of?" and "How do you determine them numerically from a set of scores?" In contrast, Question 12, which was in Phase II, stated only "Compare and contrast timbre and saturation," giving the student no specific areas. Adding the specific areas was done in order to acquaint the students with the type of answer that would be expected of them in terms of knowledge.

A total of six points was allowed for each specific area given in the question, one point being allowed for each of the six criteria. A

point was given if the subject satisfied a criterion in his answer. For example, if a student compared, conceptualized, gave examples, defined terms, excluded irrelevants, and wrote in sentence form in one area of the question he would receive six points for that area. The total number of points was used as the score for the paper.

The questions for the first day's testing had a possible score of thirty points. The questions for the second day's testing, as well as for all further days, had a possible score of twenty-four.

General guidelines were created for determining the grading of a criterion. These guidelines are listed in Appendix C. Specific guidelines for each question were also devised and are presented in Appendix D.

Phase II

As soon as all subjects had taken two examinations, Phase II began. The control group (Group A) took the four tests of Phase II with no experimental variation. Subjects in this group received only a numerical grade on each paper.

Group B received lecture instruction in techniques useful in writing answers to essay questions immediately after Phase I and before taking the four tests of Phase II. The instruction consisted of generalizations and examples of each of the six criteria used in grading the essay answers. Examples were drawn from the previous questions taken by the subjects in Phase I, as well as examples from sources outside psychology. Instruction consisted of approximately twenty minutes of lecture and class discussion.

Group C received the same type of instruction as Group B and also had comments affixed to their papers. All papers except those with a

score above 85 per cent received three comments. The first and third comments were based upon aspects of the paper which could be improved. The second comment was of a positive nature, reflecting the best aspect of the paper. Papers with a score above 85 per cent were given the single comment of "Excellent."

All comments were based upon the six criteria developed for grading the answers. The six criteria were first ranked according to their apparent difficulty. This was done by noting the frequency with which points were made on each of the criteria by the students in Phase I. Those criteria on which the most points were made by the students were assumed to be the easiest. Those criteria amassing the fewest points were assumed to be the hardest.

In giving the two negative comments, the least difficult criteria which were applicable to the paper being graded were used. In giving positive comments, the most difficult criterion which was applicable to the paper being graded was used. All comments were given in a standard form. Both the positive and the negative comments which were used are given in Appendix E.

The fourth experimental group, Group D, received the same treatment as Group C and were required to rewrite their papers as well. The rewriting was to be done in terms of the comments received on the paper. It was emphasized that it was not always necessary to change the whole paper, only those parts that could be improved by the standards suggested in the instruction portion of the treatment. Incentive was added by telling the subjects that the higher grade of the two papers would be used as their class grade.

Papers in Phase II were graded somewhat differently than those in

Phase I. Although the same criteria and the same general and specific guidelines were used, the elimination of specific areas in presenting the question necessitated a different scoring system. An upper limit of four points for each of the criteria was used, giving a possible total of twenty-four points for any one paper. For the comparison, conceptualization, example, and definition categories, one point was awarded each time one of the requirements for that criterion was fulfilled as based upon the general and specific guidelines. For example, if it was determined by the use of the guidelines that the subject had presented three examples he would be awarded three points for the criterion concerning examples.

In grading the papers using the criterion of irrelevant material, each paper was awarded four points and then one point was subtracted from this figure for every irrelevant statement made. A lower limit of zero points was set for the criterion.

For the criterion pertaining to sentence form, one point was awarded for each closing punctuation (period, question mark, dash, etc.) present in the paper, up to a limit of four points. A point was then subtracted for each instance of non-sentence form (other than those allowed by the guidelines). Again, there was a lower limit of zero points.

Phase III

Scores from essay examinations in the last two class sessions were used to compare the results of the four treatments. Prior to giving these tests for Phase III, treatment groups were sampled-back so that each group contained twelve subjects. This was done by determining the number of subjects in each group which fell within the prescribed range of scores on the Phase I tests who also had attended every class

session. Twelve subjects in each group were then randomly selected from the total number of eligible subjects.

In order to get an accurate measure of experimental error the same three questions were used both days of Phase III. On the first day each student in a group was randomly assigned to one of the three questions until four subjects were assigned to each question. Thus each of the four treatment groups had each question an equal number of times.

On the second day for each group, the four subjects having the same question were randomly divided so that two of them would have one of the questions they had not had, while the remaining two would get the other question. In this way each treatment group again received each question the same number of times, while no individual was given the same question in both sessions.

A spot-check of inter-scorer reliability was carried out on each question given in the study and on each treatment group. In Phases I and II, twenty papers on each question were randomly drawn and were graded by the experimenter and one other grader. The twenty papers were drawn from the total number of papers in all four sections since it was not completely determined until Phase III which subjects would be used in the final analysis (due to absences and sampling-back). In Phase III only the twelve papers for each group were used in the reliability check.

Three different graders were used during the study, but one was used almost exclusively during the second and third phases. All graders were instructed in the criteria used in grading. Occasional checks were made for variations in grading, and occasional sessions were held to discuss the grading. Copies of the specific and general guidelines were given to the graders as aids.

Due to the small number of graders on any one question, only the grades of the experimenter were used in the statistical analysis.

CHAPTER III

RESULTS

Some concern was felt by the experimenter as to the feasibility of using the first two tests in Phase I as a measure of essay-test writing ability. Since all groups did not receive the same questions, it was necessary to determine if any significant difference existed between treatment groups. In order to accomplish this a missing data formula was used to fill in the 4 x 4 table presented in Table III. Each cell represents the average score for the treatment group on that question in Phase I. Only the Phase I scores of the twelve subjects in each group included after the sampling-back procedure were used.

TABLE III

OBSERVED AND PROJECTED PHASE I SCORES

(Scores underlined were observed averages; scores in parentheses were generated by missing data formula)

	Group A	Group B	Group C	Group D	Question Average
Question 1	(<u>52.13</u>)	<u>40.28</u>	(<u>42.13</u>)	50.00	46.14
Question 2	<u>45.00</u>	(<u>34.83</u>)	<u>36.67</u>	(<u>42.85</u>)	39.84
Question 3	<u>41.14</u>	(<u>29.29</u>)	(<u>30.29</u>)	<u>36.46</u>	34.30
Question 4	(<u>57.73</u>)	<u>47.57</u>	<u>46.88</u>	(<u>54.74</u>)	51.73
Group Average	49.00	38.00	38.99	46.01	

An analysis of variance was then performed on the full table. The results are given in Table IV. Treatments were not significant.

TABLE IV
ANALYSIS OF VARIANCE FOR PHASE I

Source	df	Mean Square	F
Total	7		
Questions	3	29.838	5.223
Treatments (adj.)	3	29.137	5.101
Residual	1	5.712	

* Significant at .05

To further substantiate this result a test of homogeneity (Glass, 1966) was run on the Phase III scores. Table V furnishes the statistical breakdown.

TABLE V
TEST OF HOMOGENEITY FOR PHASE III

Source	df	Mean Square	F
Total	47		
Treatments	3	48.872	1.381
Residual	44	35.383	

* Significant at .05

The scores from Phase III of the experimental design were used for the final statistical analysis. A 2 x 3 x 4 factorial arrangement was tested, with two different days of testing, three different questions and four different treatments. The form of the data used was the percent of the possible score for an answer. The scores of all forty-eight subjects are presented in Appendix F.

An analysis of variance was run on the data. Table VI gives the breakdown of the analysis.

TABLE VI
ANALYSIS OF VARIANCE FOR PHASE III

Source	df	Mean Square	F
Total	95		
Times	1	30.521	1
Questions	2	707.640	4.573*
Treatments	3	738.369	4.772**
Times x Questions	2	382.074	2.469
Times x Treatments	3	110.613	1
Questions x Treatments	6	148.233	1
Times x Questions x Treatments	6	23.799	1
Residual	72	154.746	

* Significant at .025

** Significant at .005

In order to determine where the significance might lie between individual treatments, Duncan's (1955) new multiple range test was applied. The results are given in Table VII. The computations used in the multiple range test are presented in Appendix G.

TABLE VII
RESULTS FROM DUNCAN'S NEW MULTIPLE RANGE TEST

Treatment Groups	Group C	Group A	Group B	Group D
Group C		1.563	5.903	10.763**
Group A			5.103	9.200*
Group B				4.860
Group D				

* Significant at .05

** Significant at .01

Table VII indicates that Group D was significantly different from Groups A and C. No other significant differences were found.

Figure 1 offers a graphic representation of the progression of the test scores for the four treatment groups. Points represent the means of the per cent of the possible score for a question for each group in each phase. It must be noted that in Phase I the groups did not receive the same questions, therefore making the Phase I results somewhat ambiguous.

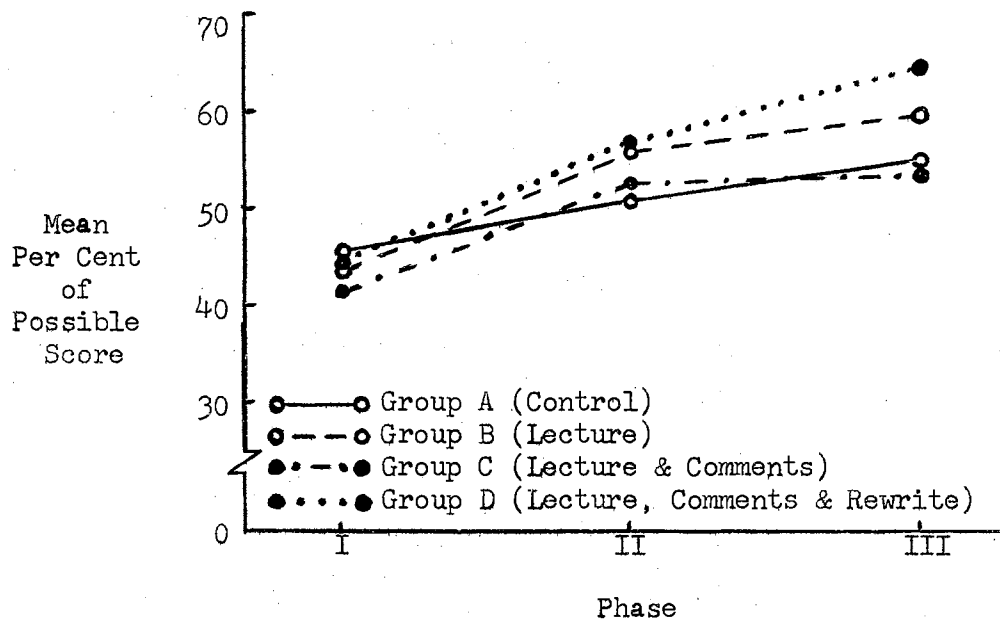


Figure 1. Progress of the experimental groups through the experiment. Phase I and III represent an average over two days of testing, while Phase II represents an average over four days of testing.

The effects of rewriting for Group D are not wholly present in Phase II. This was due to the fact that the papers from the first day's testing in Phase II were not handed back until the second week. The re-writing of the first test of Phase II, therefore, was not completed by the students until the third week of Phase II. This may to some degree account for the divergence of Group D and the other groups from Phase II to Phase III.

The spot-checks of inter-scorer reliability which were run for each question and each treatment group are given in Appendix H. The reliability of grading using all twenty-six questions was .74, with a range from .46 to .87 between questions and a range of .67 to .83 between treatment groups.

CHAPTER IV

DISCUSSION

Since the four treatment groups did not receive the same questions in Phase I, it was of some importance to show that there was no significant difference between the groups. The scores on the Phase I questions were used as indications of essay-test writing ability and were further used as a criterion for including or excluding subjects in the study. Analysis of variance, after applying a missing data formula, indicated no significant difference. This suggested that the Phase I scores could be used as a criterion for the inclusion or exclusion of subjects.

Those results were somewhat spurious, however, since only one degree of freedom was left for the residual term after the missing data formula was applied. Therefore, to further substantiate the results, a test of homogeneity was run to determine if the group samples were drawn from a common population. The test of homogeneity was also non-significant, thereby giving additional evidence that the scores on the Phase I questions could be used as a measure of essay writing ability for the experimental population.

General Results

The significant differences between treatments indicated in the analysis of variance of the Phase III data demonstrated that teaching

students how to write better answers to "compare and contrast" essay questions was more effective than having the students take the tests with no special instruction. Although the Questions variance was also significant, it should not have affected the significance of the treatments since all treatment groups had each question the same number of times and the Questions x Treatments variance was not significant.

Individual group comparisons indicated that neither the instruction (Group B) nor the instruction and comments (Group C) were significantly better than taking the test with no instruction or comments (Group A). Only the group which received instruction, comments, and had to rewrite their papers (Group D) was significantly better than the control group. Group D was also found to be significantly better than Group C, and better than but not significantly different from Group B. The most defensible conclusion would seem to be that students can be taught how to answer "compare and contrast" essay questions if they are given initial instruction in specific criteria to be included in writing answers, are provided comments concerning the criteria on the returned papers, and are asked to rewrite their answers in the light of the comments. Replication might indicate which of the steps are most relevant and worth the effort and which are least relevant and could be excluded.

Future Research

A replication of this study might determine whether it is necessary to use all three of the variables of lecture instruction, comments, and rewriting in achieving the results. It may be that the most efficient results would be obtained by elimination of one or two of these steps.

A number of areas could be examined in new research concerning the teaching of essay examination writing ability. It might be profitable to develop new criteria, other than those used in this experiment. No attempt was made in this study to cover all or even the major variables in essay writing ability.

The study of essay examination writing ability should be attempted with different populations. It may be possible that different populations would show dissimilar results. Experimentation should be carried out with younger subjects in an attempt to determine the most efficient time for teaching this ability.

A most important area of research would be the transfer of the essay examination writing skill to situations other than this experimental case. It should be determined whether a skill must be developed for each new content area or whether the skill will generalize to those areas.

CHAPTER V

SUMMARY AND CONCLUSIONS

In an effort to determine if it was possible to teach students how to write better answers to "compare and contrast" essay questions, four discussion groups in Introductory Psychology were selected for study. In Phase I all groups were given two ten-minute essay tests as a measure of essay writing ability. Each group was then randomly given a different experimental treatment in Phase II. Group A took four tests and received only a grade. Group B was given lecture instruction in writing better answers before taking the four tests, while Group C was given instruction plus comments on their four returned papers. Instruction and comments were also given to Group D, and they were asked to rewrite their answers to each of the four tests in line with the comments.

The two tests given to each group in Phase III were used to compare the effects of the various treatments.

Six criteria of good essay answers were chosen as aids for grading the essay tests. These same criteria were also used as the basis of the lecture instruction and the comments.

An analysis of variance of the Phase III results was performed. A $2 \times 3 \times 4$ factorial arrangement was tested, with two days of testing, three questions, and four treatments. The analysis indicated an overall significant difference between the experimental treatments at the .005 level. The difference between questions was also significant, but did

not seriously affect the treatment results since the Questions x Treatments interaction was not significant.

A Duncan's (1955) new multiple range test was run to determine where the significant differences lay between the individual treatment groups. This analysis showed Group D to be significantly better than Group C at the .01 level, significantly better than Group A at the .05 level and better than Group B but not significantly so.

Spot-checks of inter-scorer reliability were run for each question used in the study. An overall reliability of .74 was found for the twenty-six questions.

From these results it was concluded that it is possible to teach students how to write better answers to "compare and contrast" essay questions. According to the statistical analysis the method found best for this result was to give the students instruction in essay test writing, provide comments on the returned papers, and have the students rewrite their papers. Further research might indicate whether all of these steps are necessary for the most efficient results.

SELECTED BIBLIOGRAPHY

- Bryan, Glenn L., Joseph W. Rigney, and Charles Van Horn. An evaluation of three types of information for supplementing knowledge of results in a training technique. University of Southern California Elect. Personnel Research Technical Report, No. 19 iv, 1957.
- Duncan, D. B. Multiple range tests for correlated and heteroscedastic means. Biometrics, 1957, 13, 164-176.
- Edwards, Allen L. Experimental design in psychological research. New York: Holt, Rinehart and Winston, 1964.
- Findlayson, D. S. The reliability of the marking of essays. Brit. J. educ. Psychol., 1951, 21, 126-134.
- Glass, Gene V. Testing homogeneity of variances. Amer. educ. res. J., 1966, 3, 187-190.
- Hashimoto, Juji. Tōan henkyaku no hōhō ga gakushū-kōkā ni oyobusu eikyō no kenkyū. (Effects of various methods of returning test papers upon achievement of pupils.) Jap. J. educ. Psychol., 1956, 3, 146-156. (Information used was taken from Psychol. Abstr., 1958, 32, article 4604.)
- Page, Ellis B. Teacher comments and student performance: A seventy-four classroom experiment in school motivation. J. educ. Psychol., 1958, 49, 173-181.
- Stalnaker, J. M. Question VI, The essay. English J. (College Ed.), 1937, 26, 133-140.
- Steel, Robert G., and James H. Torrie. Principles and procedures of statistics. New York: McGraw-Hill Book Company, Inc., 1960.
- Stone, G. Raymond. The training function of examinations: retest performance as a function of the amount and kind of critique information. USAF Pers. Train. Res. Cent. Res. Rep., No. AFPTRC-TN-55-8, 1955.
- Travers, R. M. W. How to make achievement tests. New York: The Odyssey Press, 1950.
- Wood, Dorothy A. Test construction: Development and interpretation of achievement tests. Columbus, O.: Charles E. Merrill Books, Inc., 1960, Chap. 10.

APPENDIX A

APPENDIX A

QUESTIONS USED AS EXAMINATIONS

1. Compare and contrast the median and the mean. Base your answer on the following areas: (a) What are they measures of? (b) What type of statistics are each used for, descriptive or inferential? (c) How do you determine them numerically from a set of scores? (d) What difficulties might occur in using them? (e) Where do they occur in a frequency distribution?
2. Compare and contrast the range and the standard deviation. Base your answer on the following areas: (a) What are they measures of? (b) What type of statistics are each used for, descriptive or inferential? (c) How do you determine them numerically from a set of scores? (d) What difficulties might occur in using them? (e) How do they relate to a frequency distribution?
3. Compare and contrast positive and negative correlation. (a) What are the numerical limits of each? (b) Can you predict better if the correlation for each is high or low? (c) How do they look on a graph? (d) What problems might occur in interpreting them? (Also included but not used in grading: "What would each indicate in terms of reliability?")
4. Compare and contrast reliability and validity. (a) How do you get scores in order to find the reliability and validity of a test? (b) What is the relationship of reliability and validity to correlation? (c) Can a test be reliable without being valid? or valid without being reliable? (d) Why aren't reliability and validity always accurate measures of a test?
5. Compare and contrast hue and brightness.
6. Compare and contrast hue and saturation.
7. Compare and contrast the additive and subtractive methods of color mixture.
8. Compare and contrast color mixing using complementary colors and color mixing using non-complementary colors.
9. Compare and contrast intensity and brightness.
10. Compare and contrast pitch and intensity.
11. Compare and contrast pitch and timbre.
12. Compare and contrast timbre and saturation.
13. Compare and contrast the unconditioned stimulus and the unconditioned response.

14. Compare and contrast the conditioned stimulus and the conditioned response.
15. Compare and contrast the unconditioned response and the conditioned response.
16. Compare and contrast the unconditioned stimulus and the conditioned stimulus.
17. Compare and contrast stimulus generalization and response generalization.
18. Compare and contrast stimulus discrimination and response discrimination.
19. Compare and contrast stimulus discrimination and stimulus generalization.
20. Compare and contrast response discrimination and response generalization.
- 21 & 24. Compare and contrast reward conditioning and avoidance conditioning.
- 22 & 25. Compare and contrast escape conditioning and reward conditioning.
- 23 & 26. Compare and contrast escape conditioning and avoidance conditioning.

APPENDIX B

APPENDIX B

FREQUENCY OF SCORES FOR PHASE I

Scores were computed by adding each individual's score on his first question to the score on his second question. Only those subjects whose scores fell between the dotted lines were used in the study.

Score	Group A	Group B	Group C	Group D
14			1	1
15				
16			1	1
17	2		1	
18	1		3	1
19			2	2
<hr/>				
20	1	3	4	2
21	1		2	2
22	4	1	1	2
23		3	2	2
24	6	2	1	3
25	6	2	3	1
26	6	4	3	5
<hr/>				
27	1	4		1
28	1	1	2	
29	2	3	1	2
30		1	1	1
31		2		2
32		1	1	
33				
34	1	1		
35		1		
36				
37				
38		1		
39				
40				

APPENDIX C

APPENDIX C

GENERAL GUIDELINES FOR EVALUATING ESSAY ANSWERS

- A. Does the subject compare and contrast rather than describe each concept separately?
1. "Both ...," "They ...," "Each ...," (while, whereas, although, but)
 2. One statement inferring both concepts
 3. One concept affects the other
 4. One concept related to the other
 5. Both concepts related to a third concept
 6. Graphic comparisons
 7. Measurement of each
 8. Both used in the same example
 9. Results of both
- B. Does the subject conceptualize comparison categories?
1. A generalization in which both concepts are included
 2. Both concepts related to a third concept
 3. One concept related to the other
 4. One concept affects the other
 5. Results of both
 6. Graphic comparisons
 7. Measurement of each (and their relationship)
 8. Formulas of each (and their relationship)
- C. Does the subject give examples?
1. "For example ...," "... is like ...," "... is a ...,"
 2. Cites an in-class demonstration
 3. Cites an example given in the text
 4. Cites experimental evidence
 5. Gives an original example
 6. Refers to an example previously mentioned in the paper
 7. Graphic examples
 8. Drawings
 9. Diagrams
- D. Does the subject define terms?
1. "A definition of ... is ...," "... is ...," "... (defines in parentheses) ..."
 2. Any terms used should be defined
 3. Defines the parts of a formula
 4. Cites text definition
- E. Does the subject exclude irrelevant material?
1. Should include only information needed to fully and directly answer the question

2. Must link all statements to the main question in some manner
3. Should exclude general statements which could be made without any knowledge of the correct answer
4. Should exclude meaningless phrases which repeat previous statements
5. Should not "pad" answer

F. Does the subject write in sentence form?

1. Statements should be in complete sentence form
2. Must be coherent sentences
3. Exceptions:
 - (a) Where step form might be appropriate
 - (b) Within examples phrases may be used

APPENDIX D

APPENDIX D

SPECIFIC GUIDELINES FOR EVALUATING ESSAY ANSWERS

(The letters in parentheses refer to the specific part of the question where the comment might be found)

Question 1: Compare and contrast the median and the mean.

A. Comparisons and Contrasts

1. Central tendency (a)
2. Inferential statistics (b)
3. Descriptive statistics (b)
4. Numerical similarities and differences (c)
5. Common difficulties (d)

B. Conceptualizations

1. Central tendency (a)
2. Inferential statistics (b)
3. Descriptive statistics (b)
4. Dispersion (d)
5. Deviation (d)
6. Normal distribution (e)
7. Symmetrical distribution (e)

C. Examples

1. Mean (a)
2. Median (a)
3. Descriptive statistic (b)
4. Inferential statistic (b)
5. Computations of mean and/or median (c)
6. Frequency distribution (e)
7. Normal distribution (e)
8. Symmetrical distribution (e)

D. Definitions

1. Central tendency (a)
2. Inferential statistics (b)
3. Descriptive statistics (b)
4. Mean (c)
5. Median (c)
6. Normal distribution (d)
7. Symmetrical distribution (d)
8. Variability (d)
9. Frequency distribution (e)

E. Irrelevant Material

1. Should not list things from which means and medians can be found (a)
2. Not necessary to mention what kinds of scores can be used (c)
3. Not necessary to state "There are many difficulties in using the mean and the median." (d)

F. Sentence Form

Question 2: Compare and contrast the range and the standard deviation.

A. Comparisons and Contrasts

1. Variability (a)
2. Dispersion (a)
3. Descriptive statistics (b)
4. Inferential statistics (b)
5. Numerical similarities and differences (c)
6. Common difficulties (d)

B. Conceptualizations

1. Variability (a)
2. Dispersion (a)
3. Descriptive statistics (b)
4. Inferential statistics (b)
5. Central tendency (d)
6. Normal distribution (e)
7. Symmetrical distribution (e)

C. Examples

1. Range (a)
2. Standard deviation (a)
3. Descriptive statistics (b)
4. Inferential statistics (b)
5. Computation of range and/or standard deviation (c)
6. Frequency distribution (e)
7. Normal distribution (e)
8. Symmetrical distribution (e)

D. Definitions

1. Variability (a)
2. Dispersion (a)
3. Inferential statistics (b)
4. Descriptive statistics (b)
5. Range (c)
6. Standard deviation (c)
7. Normal distribution (d)
8. Symmetrical distribution (d)
9. Variability (d)
10. Frequency distribution (e)

E. Irrelevant Material

1. Should not list things from which ranges and standard deviations can be found (a)
2. Not necessary to mention what kinds of scores can be used (c)
3. Not necessary to state "There are many difficulties in using the range and standard deviation." (d)

F. Sentence Form

Question 3: Compare and contrast positive and negative correlation.

A. Comparisons and Contrasts

1. "The limits of correlation are ..." (a)
2. $+1 \rightarrow 0 \rightarrow -1$ (a)
3. Draw both on the same graph (c)
4. "On a graph ..." (c)

B. Conceptualizations

1. Perfect correlation (a)
2. Coefficient of correlation (a)
3. Draws on a graph (c)
4. Refers to slope (c)
5. Refers to there not being a cause-effect relationship between variables in correlation (d)

C. Examples

1. Positive correlation (a)
2. Negative correlation (a)
3. Correlation (a)
4. Perfect correlation (a)
5. Good predictability (b)
6. Poor predictability (b)
7. Labels axes of graph (c)
8. ".25 is not half of .50." (d)

D. Definitions

1. Positive correlation (a)
2. Negative correlation (a)
3. Correlation (a)
4. Perfect correlation (a)
5. Coefficient of correlation (a)
6. Slope (c)

E. Irrelevant Material

1. Does not talk about prediction (b)
2. Does not indicate if one is better than the other (b)
3. Should refer to the graphing of the parts (c)

4. Should not say "There are many problems in interpreting correlation." (d)
5. Should refer to interpretation (d)

F. Sentence Form

Question 4: Compare and contrast reliability and validity.

A. Comparisons and Contrasts

1. "Reliability is the test compared to its retest while validity is the test compared to an outside criterion." (a)
2. Both correlations (b)

B. Conceptualizations

1. Outside criterion (a)
2. Test-retest (a)
3. "Real world" (a)
4. Correlations (b)
5. Concept of validity (c)
6. Predictability (d)
7. Test conditions (d)

C. Examples

1. Test-retest (a)
2. Outside criterion (a)
3. Correlation as reliability (b)
4. Correlation as validity (b)
5. Reliability without validity (c)
6. Validity without reliability (c)
7. Examples of problems (d)

D. Definitions

1. Reliability (a)
2. Validity (a)
3. Outside criterion (a)
4. Test-retest (a)
5. Positive correlation (b)
6. Negative correlation (b)
7. Correlation (b)

E. Irrelevant Material

1. Should not just say "By collecting data" or "By testing." (a)

F. Sentence Form

Question 5: Compare and contrast hue and brightness.

A. Comparisons and Contrasts

1. Both deal with color
2. Both can be observed by the eye
3. Both can be changed through color mixture
4. Both are degrees of color
5. Both are related to "pure" color
6. Both are related to frequency

B. Conceptualizations

1. Color
2. Wave length
3. Depth
4. "Pure" color
5. Measurement
6. Color mixture
7. Sensation
8. Perception
9. Frequency

C. Examples

1. Hue
2. Brightness
3. Saturation
4. Amplitude
5. Wave length

D. Definitions

1. Hue
2. Brightness
3. Saturation
4. Color
5. Amplitude
6. Wave length

E. Irrelevant Material

1. Should not go into detail about saturation

F. Sentence Form

Question 6: Compare and contrast hue and saturation.

A. Comparisons and Contrasts

1. Both deal with color
2. Both can be observed by the eye
3. Both can be changed through color mixture
4. Both are degrees of color

5. Both are related to "pure" color
6. Both are related to frequency

B. Conceptualizations

1. Color
2. Wave length
3. Depth
4. "Pure" color
5. Measurement
6. Color mixture
7. Sensation
8. Perception
9. Frequency

C. Examples

1. Hue
2. Saturation
3. Brightness
4. Amplitude
5. Wave length

D. Definitions

1. Hue
2. Saturation
3. Brightness
4. Color
5. Amplitude
6. Wave length

E. Irrelevant Material

1. Should not go into detail about brightness

F. Sentence Form

Question 7: Compare and contrast the additive and subtractive methods of color mixture.

A. Comparisons and Contrasts

1. Results of mixing
2. Purity of colors
3. Effects on hue
4. Effects on brightness
5. Effects on saturation
6. Materials used in mixing

B. Conceptualizations

1. Complementary colors

2. Non-complementary colors
3. Purity of colors
4. Hue
5. Brightness
6. Saturation
7. Differences in method
8. Results of mixing

C. Examples

1. Subtractive method
2. Additive method
3. Subtractive mixture
4. Additive mixture

D. Definitions

1. Additive law
2. Subtractive law
3. Complementary colors
4. Non-complementary colors
5. Color mixing
6. Hue
7. Brightness
8. Saturation

E. Irrelevant Material

F. Sentence Form

Question 8: Compare and contrast color mixing using complementary colors and color mixing using non-complementary colors.

A. Comparisons and Contrasts

1. Results of mixing
2. Both on color wheel
3. Both under additive or subtractive laws

B. Conceptualizations

1. Color mixing using pigments
2. Color mixing using lights
3. Positions on color wheel
4. Subtractive law
5. Additive law
6. Color transmission

C. Examples

1. Complementary color mixing
2. Non-complementary color mixing
3. Color wheel

4. Additive law
5. Subtractive law
6. Color mixing using pigments
7. Color mixing using lights

D. Definitions

1. Complementary colors
2. Non-complementary colors
3. Primary colors
4. Additive law
5. Subtractive law
6. Color wheel

E. Irrelevant Material

1. Should not talk about "true colors"
2. Should not talk about pastels

F. Sentence Form

Question 9: Compare and contrast intensity and brightness.

A. Comparisons and Contrasts

1. Relation to "pure" waves
2. Length of wave
3. Frequency of wave
4. Amplitude of wave
5. Change in wave
6. Both are perceptions
7. Both are sensations
8. Measurement of each
9. Both are waves
10. Chart of relationship

B. Conceptualizations

1. Waves of energy
2. Frequency
3. Amplitude
4. Change in wave
5. Perception
6. Sensation
7. Measurement of each
8. Waves
9. Chart of relationships

C. Examples

1. Intensity
2. Brightness
3. Wave

4. Amplitude
5. Frequency

D. Definitions

1. Intensity
2. Brightness
3. Wave
4. Amplitude
5. Frequency

E. Irrelevant Material

1. Should not dwell on the other relationships on the chart of relationships

F. Sentence Form

Question 10: Compare and contrast pitch and intensity.

A. Comparisons and Contrasts

1. Both are sound waves
2. Frequency
3. Amplitude
4. Results when one is constant while the other varies
5. Both part of hearing (Auditory)
6. Change in sound
7. Physical vs. psychological
8. Chart of relationships

B. Conceptualizations

1. Sound waves
2. Frequency
3. Amplitude
4. Results when one is constant while the other varies
5. Hearing (Audition)
6. Change in sound
7. Physical vs. psychological
8. Chart of relationships

C. Examples

1. Pitch
2. Intensity
3. Frequency
4. Amplitude
5. Timbre
6. Sound wave

D. Definitions

1. Pitch
2. Intensity
3. Frequency
4. Amplitude
5. Timbre

E. Irrelevant Material

1. Should not be greatly concerned with relationship to color

F. Sentence Form

Question 11: Compare and contrast pitch and timbre.

A. Comparisons and Contrasts

1. Both are sound waves (or sound)
2. Overtones
3. Audition
4. Both are sensations
5. Both are perceptions
6. Frequency
7. Amplitude
8. Chart of relationships

B. Conceptualizations

1. Sound waves
2. Overtones
3. Hearing (Audition)
4. Sensation
5. Perception
6. Frequency
7. Amplitude
8. Chart of relationships

C. Examples

1. "Pure" wave
2. Pitch
3. Timbre
4. Intensity
5. Overtones
6. Frequency
7. Amplitude
8. Pitch (being similar to hue)
9. Timbre (being similar to saturation)

D. Definitions

1. Pitch
2. Timbre
3. Intensity

4. "Pure" wave
5. Overtones
6. Frequency
7. Audition
8. "White" sound (or "white" noise)
9. Threshold

E. Irrelevant Material

1. Should not be too concerned with relationship to color

F. Sentence Form

Question 12: Compare and contrast timbre and saturation.

A. Comparisons and Contrasts

1. Both are waves
2. Color vs. sound
3. Visual vs. auditory
4. Both are sensations
5. Both are perceptions
6. Both are mixtures
7. Both deal with purity
8. Psychological response
9. Both deal with depth
10. Chart of relationships

B. Conceptualizations

1. Mixture
2. Purity
3. Psychological response
4. Depth
5. Sensation
6. Perception
7. Wave
8. Chart of relationships

C. Examples

1. Timbre
2. Saturation
3. Mixture
4. Purity

D. Definitions

1. Timbre
2. Saturation
3. Mixture
4. Purity

D. Definitions

1. Timbre
2. Saturation
3. Mixture
4. Purity

E. Irrelevant Material

1. Should not dwell on the other relationships on the chart of relationships

F. Sentence Form

Question 13: Compare and contrast the unconditioned stimulus and the unconditioned response.

A. Comparisons and Contrasts

1. Both are motivational
2. Both based on previous experience
3. Both are "natural" or "normal"
4. Both are unconditioned
5. Both are part of classical conditioning
6. Classical conditioning paradigm

B. Conceptualizations

1. Motivation
2. Previous or past experience
3. "Naturalness" or "Normalness"
4. Unconditioned
5. Classical conditioning paradigm
6. Classical conditioning

C. Examples

1. Unconditioned stimulus
2. Unconditioned response
3. Conditioned stimulus
4. Conditioned response
5. Pavlov's experiment
6. Class experiment

D. Definitions

1. Unconditioned stimulus
2. Unconditioned response
3. Conditioned stimulus
4. Conditioned response
5. Stimulus
6. Response

7. Unconditioned
8. Conditioned

E. Irrelevant Material

1. Stimulus generalization
2. Response
3. Spontaneous recovery
4. Extinction
5. Should not dwell on conditioned stimulus and response

F. Sentence Form

Question 14: Compare and contrast the conditioned stimulus and the conditioned response.

A. Comparisons and Contrasts

1. Both are motivational
2. Both are based on previous experience
3. Both are "natural" or "normal"
4. Both are conditioned
5. Both part of classical conditioning
6. Classical conditioning paradigm
7. Both part of learning

B. Conceptualizations

1. Motivation
2. Previous experience
3. "Naturalness" or "Normalness"
4. Conditioned
5. Classical conditioning
6. Classical conditioning paradigm
7. Learning

C. Examples

1. Unconditioned stimulus
2. Unconditioned response
3. Conditioned stimulus
4. Conditioned response
5. Pavlov's experiment
6. Class experiment

D. Definitions

1. Unconditioned stimulus
2. Unconditioned response
3. Conditioned stimulus
4. Conditioned response
5. Stimulus
6. Response

7. Conditioned
8. Unconditioned

E. Irrelevant Material

1. Should not dwell on the unconditioned stimulus and response
2. Stimulus generalization
3. Response generalization
4. Spontaneous recovery
5. Extinction

F. Sentence Form

Question 15: Compare and contrast the unconditioned response and the conditioned response.

A. Comparisons and Contrasts

1. Both result from a stimulus
2. Both part of classical conditioning
3. Amount of response
4. One "natural," other not "natural"
5. One learned, other not learned
6. Classical conditioning paradigm

B. Conceptualizations

1. Stimuli
2. Classical conditioning
3. Amount of response
4. "Naturalness"
5. Learning
6. Classical conditioning paradigm

C. Examples

1. Unconditioned stimulus
2. Unconditioned response
3. Conditioned stimulus
4. Conditioned response
5. Pavlov's experiment
6. Class experiment

D. Definitions

1. Unconditioned stimulus
2. Unconditioned response
3. Conditioned stimulus
4. Conditioned response
5. Stimulus
6. Response
7. Conditioned
8. Unconditioned

E. Irrelevant Material

1. Stimulus generalization
2. Response generalization
3. Spontaneous recovery
4. Extinction
5. Should not dwell on unconditioned and conditioned stimulus

F. Sentence Form

Question 16: Compare and contrast the unconditioned stimulus and the conditioned stimulus.

A. Comparisons and Contrasts

1. Both motivational
2. Both part of classical conditioning
3. Both elicit a response
4. Amount of response
5. One "natural," other not "natural"
6. One learned, other not learned
7. Classical conditioning paradigm
8. Both part of learning

B. Conceptualizations

1. Motivation
2. Classical conditioning
3. Eliciting responses
4. Amount of response
5. "Naturalness"
6. Learning
7. Classical conditioning paradigm

C. Examples

1. Unconditioned stimulus
2. Unconditioned response
3. Conditioned stimulus
4. Conditioned response
5. Pavlov's experiment
6. Class experiment

D. Definitions

1. Unconditioned stimulus
2. Unconditioned response
3. Conditioned stimulus
4. Conditioned response
5. Stimulus
6. Response
7. Conditioned
8. Unconditioned

E. Irrelevant Material

1. Stimulus generalization
2. Response generalization
3. Spontaneous recovery
4. Extinction
5. Should not dwell on unconditioned and conditioned response

F. Sentence Form

Question 17: Compare and contrast stimulus generalization and response generalization.

A. Comparisons and Contrasts

1. Both part of classical conditioning
2. Both generalizations
3. Both part of learning
4. Both used as basis for response
5. Both affect the response
6. Diagrams of relationships

B. Conceptualizations

1. Classical conditioning
2. Generalization
3. Learning
4. Response
5. Diagrammatical relationships

C. Examples

1. Stimulus generalization
2. Response generalization
3. Pavlov's experiment
4. Class experiment
5. Unconditioned stimulus
6. Unconditioned response
7. Conditioned stimulus
8. Conditioned response

D. Definitions

1. Stimulus generalization
2. Response generalization
3. Stimulus
4. Response
5. Generalization
6. Conditioning
7. Unconditioned stimulus
8. Unconditioned response
9. Conditioned stimulus
10. Conditioned response

E. Irrelevant Material

1. Should not dwell on stimulus and response discrimination

F. Sentence Form

Question 18: Compare and contrast stimulus discrimination and response discrimination.

A. Comparisons and Contrasts

1. Both part of classical conditioning
2. Both discriminations
3. Both part of learning
4. Both used as basis for response
5. Both affect the response
6. Diagrams of relationships

B. Conceptualizations

1. Classical conditioning
2. Discrimination
3. Learning
4. Response
5. Diagrammatical relationship

C. Examples

1. Stimulus discrimination
2. Response discrimination
3. Stimulus
4. Response
5. Unconditioned stimulus
6. Unconditioned response
7. Conditioned stimulus
8. Conditioned response
9. Class experiment

D. Definitions

1. Stimulus discrimination
2. Response discrimination
3. Stimulus
4. Response
5. Conditioning
6. Unconditioned stimulus
7. Unconditioned response
8. Conditioned stimulus
9. Conditioned response
10. Discrimination
11. Extinction
12. Inhibition

E. Irrelevant Material

1. Should not dwell on stimulus and response generalization

F. Sentence Form

Question 19: Compare and contrast stimulus discrimination and stimulus generalization.

A. Comparisons and Contrasts

1. Both used as basis for response
2. Both part of learning
3. Both elicit responses
4. May or may not inhibit (extinguish) responses
5. Both deal with stimuli
6. Both part of classical conditioning
7. Diagrams of relationships

B. Conceptualizations

1. Responses
2. Learning
3. Inhibition
4. Extinction
5. Stimuli
6. Classical conditioning
7. Diagrammatical relationship

C. Examples

1. Stimulus discrimination
2. Stimulus generalization
3. Stimulus
4. Response
5. Unconditioned stimulus
6. Unconditioned response
7. Conditioned stimulus
8. Conditioned response
9. Pavlov's experiment
10. Class experiment

D. Definitions

1. Stimulus discrimination
2. Stimulus generalization
3. Stimulus
4. Response
5. Conditioning
6. Unconditioned stimulus
7. Unconditioned response
8. Conditioned stimulus
9. Conditioned response

10. Discrimination
11. Inhibition
12. Extinction

E. Irrelevant Material

1. Should not dwell on response discrimination and generalization

F. Sentence Form

Question 20: Compare and contrast response discrimination and response generalization.

A. Comparisons and Contrasts

1. Both types of response
2. Both part of learning
3. Refers to type or kind of response
4. Refers to inhibition and noninhibition of response
5. Both part of classical conditioning
6. Diagrams of relationships

B. Conceptualizations

1. Responses
2. Learning
3. Inhibition
4. Classical conditioning
5. Diagrammatical relationship

C. Examples

1. Response discrimination
2. Response generalization
3. Stimulus
4. Response
5. Unconditioned stimulus
6. Unconditioned response
7. Conditioned stimulus
8. Conditioned response
9. Pavlov's experiment
10. Class experiment

D. Definitions

1. Response discrimination
2. Response generalization
3. Stimulus
4. Response
5. Conditioning
6. Unconditioned stimulus
7. Unconditioned response
8. Conditioned stimulus

9. Conditioned response
10. Discrimination
11. Generalization
12. Inhibition
13. Extinction

E. Irrelevant Material

1. Should not dwell on stimulus discrimination and generalization

F. Sentence Form

Question 21 & 24: Compare and contrast reward conditioning and avoidance conditioning.

A. Comparisons and Contrasts

1. Reward vs. punishment
2. Both conditioning (instrumental)
3. Both learning
4. Different results from responding

B. Conceptualizations

1. Reward vs. punishment
2. Instrumental conditioning
3. Learning
4. Results

C. Examples

1. Reward conditioning
2. Escape conditioning
3. Avoidance conditioning
4. Specific examples on graphs
5. Stimuli
6. Responses
7. Goals
8. Rewards
9. Aversive stimuli
10. Experimental situations
11. Negative reinforcement

D. Definitions

1. Reward conditioning
2. Avoidance conditioning
3. Escape conditioning
4. Instrumental conditioning
5. Reinforcement
6. Positive reinforcement
7. Negative reinforcement

E. Irrelevant Material

1. Should not dwell on escape conditioning

F. Sentence Form

Question 22 & 25: Compare and contrast escape conditioning and reward conditioning.

A. Comparisons and Contrasts

1. Reward vs. punishment
2. Both conditioning (instrumental)
3. Both learning
4. Different results from responding

B. Conceptualizations

1. Reward vs. punishment
2. Instrumental conditioning
3. Learning
4. Results

C. Examples

1. Reward conditioning
2. Escape conditioning
3. Avoidance conditioning
4. Specific examples on graphs
5. Stimuli
6. Responses
7. Goals
8. Rewards
9. Aversive stimuli
10. Experimental situations
11. Negative reinforcement

D. Definitions

1. Reward conditioning
2. Avoidance conditioning
3. Escape conditioning
4. Instrumental conditioning
5. Reinforcement
6. Positive reinforcement
7. Negative reinforcement

E. Irrelevant Material

1. Should not dwell on avoidance conditioning

F. Sentence Form

Question 23 & 26: Compare and contrast escape conditioning and avoidance conditioning.

A. Comparisons and Contrasts

1. Both deal with punishment
2. Both conditioning (instrumental)
3. Both learning
4. Different results from responding

B. Conceptualizations

1. Reward vs. punishment
2. Instrumental conditioning
3. Learning
4. Results

C. Examples

1. Reward conditioning
2. Escape conditioning
3. Avoidance conditioning
4. Specific examples on graphs
5. Stimuli
6. Responses
7. Goals
8. Rewards
9. Aversive stimuli
10. Experimental situations
11. Negative reinforcement

D. Definitions

1. Reward conditioning
2. Avoidance conditioning
3. Escape conditioning
4. Instrumental conditioning
5. Reinforcement
6. Positive reinforcement
7. Negative reinforcement

E. Irrelevant Material

1. Should not dwell on reward conditioning

F. Sentence Form

APPENDIX E

APPENDIX E

COMMENTS GIVEN TO TREATMENT GROUPS C AND D

Favorable

1. Good conceptualization of comparisons
2. Good comparisons
3. Good definitions
4. Good examples
5. Excluded irrelevants well
6. Good sentence form

Unfavorable

1. Write in sentence form
2. Exclude irrelevant material
3. Give examples
4. Define terms
5. Try to compare and contrast
6. Try to conceptualize your comparisons

APPENDIX F

APPENDIX F

PHASE III SCORES USED FOR STATISTICAL ANALYSIS

Time	Question	Subject	Group A	Group B	Group C	Group D
1	1	1	50.00	54.17	45.83	70.83
		2	45.83	45.83	54.17	66.67
		3	58.33	50.00	66.67	62.50
		4	58.33	41.67	45.83	62.50
	2	1	62.50	66.67	62.50	70.83
		2	58.33	62.50	54.17	70.83
		3	50.00	79.17	58.33	83.33
		4	75.00	70.83	41.67	87.50
	3	1	45.83	58.33	45.83	41.67
		2	37.50	70.83	45.83	58.33
		3	62.50	54.17	50.00	45.83
		4	54.17	33.33	66.67	83.33
2	1	1	75.00	50.00	58.33	87.50
		2	37.50	79.17	54.17	58.33
		3	75.00	70.83	58.33	70.83
		4	41.67	50.00	45.83	62.50
	2	1	58.33	58.33	50.00	54.17
		2	66.67	79.17	29.17	70.83
		3	58.33	75.00	54.17	50.00
		4	41.67	45.83	58.33	79.17
	3	1	50.00	70.83	66.67	45.83
		2	50.00	41.67	41.67	41.67
		3	54.17	62.50	62.50	45.83
		4	54.17	54.17	66.67	70.83

APPENDIX G

APPENDIX G

DUNCAN'S NEW MULTIPLE RANGE TEST

$$1. s_{\bar{x}} = \sqrt{(\text{error mean square})/r} = \sqrt{\frac{154.741}{24}} = 2.5392$$

2. Significant Studentized Ranges

alpha level \ no. of means in range	2	3	4
.05	2.82	2.97	3.07
.01	3.745	3.905	4.015

3. Least Significant Ranges

alpha level \ no. of means in range	2	3	4
.05	7.161	7.541	7.795
.01	9.509	9.916	10.195

APPENDIX H

APPENDIX H

RAW SCORES USED IN SPOT CHECKS OF RELIABILITY

1. By Questions:

Question	Grader	Subject																				ΣE	ΣX	ΣE ²	ΣX ²	ΣEX	Reliability
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20						
1	E	12	10	13	17	15	12	9	18	10	13	22	15	10	12	16	15	12	13	13	7	264	198	3,710	2,330	2,850	.82
	X	4	11	8	14	12	4	9	17	5	10	19	7	7	10	14	14	9	11	11	2						
2	E	11	10	10	16	14	17	10	13	12	11	7	8	12	9	14	10	13	10	13	16	236	173	2,924	1,731	2,185	.79
	X	9	4	4	10	12	12	9	8	8	7	4	4	12	9	11	6	10	5	12	17						
3	E	12	10	12	11	10	11	8	15	10	15	7	10	10	11	7	9	6	11	9	8	202	176	2,146	1,640	1,844	.68
	X	11	7	8	12	11	13	6	10	8	11	6	10	8	8	7	8	5	11	8	8						
4	E	13	13	10	14	10	16	11	10	13	10	14	12	14	6	10	13	6	14	16	9	233	194	2,879	1,966	2,361	.86
	X	10	8	7	11	8	13	8	8	11	10	11	10	11	8	9	12	5	12	13	9						
5	E	11	16	14	13	17	15	8	7	14	13	16	15	13	15	17	10	13	8	14	12	261	251	3,571	3,341	3,413	.77
	X	12	18	13	13	12	15	10	7	13	8	18	15	11	16	15	10	15	8	12	10						
6	E	12	13	14	11	8	15	13	6	4	9	13	9	15	13	11	8	8	10	10	8	210	227	2,378	2,737	2,514	.78
	X	12	14	15	12	9	18	14	8	10	9	10	9	13	16	11	7	11	10	11	8						
7	E	10	9	12	9	10	7	9	6	11	11	13	9	9	12	11	10	13	15	15	15	216	200	2,454	2,060	2,204	.51
	X	8	8	14	9	8	9	8	11	11	8	11	11	8	11	9	10	11	11	13	11						
8	E	13	5	4	7	8	11	15	11	13	7	16	16	16	6	10	14	5	15	10	11	213	192	2,579	1,924	2,149	.66
	X	12	7	4	7	9	11	10	10	10	6	10	12	11	11	11	10	10	10	10	10						
9	E	8	11	8	16	20	10	6	11	5	10	11	13	13	11	11	18	8	12	11	13	224	215	2,742	2,603	2,612	.78
	X	12	10	6	19	19	11	7	10	5	7	11	6	13	12	9	13	9	14	8	14						
10	E	15	5	10	16	8	6	13	14	16	10	14	16	17	10	12	6	9	16	10	16	240	240	3,158	3,189	3,116	.82
	X	18	6	12	15	8	6	11	14	14	7	9	17	18	12	10	12	9	16	9	17						
11	E	13	11	13	15	14	16	14	9	15	13	12	5	21	13	11	14	12	16	16	15	268	257	3,784	3,617	3,658	.87
	X	11	9	12	14	16	15	12	8	14	17	12	5	21	14	5	14	11	16	15	17						
12	E	10	12	13	12	17	14	9	14	9	19	14	11	17	12	12	11	15	16	15	12	264	253	3,626	3,349	3,444	.72
	X	9	14	7	11	16	13	10	16	9	16	16	10	15	12	10	13	16	13	14	13						
13	E	16	11	13	13	15	12	16	16	16	11	14	17	14	16	17	13	9	14	15	13	281	290	4,039	4,352	4,168	.81
	X	18	11	12	14	14	10	18	19	18	14	16	18	14	15	17	11	11	16	15	11						
14	E	19	18	13	15	16	13	13	10	14	13	17	20	15	12	10	11	15	16	15	16	291	282	4,379	3,530	3,832	.73
	X	19	14	14	10	15	9	18	13	14	12	16	19	8	9	3	4	13	15	13	16						
15	E	15	10	13	13	14	17	10	13	17	12	14	11	11	17	11	16	11	14	10	11	282	257	3,560	3,461	3,466	.70
	X	12	9	10	12	12	16	9	11	18	15	13	15	12	16	10	18	9	14	15	11						
16	E	15	15	11	10	16	14	15	18	18	13	14	12	17	17	16	11	15	16	15	21	303	257	4,743	3,479	3,998	.64
	X	12	14	10	11	18	13	12	15	12	13	13	14	15	13	14	7	11	7	13	20						
17	E	11	12	9	8	11	8	8	12	12	12	8	11	10	14	14	12	11	13	11	12	219	193	2,467	1,951	2,149	.46
	X	11	10	8	10	12	6	8	12	7	11	8	12	8	9	8	12	11	13	8	11						
18	E	9	12	9	16	10	17	12	20	17	15	13	10	11	15	6	12	19	15	10	19	258	228	3,590	2,736	3,093	.80
	X	7	12	10	13	11	15	12	19	12	11	9	10	10	10	8	10	15	12	12	10						
19	E	9	9	13	4	7	12	11	13	15	9	12	12	12	16	9	17	13	10	10	13	226	204	2,732	2,340	2,481	.82
	X	9	8	10	0	3	13	15	12	15	9	11	11	10	13	9	14	13	10	11	8						
20	E	12	17	17	13	10	15	13	15	7	4	12	16	12	9	10	12	6	15	16	6	237	191	3,097	2,045	2,464	.79
	X	13	9	14	11	8	14	11	11	6	4	10	9	10	7	6	8	8	15	14	3						
21	E	17	16	14	11	11	13	11	12	12	11	16	15	10	13	15	14	211	217	2,853	2,985	2,900	.70				
	X	16	17	15	13	11	14	13	13	15	12	15	14	12	12	13	12										
22	E	18	12	15	10	14	15	16	13	15	17	17	21	20	19	14	17	253	227	4,129	3,303	3,671	.79				
	X	14	11	13	12	12	15	14	13	12	14	15	16	19	17	12	18										
23	E	11	16	20	11	11	9	12	15	11	14	8	14	17	13	13	10	205	210	2,773	2,944	2,808	.71				
	X	11	13	18	8	13	6	15	14	13	16	12	18	19	11	11	12										
24	E	14	13	18	15	17	12	12	10	19	14	14	17	21	9	18	11	234	216	3,600	3,012	3,264	.80				
	X	15	12	17	17	15	12	12	11	13	12	12	14	17	10	17	10										
25	E	13	18	13	7	14	11	19	14	12	12	14	19	14	10	17	16	223	209	3,271	2,869	2,996	.55				
	X	13	12	12	13	11	7	18	10	13	12	15	18	16	11	17	11										
26	E	13	16	10	11	12	11	16	17	10	17	15	10	13	13	12	15	211	189	2,877	2,389	2,569	.62				
	X	12	13	10	6	13	15	15	16	11	15	10	5	15	10	10	13										
Total																						48245	5716	84,061	71,874	76,208	.74

2. By Treatment Groups:

Group	Grader	Scores																								ΣE	ΣX	ΣE^2	ΣX^2	ΣEX	Reliability											
A	E	11	10	10	16	14	17	10	13	12	11	7	8	12	12	10	12	11	10	11	8	15	10	15	8	7	16	11	15	10	9	11	13									
	X	9	4	4	10	12	12	9	8	8	7	4	4	12	11	7	8	12	11	13	6	10	8	11	8	6	12	8	11	8	9	10	14									
	E	15	9	4	10	6	14	13	8	16	15	12	9	13	9	13	10	10	9	16	15	5	19	15	9	8	11	14	12	10	12	13	16									
	X	13	8	10	11	8	11	11	8	18	10	10	9	10	8	14	12	9	8	15	14	5	16	14	10	9	10	16	10	7	10	13	13									
	E	20	14	14	12	16	13	15	16	11	15	17	14	14	14	15	13	11	15	18	16	11	10	17	15	10	17	11	12	8	12	9	10									
	X	19	13	14	10	7	12	12	15	12	13	16	13	13	16	8	16	9	11	18	18	10	9	16	14	9	15	12	13	10	11	9	11									
	E	16	6	9	12	11	10	12	19	11	12	17	13	12	15	11	14	14	11	12	18	9	15	14	13	14	12	12	18	16	9	13	18									
	X	13	8	8	12	11	11	10	15	12	10	9	9	15	12	12	12	12	8	11	14	6	14	15	11	16	13	10	17	11	10	12	17									
	E	10	14	13	10																																					
	X	11	15	15	11																														1,635	1,468	21,517	17,624	19,096	.71		
B	E	12	10	13	17	15	12	9	18	10	13	22	15	10	13	13	10	14	10	16	11	10	13	10	14	12	11	5	10	15	11	14	13									
	X	4	11	8	14	12	4	9	17	5	10	19	7	7	10	8	7	11	8	13	8	8	11	10	11	10	12	10	10	11	11	15	15									
	E	13	13	11	15	10	13	16	10	15	9	9	13	11	8	15	13	13	13	11	11	13	16	17	16	11	11	16	12	10	17	16										
	X	12	14	10	13	15	11	13	18	10	10	8	9	6	8	6	14	7	12	11	5	10	17	16	15	16	12	13	19	13	9	18	15									
	E	13	14	12	10	15	17	12	13	11	14	13	18	15	18	17	14	13	20	14	16	10	16	10	13	15	11	12	15	13	14	10	10									
	X	11	16	9	15	13	17	15	12	10	16	13	12	14	18	18	14	11	19	14	19	13	18	11	9	12	15	10	10	11	9	8	12									
	E	12	15	11	15	4	9	17	15	12	13	15	13	12	15	12	10	13	13	14	11	12	10	17	13	16	15	19	17	8	14	19	11									
	X	12	12	8	11	4	7	14	14	11	13	15	13	12	15	11	8	12	12	18	13	13	12	19	11	14	15	17	14	12	10	18	7									
	E	17	10	12	18	15	13	18	17	12																																
	X	15	11	12	12	10	10	13	14	12																										1,806	1,641	24,964	21,289	22,548	.67	
C	E	9	14	10	13	10	13	16	14	6	10	13	5	14	16	9	9	10	12	8	6	9	13	4	8	11	12	7	8	14	5	7	11									
	X	9	11	6	10	5	12	17	11	8	9	12	5	12	13	9	9	8	12	11	11	11	16	4	9	11	14	9	10	13	7	7	12									
	E	8	11	7	13	11	16	12	9	12	10	11	14	6	15	14	9	8	11	16	12	16	18	16	14	9	18	13	11	16	15	18										
	X	7	11	7	11	10	17	14	9	11	12	11	14	6	18	12	8	12	9	17	14	13	14	15	13	11	15	11	7	16	15	13	14									
	E	11	11	17	13	13	14	13	15	16	15	10	10	8	11	17	10	14	12	6	6	12	9	12	9	9	13	9	11	13	16	15	13									
	X	11	11	18	14	10	14	12	10	13	11	6	10	8	10	8	10	12	10	8	13	8	3	12	9	7	8	7	10	9	11	14	17	13	13							
	E	11	11	11	12	14	10	16	12	10	14	15	13	7	14	13	14	16	11	16																						
	X	13	13	13	15	12	12	13	13	5	12	13	12	13	12	12	11	15	10	13																	1,354	1,284	17,114	15,392	15,939	.74
D	E	12	16	15	12	13	13	7	7	10	10	11	7	9	6	11	9	8	10	16	13	8	11	13	10	17	12	15	13	15	16	14	15									
	X	10	14	14	9	11	11	2	6	10	8	8	7	8	5	11	8	8	10	10	8	9	9	11	10	12	11	16	10	15	11	12	18									
	E	6	17	13	16	15	12	10	16	5	6	6	11	15	12	10	14	12	21	14	5	17	13	14	11	15	17	15	12	16	11	17	21									
	X	11	15	14	15	16	12	7	14	5	7	12	8	17	11	11	16	12	21	9	6	16	14	12	14	15	13	12	14	16	4	16	20									
	E	10	11	13	19	12	11	10	13	7	8	7	13	12	9	17	12	16	12	16	4	8	20	17	17	16	15	10	17	20	21	15										
	X	3	15	11	19	11	11	10	11	6	6	3	8	8	10	15	10	14	9	10	14	0	6	19	18	16	15	14	12	15	19	16	13									
	E	14	11	20	21	13	11	10	14	11	17	17	17	15	12	19	14																									
	X	16	11	18	17	13	15	10	15	6	16	15	17	17	12	18	13																			1,450	1,323	20,466	17,569	18,625	.83	
Total																										6,245	5,716	84,061	71,874	76,208	.74											

VITA

David Murray DeFrain

Candidate for the Degree of

Master of Science

Thesis: THE EFFECTS OF INSTRUCTION, COMMENTS, AND PRACTICE UPON TEACHING STUDENTS HOW TO WRITE BETTER ANSWERS TO ESSAY QUESTIONS

Major Field: Psychology

Biographical:

Personal Data: Born in Harbor Beach, Michigan, November 16, 1943, the son of Murray A. and Vivian L. DeFrain.

Education: Attended grade school in Cass City, Michigan, and Holden, Missouri; graduated from Warrensburg High School, Warrensburg, Missouri in 1961; received the Associate in Arts degree from Graceland College, Lamoni, Iowa, with a major in Liberal Arts, in June, 1963; received the Bachelor of Arts degree from Central Missouri State College, Warrensburg, Missouri, with a major in Psychology, in August, 1965; completed requirements for the Master of Science degree in May, 1967.

Professional experience: Teaching assistantship in Reading, September, 1966, to the present; teaching Reading Improvement at Federal Aviation Authority, March, 1967, to the present.