

THE EFFECT OF TEST-MODE EXPECTANCY
ON COGNITIVE STRUCTURE

By

PHILIP M. HOLMES
"

Bachelor of Arts
Northeastern State University
Tahlequah, Oklahoma
1975

Master of Education
Northeastern State University
Tahlequah, Oklahoma
1978

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 1985

Thesis
1985D
H752e
cop. 2



THE EFFECT OF TEST-MODE EXPECTANCY
ON COGNITIVE STRUCTURE

Thesis Approved:

Kary T. Brun

Thesis Adviser

Robert J. Stammers

James M. Pica

Joseph P. Pica

Norman D. Mumken

Dean of the Graduate College

ACKNOWLEDGMENTS

I realize that I could not have completed an undertaking of this nature alone and I wish to express my sincere appreciation to some very special people for their support, assistance, and caring.

To Dr. Larry T. Brown, my mentor, my committee chairperson, and finally my friend; a "true" teacher who set an example that I can follow throughout my college teaching career.

I am also indebted to the other members of my committee. To Dr. Robert F. Stanners for his expert advise and guidance on multidimensional scaling procedures. To James M. Price for his expert advise on statistical procedures as well as for his patience and his explanations (and re-explanations) of statistical concepts. To Dr. Joseph Pearl for challenging group opinion and for providing valuable insights.

Several fellow graduate students are also due a special thank-you. To Rick Gowdy and Rick Frederick, who provided me with needed assistance in my "final hours." To David Johnsen and his wife Kathy, special friends who provided constant support and who made the whole experience even more worthwhile.

I cannot say enough concerning the support and assistance I received from my parents, Phil and Marjorie Holmes. Their support made all of this possible and their confidence kept me going.

To my children, Angie and Lindsay--two special people in my life who gave and continue to give me purpose.

Finally, to the woman in my life--my wife, Sally. We have shared and endured many difficult times together. I am sure that I will never fully appreciate all of the sacrifices that you have made. And, for what I am today, I owe so much to you.

TABLE OF CONTENTS

Chapter	Page
I. REVIEW OF RELEVANT LITERATURE.	1
II. METHOD	12
Subjects.	12
Materials	12
Procedure	14
III. RESULTS.	20
IV. DISCUSSION	25
REFERENCES.	31
APPENDIXES.	36
APPENDIX A - EXPERIMENTAL MULTIPLE-CHOICE TEST	37
APPENDIX B - EXPERIMENTAL ESSAY TEST	42
APPENDIX C - CONCEPT-COMPARISON RATING TASK.	50
APPENDIX D - SYLLABUS FOR CLASS M.	57
APPENDIX E - SYLLABUS FOR CLASS E.	62
APPENDIX F - RESULTS OF THE ANALYSES OF THE THREE- DIMENSIONAL SOLUTIONS	68

LIST OF FIGURES

Figure	Page
1. Testing Sequences for Class E and Class M	73
2. Class E Two-Dimensional Coordinate System	74
3. Class M Two-Dimensional Coordinate System	75
4. TA Reference Two-Dimensional Coordinate System.	76
5. Faculty Reference Two-Dimensional Coordinate System	77
6. Class E Three-Dimensional Coordinate System	78
7. Class M Three-Dimensional Coordinate System	79
8. TA Reference Three-Dimensional Coordinate System.	80
9. Faculty Reference Three-Dimensional Coordinate System	81

CHAPTER I

REVIEW OF RELEVANT LITERATURE

When evaluating students on their degree of learning, educators must first decide upon what kind of evaluation instrument to use. The basic problem is deciding between the use of the essay test or the objective test (e.g., multiple-choice, true-false, and matching). Both types of tests have their advantages and limitations (e.g., Mehrens & Lehmann, 1978, pp. 208-213). Regardless of any of the other advantages and limitations, the amount and nature of learning promoted by the type of test should be a major concern. Unfortunately, the research studies to date have provided few answers.

Students often show a concern for knowing what kind of test will be used to evaluate their learning. Is this concern idle curiosity or does knowledge of the test mode affect learning? Balch (1964) concluded that of the various factors related to testing, the student's expectation of the test mode appears to play a major role in learning.

Several studies have shown that students report differences in their preparation according to the type of examination expected (Class, 1935; Gustav, 1964; Silvey, 1951; Terry, 1934; White, 1932). Students say that they tend to memorize details for objective tests, whereas they attempt to understand general principles for essay tests (Gustav, 1964; Silvey, 1951; Terry, 1934). Students also report studying less thoroughly and for a shorter period of time for objective tests (Gustav,

1964). Silvey (1951) points out that while the students' beliefs and attitudes toward essay and objective examinations may not always be justified on the basis of fact or critical analysis, these beliefs and attitudes appear to be firmly implanted as if they had the support of conclusive evidence.

Rather than rely on student reports regarding differences in preparation, some studies have attempted to objectively measure possible differences in preparation (d'Ydewalle, Swerts, & de Corte, 1983; Hakstian, 1971; Meyer, 1935). D'Ydewalle et al. (1983) found that students expecting open questions on an exam used significantly more time to prepare than did students expecting multiple-choice questions. Meyer (1935) found that students expecting a multiple-choice test used mostly underlining of the reading material as their means of preparation, whereas the methods of notetaking and the writing of summaries were most used by students expecting an essay exam.

Hakstian (1971), however, found no difference in the manner of preparation for different test modes. If there are no differences in the study habits of students expecting different test modes, then it would seem logical to assume that there would be no difference between these students in their performance on any type of test. In addition to finding no difference in preparation, Hakstian (1971) also found no difference in performance between students anticipating either objective or essay examinations.

Other studies have also concluded that there is no difference in retention regardless of the type of examination for which the student has prepared (Kulhavy, Dyer, & Silver, 1975; Sax & Collet, 1968; Vallance, 1947). In the study by Sax and Collet (1968), two classes of an intro-

ductory college course were used. Class 1 received three essay tests during the semester and were told to expect the same type of final examination. Class 2 received three multiple-choice examinations during the semester and were told to expect the same type of final examination. The final examination was a combination multiple-choice and essay test. There were no differences in performance between the two classes on either the multiple-choice or the essay part of the final exam. Sax and Collet concluded that their results were probably due to the fact that the multiple-choice examinations were of the type that required an application of knowledge rather than simple recognition of a memorized response. They reasoned that their results might have supported the essay test had they used only multiple-choice items requiring simple recognition.

Several studies, however, have found that when students have prepared for a recall test--e.g., essay, short-answer, completion--they show better retention of the studied material than students who have prepared for a recognition test--e.g., multiple-choice, matching, true-false (d'Ydewalle et al., 1983; Gay, 1980; Meyer, 1936). Meyer (1936) investigated the expectation or set effects of four types of tests: multiple-choice, true-false, completion, and essay. He found that students who were expecting an essay examination performed better on all types of tests than students with multiple-choice, true-false, or completion test-mode expectations.

Gay (1980) used two treatment groups in her study; a multiple-choice (MC) group took six multiple-choice tests during the term of the course and a short-answer (SA) group took six short-answer tests during the term of the course. At the end of the term, both groups were given the same

final examination, which consisted of fifteen multiple-choice items and fifteen short-answer items. She found no significant difference between the two groups in their performance on the multiple-choice items. The SA group, however, did perform significantly better than the MC group on the short-answer items.

In addition to investigating test-mode expectancy effects upon the amount and nature of learning, some studies have examined the relationship between test-mode expectancy and length of retention (Hakstian, 1971; Meyer, 1934). Meyer (1934) found that the rate of forgetting for true-false and multiple-choice tests was greater than for completion and essay tests. Voss (1974) may offer one plausible explanation for such a finding. Based on his findings, Voss reasoned that the incorrect alternatives in a multiple-choice test may cause interference which in turn may be detrimental to long-term retention. Hakstian (1971) found no difference between the objective and essay examinations regarding the "breakdown" (forgetting) of material over time. This is not surprising since Hakstian also found no difference in student preparation for either type of test.

It is difficult at first glance to understand why the research on the effects of test-mode expectancy has produced such conflicting results. The number of variables that could possibly affect the results, however, are many: student motivation, classroom vs. nonclassroom environments, level of learning required (Bloom, 1956), and the student's prior experience with the test modes being investigated illustrate a few. No studies to date have made any serious attempts at examining possible reasons for the conflicting results in this area of investigation. The question of "Which type of test promotes 'better' learning?" remains

unanswered.

The studies discussed so far have been concerned with defining "better learning" in terms of efficiency (the ratio of output to input or of gain to effort). The studies that found no difference in performance regardless of the test mode anticipated or administered (Kulhavy et al., 1975; Sax & Collet, 1968; Vallance, 1947) would suggest that knowledge of test mode provides no advantage in terms of efficiency. If one considers these studies in relation to students' views regarding the difference between objective and essay examinations in terms of preparation effort (Class, 1935; Gustav, 1964; Silvey, 1951; Terry, 1934; White, 1932), then the objective test would be more efficient because of the lesser amount of preparation effort required.

On the other hand, although difficult to measure in terms of maximum efficiency, the studies of d'Ydewalle et al. (1983) and Meyer (1936) would support essay test preparation as providing maximum gain regardless of the type of test expected. For maximum efficiency, Gay's (1980) results would suggest that the student prepare for a known test mode and, in situations in which the test mode is unknown, essay test preparation would be the safest course of action.

The student's desire to know the mode of an examination prior to preparation would seem to indicate that the student is primarily interested in efficiency of learning. Ausubel (1963), however, contends that the goal of the educative process should be to maximize meaningful learning and retention. To Ausubel, meaningfully learned materials are "related to existing concepts in cognitive structure in ways making possible the understanding of various kinds of significant (e.g., derivative, correlative, qualifying) relationships" (p. 217). This understand-

ing of relationships has been referred to as internal connectedness, structural knowledge, and conceptual knowledge.

Early researchers promoting structural knowledge as an important goal for education included such individuals as Bruner (1960) and Ausubel (1963). There has been an increased level of interest in the structural aspect of knowledge in recent years. One reason for this may be the recent proposals for representing memory structure in the form of a propositional structure (e.g., Anderson, 1976; Kintsch & van Dijk, 1978; Norman & Rumelhart, 1975).

Propositions are abstract codes for the meaning of information acquired from written or spoken discourse. Information is stored as a network of interconnected propositions and, therefore, storage or retrieval of information requires an evaluation of the propositional interconnections. Propositions have a setlike structure: "People learn from experience which aspects or higher-order properties of an event are significant, and to represent these they develop a code, which is more direct and efficient than storing the details" (Anderson, 1983, p. 75). Memorization of material would be an example of "storing the details" which could be accomplished without coding the meaning of the material.

Quantitative attempts to measure and depict the nature of cognitive (or knowledge) structures have involved a variety of techniques. One technique that has received considerable attention is what has been termed the concept comparison task (Stanners, Brown, Price, & Holmes, 1983). This method requires subjects to make a judgment of the degree of similarity between concept labels presented in pairs. The comparative judgments are made in the form of a rating (e.g., the numbers from "1" to "7"). This approach has been used to rate concepts from short pieces of

fictional prose (Bisanz, LaPorte, Vesonder, & Voss, 1978; Laporte & Voss, 1979; Stanners, Price, & Painton, 1982), animal concept labels (Henley, 1969), concepts important for understanding experimental design and measurement scales (Fenker, 1975), concepts in the area of developmental psychology (Wainer & Kaye, 1974), in the area of personality theory (Stanners & Brown, 1982), and in learning (Brown & Stanners, 1983). Rating data from the concept comparison task are often analyzed by a multidimensional scaling (MDS) procedure to provide information concerning the structure of the interrelationships.

As MDS is composed of a variety of mathematical techniques for analyzing the structure of data, it is difficult to provide a single, comprehensive definition for it. The common purpose of MDS, however, is to represent whatever pattern or structure that may lie hidden in a matrix of empirical data in a form that makes the data much easier to comprehend--namely, as a geometrical model or picture (Shepard, 1972, p. 1). "When multidimensional scaling yields useful insights, these generally result from examining the configuration. One of the most important methods of examination is simply to look at the arrangement of points (e.g., concept labels)..." (Kruskal & Wish, 1978).

One important consideration for any method that may be used to assess structural knowledge is its validity. Does the concept comparison task provide a meaningful assessment of conceptual knowledge? The results of a number of studies would appear to support the concept comparison task as a valid means of assessing conceptual knowledge (Brown & Stanners, 1983; Fenker, 1975; Stanners & Brown, 1982; Wainer & Kaye, 1974).

Fenker (1975) used the concept comparison task to examine the ef-

fects of instruction on cognitive structure. His subjects were undergraduate students in an experimental design and measurement course. The stimuli used in the concept comparison task were basic concepts underlying the topic areas of experimental design and measurement. The students were administered the concept comparison task before and after a period involving standard classroom learning experiences (Fenker used the term "standard" to denote that no special attention was given to the particular concepts used for the study) and a MDS procedure was used to analyze the data. In order to assess the students' understanding of the material, their scaling results were compared to the optimal organization or formal structure. The formal structure was produced by scaling the judgments of an expert or reference group (faculty and graduate students with extensive experience in the topic area). The intervening instruction (i.e., standard classroom learning experiences) produced no significant effect on the post-intervention scaling solution and, therefore, no improvement in correspondence between the formal structure and the students' cognitive structure.

In a second experiment in Fenker's (1975) study, the students were given a list of relevant concepts and were encouraged to not only learn the concepts, but to also consider how related or associated the concepts were to each other. The results of this intervention were greater improvement in the correspondence between student and formal structure and the production of more meaningful and interpretable student cognitive maps.

The ability of the concept comparison task to distinguish between two groups of students based on their exposure to a topic area was evidenced by Stanners and Brown (1982). Using concepts from the area of

personality theory as the stimuli for the concept comparison task, Stanners and Brown found significant differences between the MDS solutions of the rating data for undergraduate students enrolled in an introductory psychology course and graduate students in psychology. An inspection of the undergraduates' MDS "map" produced from the rating data showed evidence of misconceptions or a lack of knowledge regarding the interrelationships of some of the concepts used. In addition, intersubject agreement of the concept interrelationships was significantly greater for the graduate student group.

If one of the goals of the education process is to promote the understanding of the meaning of concepts and their interrelationships, then knowledge of possible differences between essay and objective examinations in promoting this goal would be valuable. Many educators appear to have attitudes toward the objective and essay examinations similar to that of Kinney and Eurich (1932):

...the use of the subjective examination stimulates the pupil to study in order to acquire an organized body of information, and observe the relationships and implications of facts thus learned....pupils expecting to be tested with an objective examination are more apt to memorize unrelated facts without a consideration of their interrelationships. (p. 544)

As has been noted, most of the research concerned with possible differences brought about by different test-mode expectations has dealt only with the issue of test-mode differences in terms of learning efficiency. The major purpose of the present research was to investigate several possible differences brought about by different test-mode expectations (essay and multiple-choice). Three areas of possible

differences were examined.

The first area of possible differences examined was concerned with the issue of efficient learning. Would test performance differ between students who received the expected test mode and those who received the unexpected test mode? Research has yielded conflicting results related to this question (e.g., Gay, 1980; Sax & Collet, 1968). It would appear, however, that students do hold definite beliefs about how they prepare for objective and essay tests (Gustav, 1964; Silvey, 1951). Because of these beliefs and because students possibly spend more time preparing for an essay test than for an objective test (d'Ydewalle et al., 1983), the hypothesis for this area of investigation was that students expecting an essay test would perform better than those expecting a multiple-choice test, regardless of the test mode received.

The second area of possible differences examined was concerned with test-mode expectation and conceptual knowledge. Related to this area is the apparently common belief of educators that essay tests are better than objective tests at promoting the learning of material as an organized whole and at promoting the understanding of relationships within the organized whole (e.g., Mehrens & Lehmann, 1978, p. 210). The hypothesis in this case was that students expecting an essay test would show a closer correspondence in conceptual knowledge with an "expert" or reference group than would students expecting a multiple-choice test.

The final area of possible differences examined was concerned with differences in length of retention as a function of test-mode expectation. Meyer (1934) and Hakstian (1971) differed in their findings with to this issue. But, if students do prepare more thoroughly and for a longer period of time for essay tests than for objective tests

(d'Ydewalle et al., 1983; Gustav, 1964), this might favor essay test preparation in terms of promoting better retention of learned material over time. In addition, multiple-choice questions, because of their structure, may interfere with the retention of learned material (Voss, 1974). The hypothesis for this area of investigation was that students who expected and received an essay test would exhibit less change in correspondence of conceptual knowledge with an "expert" or reference group three weeks following the test than would students expecting and receiving a multiple-choice test.

CHAPTER II

METHOD

Subjects

The subjects were 114 undergraduate students from two classes of an introductory psychology course at a large Southwestern state university.

Materials

Three different modes of assessing learning were used in this study: multiple-choice, essay, and concept comparisons. A multiple-choice test consisting of 15 multiple-choice questions and a separate answer sheet constituted the multiple-choice assessment mode (see Appendix A). An essay test consisting of a cover/instructions page and six question/response pages made up the essay assessment mode (see Appendix B). The cover/instructions page provided a place for the student's name and class (section) number and included the instructions for answering the essay questions. The instructions asked the student to write neatly, to use complete sentences for their answers, and to respond to only one question per page. The question/response pages contained one question per page, leaving the remaining space on each page available for a written response to the question. The multiple-choice and essay tests were a mixture of knowledge, application, and analysis items (Bloom, 1956).

A concept-comparison rating task constituted the concept-comparison mode (see Appendix C). Twelve "concepts" from the area of personality theory were selected for the concept-comparison rating task. The con-

cepts were Freud, Skinner, Rogers, Unconscious, Free Will, Self, Learning, Instinct, External Influences, Early Experiences, Traits, and Testable. The concept-comparison rating form had each concept name paired with every other concept name, yielding 66 pairs. Three versions of the form were constructed by arranging the 66 pairs in three different orders. Each of the orders was random except that no concept name was allowed to be a member of more than two successive pairs. The right-left positions of the names in the pairings were balanced so that each name appeared in the right and left positions an equal number of times. To the left of each pairing was a blank for the student's rating. To provide a means of assessing the reliability of a student's ratings, 15 of the 66 pairs were randomly selected, repeated, and added to the end of the list of the 66 pairs, yielding a total of 81 pairs to be rated. In order to disguise the repetitions, the right-left positions of the 15 additional pairs were reversed and the pairs were separated from their original presentations by at least 27 items.

A set of instructions accompanied the rating form and asked the students to use the numbers "1" through "7" to rate the concept names of each pair on their "closeness of connection." The students were instructed to use a rating of "1" if a pair of concepts was judged to be "very closely connected" and to use a rating of "7" if a pair of concepts was judged to be "very unconnected." The students were told to use the numbers "2," "3," "4," "5," and "6" to indicate degrees of "connectedness" falling between the two extremes. The students were further instructed to base their decisions on their knowledge of the concepts and that just because a given pair of concepts was not treated as being "closely connected" in the textbook did not necessarily mean

that they were "very unconnected."

Procedure

The multiple-choice and essay modes of assessment were used in two "training" tests and all three modes were used in an "experimental" test. Each training test and the experimental test consisted of two parts. The first part of each training test was a 30-item multiple-choice test, and the second part was a 5-item essay test. The first part of the experimental test consisted of an 81-item concept-comparisons rating task, a 15-item multiple-choice test, or a 6-item essay test; the second part consisted of either 30 multiple-choice questions or 9 essay questions (see Figure 1 for the testing sequence).

Insert Figure 1

about here

During the first class meeting of the semester, a course syllabus was distributed to the students of two introductory psychology classes (see Appendixes D and E). The syllabus informed the students that they would be receiving five unit tests and a comprehensive final examination. Each unit test was worth 30 points and the final exam was worth 50 points. The student's grade for the course was based on the total number of points accumulated on all tests. The syllabus stated that the first two tests would consist of 30 multiple-choice questions and 5 essay questions (see Figure 1). It was explained that, although the answers to the essay questions would not contribute to any part of the student's grade, the questions should be answered "as though" they did. In addition, the students were told that, while their performance would not affect their

grade, the essay answers would be returned with a "score." The syllabi for the two classes gave a general education rationale for including essay questions on the tests. It was explained that the essay test was seen by many as an important tool for helping students to learn to organize and communicate learned material.

Information concerning the assessment mode of the third test was different for the two classes. One class (hereafter referred to as "Class M") was told that the essay questions would occur only on the first two tests and that the third test would have multiple-choice items, exclusively. The other class (hereafter referred to as "Class E") was told that their third test would consist of only essay items and that their score on this test would constitute their grade for Unit 3. Both classes were given the same reading assignments in their text, Psychology (Rathus, 1981), and both classes received the same lectures from the same instructor for the first two tests (training tests).

The first test was scheduled to follow two and one-half weeks of reading assignments and class lectures. One week before the first test each class was randomly divided into three groups of equal size. Each group was permanently assigned (for the term of the study) to different rooms for the test days. On the day of the test, the instructor administered the test in one of the assigned rooms and each of two assistants administered the test in one of the other rooms. Test items and test procedures were the same for both classes.

The multiple-choice part of the test was distributed first. The students were told that they would have 30 min to complete this part of the test and that if they finished before this time, they were to remain seated. At the end of the 30-min period, the multiple-choice part of

the test was collected and the essay part of the test was distributed. The students were told that they would have 20 min for this part of the test and that if they finished before this time, they were to remain seated. At the end of 20 min the essay part of the test was collected and the students were dismissed.

Before the essay tests were scored by the instructor, an assistant assigned a code number to each student's name and wrote this number on each page of the student's test. The assistant then removed all cover sheets, randomly ordered all essay tests (the essay tests from both classes were combined), and delivered them to the instructor.

One question was scored on all essay tests, the tests were randomly ordered, and then the next question was scored. The score for each question was written on each question's response page. This process continued until all questions were scored. The assistant then reattached the cover sheets.

The multiple-choice part of the test was computer graded and the scores were recorded and kept by the instructor. The results of the essay portion of the training test were not recorded nor did they constitute any part of the students' grades.

During the second class period following the test, the multiple-choice tests and results were distributed to the students and discussed by the instructor. Then the essay tests were handed back and the instructor discussed these items.

The second training test was administered to both classes two and one-half weeks following the first training test. The administration and scoring procedures for the second test were the same as for the first test.

After the second training test, there was a week and a half of reading assignments and standardized class demonstrations on personality theory. No lectures were given during this time. On the day of the test the students were told that they would be administered an experimental test prior to their "expected" test. They were told that the results of this test would constitute data for a research study. The students were informed that their participation was voluntary and that participation would be rewarded with extra-credit points. Only one student chose not to participate. Each of the three groups of each class was given information concerning the mode of their particular experimental test.

For the first part of the experimental test one group in each class received multiple-choice items, one group received essay items, and one group received the concept-comparisons rating task (see Figure 1). For the experimental multiple-choice and essay tests, the instructions were included on the tests. The instructor/assistant read the instructions aloud and then answered questions about the test. The students were informed that they would have 20 min to complete the test and that if they finished before this time, they were to remain seated.

For the rating task, the students were given a booklet containing the instructions and the list of concept pairs. The instructor read the instructions aloud and gave a brief demonstration (using names of animals) on how to perform the rating task. Questions concerning the task were then answered by the instructor. The students were allowed 3 min to complete 12 ratings. This pacing was used in an effort to discourage students from hurrying to complete the task without carefully considering each concept pair. The students were told not to work ahead and

that, if they finished their set of ratings before being told to go on to the next set, they could look over previous ratings and make any desired changes.

After the first part of the experimental test was collected students were administered their expected tests. The students were allowed the remainder of the class period to complete this test.

The procedure for grading the essay portion of the first part of the experimental test was the same as for the essay portion of the training tests except that the scores on each question were recorded next to student code numbers on a separate score sheet instead of on the test itself. After all essay tests were scored, 20 tests were randomly selected for rescoring. These 20 tests were rescored so as to provide an index of intrascorer reliability.

Three weeks after the test on personality theory, the students who received the kind of test they expected as the first part of the experimental test were administered the concept-comparisons rating task and the remaining students in each class were dismissed (see Figure 1).

The groups (one from each class) that performed the rating task as their experimental test will hereafter be referred to as the "immediate-rating groups." The groups performing the rating task three weeks after the experimental test will hereafter be referred to as the "post-rating groups."

The data analyzed for this study were the results of the rating data and the first part of the experimental test. The results of the second part of the experimental test composed a part of the students' grades.

A "reference group" of five graduate students in psychology (here-

after referred to as the "TA reference group"), all instructors in two classes of the general psychology course, were given the concept-comparisons rating task. Another reference group of three professors in psychology (hereafter referred to as the "faculty reference group"), all having taught the general psychology course a number of times as well as other undergraduate and graduate courses in psychology, were given the concept-comparisons rating task. Both reference groups were allowed to perform the rating task on their own time.

CHAPTER III

RESULTS

Intrascorer reliability on the experimental essay test was .98. The mean scores of Class E and Class M on the experimental essay test were 9.84 ($\underline{SD} = 3.55$) and 9.53 ($\underline{SD} = 3.15$), respectively. The difference between these means was not significant, $t(36) = .2901$. The mean scores of Class E and Class M on the experimental multiple-choice test were 10.42 ($\underline{SD} = 2.80$) and 10.89 ($\underline{SD} = 2.02$), respectively. The difference between these means was not significant, $t(36) = .5982$.

Before comparisons involving the paired-comparisons rating task were made the ratings were scaled by the COSPA MDS computer program (Schönemann, James, & Carter, 1979). The rating data from each of the six groups performing the rating task (two immediate-rating groups, two post-rating groups, and two reference groups) were scaled separately. The average reliability between the ratings of 15 concept pairs and the later repetition of the same 15 concept pairs was calculated for each student group. The average reliabilities were .491, .562, .455, and .501 for Class E and M immediate-rating groups and Class E and M post-rating groups, respectively. Fenker (1975) used low reliabilities as indicators of careless rating and discarded the data of subjects with low reliabilities. It has since been argued that low reliability may also be indicative of a low level of knowledge (Diekhoff, 1983). Data in the present study were not discarded on the basis of reliability.

The Schönemann et al. program is based on Horan's (1969) model and provides a \underline{v} -statistic for each individual. The \underline{v} -statistic is a measure of the proportion of variance in each individual's coordinate system that can be accounted for by the coordinate system derived for the entire group. Each \underline{v} -statistic can be used to test the hypothesis that there is a random relationship between the individual's data and the group's data. A \underline{v} -statistic falling in the upper decile of empirical norms developed by Schönemann et al. leads to a rejection of the randomness hypothesis.

A test of Horan's common-space assumption for the group is based on the number of individual tests for which the randomness hypothesis can be rejected. The critical number of rejections necessary to assume common space for the group can be determined by using standard binomial tables. If the common-space assumption is met, some intersubject consistency in the judgments is indicated. If, on the other hand, the common-space assumption is not met, then judgments between individuals are assumed to be independent and the group scaling results cannot be interpreted in a meaningful way.

The test of Horan's common-space assumption was made on the \underline{v} -statistics from the two-dimensional solutions of the immediate and post conditions of Class E and Class M. The test was significant in all cases, $\underline{p} < .01$. The ratio of students to the total in each group whose \underline{v} -statistics met or exceeded the .10 level were 12/19, 15/19, 8/19, and 12/19 for the Class E immediate-rating group, the Class M immediate-rating group, the Class E post-rating group, and the Class M post-rating group, respectively.

Both two- and three-dimensional solutions were analyzed because an

important feature of MDS is that it provides a graphical depiction (map) of the data, and two- and three-dimensional representations are much easier to depict than are higher-order dimensional representations. Furthermore, because the same lower dimensions are retained by the COSPA procedure when a higher dimensional solution is produced, higher-order solutions do not invalidate--but may supplement--the lower-order solutions. The outcomes of the analyses for the three-dimensional solutions (Horan's common-space assumption and the analysis of variance) were not different in any substantive way from those of the two-dimensional solutions. Therefore, the analyses of the two-dimensional solutions are reported here (outcomes of the analyses for the three-dimensional solutions are reported in Appendix F).

Information concerning possible differences between Class E and Class M in knowledge of interrelationships may be gained by examining and comparing the maps of the scaling results of the two immediate-rating groups (see Figures 2 and 3). There are some similarities between the two maps, particularly the represented relationships between the concept labels "Freud," "Skinner," and "Rogers." Differences between the depicted interrelationships of the other concept labels and these three individuals, however, are quite apparent.

Insert Figures 2 and 3

about here

The similarity between the maps of the TA reference group and the faculty reference group (see Figures 4 and 5) appears to be greater than the similarity between the maps of Class E and Class M. The most notice-

able difference between the two reference maps is the tighter and more distinct clustering of the concept labels in the faculty reference map. In a comparison of Class E (Figure 2) and Class M (Figure 3) maps with the reference maps (Figures 4 and 5), the map of Class M appears to be more similar to the maps of both reference groups.

Insert Figures 4 and 5

about here

The question of possible statistical differences between maps--and, therefore, conditions--was tested using a feature of the COSPA program which allows the substitution of an external coordinate system for the one derived from the data. The resulting χ^2 -statistics from the rescaled data sets then provide a measure of agreement with the coordinate system that is externally supplied. For comparison purposes, the larger the χ^2 -statistic for an individual, or mean χ^2 -statistic for a group, the higher the degree of agreement with the externally supplied coordinate system. In the present study the coordinate systems of the two reference groups were used for purposes of rescaling. The ratings of the graduate instructors of the introductory psychology course (TA reference group) provided one "expert" coordinate system and the ratings of the three faculty members (faculty reference group) provided a second (and possibly even "more expert") external coordinate system.

The rescaling, using the two "expert" reference coordinate systems yielded eight new sets of χ^2 -statistics (2 classes X 2 time conditions X 2 reference groups; $n = 19/\text{set}$). The χ^2 -statistics were then analyzed using a 2 X 2 X 2 analysis of variance. The expected-test-mode main

effect was significant, $F(1, 72) = 6.50$, $p < .05$. A comparison of the \underline{v} -statistics of Class E ($\underline{M} = .239$) with Class M ($\underline{M} = .292$) shows that the expected-test-mode main effect was in the opposite direction of that which was hypothesized. This means that the scaling solution of Class M has greater similarity--than that of Class E--to the coordinate systems of the reference groups (TA and faculty).

Neither of the other two main effects (reference groups nor time condition) was significant, and there were no significant interactions. The mean \underline{v} -statistic for the TA reference group condition was .266, and .265 for the faculty reference condition. The mean \underline{v} -statistics for the time conditions were .271 and .259 (immediate and post conditions, respectively). The mean \underline{v} -statistics for Class E immediate and post conditions and Class M immediate and post conditions were .245, .232, .298, and .286, respectively.

CHAPTER IV

DISCUSSION

This research investigated three areas of possible differences brought about by differences in test-mode expectation. One area of investigation was concerned with the issue of efficient learning. The issue of efficient learning actually involves two questions: (a) Is performance always better for students receiving the expected test mode than for students receiving the unexpected test mode? and (b) If not, is there one type of test-mode expectation that will lead to better performance regardless of the test mode received? Based upon the findings of d'Ydewalle et al. (1983) and Meyer (1936) it was hypothesized that students expecting an essay test would perform better than those expecting a multiple-choice test, regardless of the test mode received. As no difference in performance was found between students receiving the expected test mode and those receiving the unexpected test mode, this hypothesis was not supported.

A second area of investigation was concerned with test-mode expectation and conceptual knowledge. The apparently common belief among many educators (and students, as well) is that the objective test tends to promote the learning of isolated facts whereas the essay test promotes the learning of material as an organized whole and, in addition, promotes the learning of interrelationships within this organized whole. Or, to state this view in another way, essay tests are better than objective

tests at promoting conceptual knowledge—the understanding of concepts and their interrelationships. Based upon this common belief, it was hypothesized that students expecting an essay test would show closer correspondence in conceptual knowledge with an "expert" group than would students expecting a multiple-choice test.

An examination of the maps produced by the COSPA MDS procedure (Schönemann et al., 1979) from the concept-comparison task ratings of Class E and Class M immediate-rating groups and of the TA and faculty reference groups (see Figures 2, 3, 4, and 5, respectively) provides strong support for the analysis of variance results and vice versa. Both show that the expected test-mode main effect was in the opposite direction from that which was hypothesized.

The coordinate system of the Class E immediate-rating group appears to be quite different from that of the Class M immediate-rating group, and also quite different from the coordinate systems of the two reference groups. While all four maps (Figures 2, 3, 4, and 5) show a clear separation of "Freud," "Skinner," and "Rogers," Class E's map differs from the other three primarily in the location of "Traits" relative to the other concept labels (the map also differs in the greater dispersion of "Instinct," "Unconscious," and "Freud" relative to one another). The concept of "Traits" is not readily identifiable as belonging to any of the three theories represented—Freud, Skinner, and Rogers—and in the maps of Class M and the two reference groups it is more centrally located relative to the concept labels of "Freud," "Skinner," and "Rogers." Given the three basic theories represented, one might expect a concept label to be centrally located because it is perceived as being either equally related or equally unrelated to all three theories. "Early Ex-

periences is an example of another centrally located concept label that is related equally to all three theories.

The nonsignificant reference-groups main effect is reflected in the high degree of similarity between the coordinate systems of the two reference groups (see Figures 4 and 5). As was previously noted, the primary difference between the two reference maps is the tighter and more distinct clustering of the concept labels in the faculty reference map. The tendency for a "tighter" grouping of concepts to occur in the map of a group with the supposedly higher level of knowledge was also found by Stanners and Brown (1982).

The third and final area of investigation was concerned with test-mode expectancy and length of retention. It was hypothesized that three weeks following the test, students who had expected and had received the essay test would show lesser change in correspondence of conceptual knowledge with an expert group than would students who had expected and had received the multiple-choice test. This hypothesis was based upon Meyer's (1934) study regarding the rate of forgetting for various types of tests and upon the findings that students prepare more thoroughly and for a longer period of time for essay tests than for objective tests (d'Ydewalle et al., 1983; Gustav, 1964). The results of the analysis of variance do not support the hypothesis. The time condition main effect was not significant and, more directly related to the hypothesis, there was no significant class X time interaction.

The maps of Class E and Class M post-rating groups were almost identical to their respective counterparts in the immediate condition. The most apparent difference between the coordinate systems of the immediate and post conditions of both classes was the change in location of "Traits"

relative to the other concept labels. "Traits" may be considered an outlier with respect to its relationship to all other concept labels used in this study and this may be the reason for its change in location--relative to the other concept labels--over time. The change over time may be reflecting the students' lack of a well-formed idea regarding the interrelationship of "Traits" with the other eleven concept labels.

The results of this study would appear to support the use of the multiple-choice test over the essay test as the better test mode for promoting conceptual knowledge. The obvious question that should now be considered is, why would multiple-choice tests be better than essay tests at promoting conceptual knowledge? The answer may lie in the level of questions expected (Bloom, 1956) rather than in the type of test expected (or maybe in addition to the type of test expected). The level of question as a variable has not been closely examined by studies on the effects of test-mode expectancy, although it was briefly alluded to by Sax and Collet (1963).

It is generally considered more difficult to write "good" multiple-choice questions that measure learning at the analysis level or higher. Therefore the reported differences in student preference, as well as differences in preparation, for different test modes may be based on students' previous exposure to objective tests that contain all or most all lower-level questions. The present study did not control for question level, but it did use a mixture of knowledge-, application-, and analysis-level questions in both test modes. Both Class E and Class M were exposed to the same multiple-choice and essay items on the first two unit tests. It may be that both classes adjusted their preparation for

the third unit test (Class E for essay only and Class M for multiple-choice only) based upon their exposure to the previous two tests.

Are multiple-choice tests that contain application- and analysis-level (or higher) questions perceived as more difficult than essay tests containing the same level questions? Silvey (1951) found that students who preferred essay over objective tests gave their reason as being, "with essay tests, a little knowledge goes a long way" (p. 378). It may be that multiple-choice tests with higher-level questions would be considered more difficult. If this is the case, it may be due, in part, to the greater number of questions that can be included on a multiple-choice test as compared to an essay test.

A perceived difference in the difficulty between essay and objective tests may in itself lead to differences in the amount of learning promoted by each type of test. Students anticipating a "hard" exam show evidence of greater learning than students anticipating an "easy" exam (Sax & Reade, 1964). LaPorte and Nath (1976) found that hard goals (for learning) produced a significant increase in performance over easy goals and "do your best" goals. The question of whether the increase in performance was due only to more preparation time being used for hard goals was answered by keeping preparation time constant for the three groups (hard goals, easy goals, and do your best goals). As a result of this, LaPorte and Nath were able to conclude that the increase in performance of the hard goals group was due to something other than just the total amount of time in contact with the reading material and that the different goals produced differences in learning behaviors.

The relationship between the levels of questions on a test and the promotion of conceptual knowledge would appear to be a fairly obvious

one. The relationship between the levels of questions and type of test, however, needs to be considered. An investigation of the effects of test-mode expectancy, in which the levels of questions are controlled across test modes, might yield interesting and useful results.

REFERENCES

- Anderson, J. R. (1976). Language, memory, and thought. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.
- Ausubel, D. P. (1963). Cognitive structure and the facilitation of meaningful verbal learning. Journal of Teacher Education, 14, 217-222.
- Balch, J. (1964). The influence of the evaluating instrument on students' learning. American Educational Research Journal, 1, 169-182.
- Bisanz, G. L., LaPorte, R. E., Vesonder, G. T., & Voss, J. F. (1978). On the representation of prose: New dimensions. Journal of Verbal Learning and Verbal Behavior, 17, 337-357.
- Bloom, B. S. (Ed.) (1956). Taxonomy of educational objectives: Handbook I, cognitive domain. New York: David McKay.
- Brown, L. T., & Stanners, R. F. (1983). The assessment and modification of concept interrelationships. Journal of Experimental Education, 52, 11-21.
- Bruner, J. S. (1960). The process of education. Cambridge, MA: Harvard University Press.
- Class, E. C. (1935). The effect of the kind of test announcement on students' preparation. Journal of Educational Research, 28, 358-361.

- Diekhoff, G. M. (1983). Testing through relationship judgments. Journal of Educational Psychology, 75, 227-233.
- D'Ydewalle, G., Swerts, A., & DeCorte, E. (1983). Study time and test performance as a function of test expectations. Contemporary Educational Psychology, 8, 55-67.
- Fenker, R. M. (1975). The organization of conceptual materials: A methodology for measuring ideal and actual cognitive structures. Instructional Science, 4, 33-57.
- Gay, L. R. (1980). The comparative effects of multiple-choice versus short-answer tests on retention. Journal of Educational Measurement, 17, 45-50.
- Gustav, A. (1964). Students' preferences for test format in relation to their test scores. Journal of Psychology, 57, 159-164.
- Hakstian, A. R. (1971). The effects of the type of examination anticipated on test preparation and performance. Journal of Educational Research, 64, 319-324.
- Henley, N. M. (1969). A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behavior, 8, 176-184.
- Horan, C. B. (1969). Multidimensional scaling: Combining observations when individuals have different perceptual structures. Psychometrika, 34, 139-165.
- Kinney, L. B., & Eurich, A. C. (1932). A summary of investigations comparing different types of tests. School and Society, 36, 540-544.
- Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

- Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling. Sage University Paper series on Quantification Applications in the Social Sciences, 07-001. Beverly Hills, CA: Sage Publications.
- Kulhavy, R. W., Dyer, J. W., & Silver, L. (1975). The effects of note-taking and test expectancy on the learning of text material. Journal of Educational Research, 68, 363-365.
- LaPorte, R. E., & Nath, R. (1976). Role of performance goals in prose learning. Journal of Educational Psychology, 68, 260-264.
- LaPorte, R. E., & Voss, J. F. (1979). Prose representation: A multidimensional scaling approach. Multivariate Behavioral Research, 14, 39-56.
- Mehrens, W. A., & Lehmann, I. J. (1978). Measurement and evaluation in education and psychology. New York: Holt, Rinehart, and Winston.
- Meyer, G. (1934). An experimental study of the old and new types of examination: I. The effect of the examination set on memory. Journal of Educational Psychology, 25, 641-661.
- Meyer, G. (1935). An experimental study of the old and new types of examination: II. Methods of study. Journal of Educational Psychology, 26, 30-40.
- Meyer, G. (1936). The effect on recall and recognition of the examination set in classroom situations. Journal of Educational Psychology, 27, 81-99.
- Norman, D. A., & Rumelhart, D. E. (1975). Explorations in cognition. San Francisco: Freeman.
- Rathus, S. A. (1981). Psychology. New York: Holt, Rinehart, and Winston.

- Sax, G., & Collet, L. S. (1968). An empirical comparison of the effects of recall and multiple-choice tests on student achievement. Journal of Educational Measurement, 5, 169-173.
- Sax, G., & Reade, M. (1964). Achievement as a function of test difficulty level. American Educational Research Journal, 1, 22-25.
- Schonemann, P. H., James, W. L., & Carter, F. S. (1979). Statistical inference in multidimensional scaling: A method for fitting and testing Horan's model. In J. D. Lingoes, E. E. Roskam, & I. Borg (Eds.), Geometric representation of relational data. Ann Arbor, MI: Mathesis Press.
- Shepard, R. N. (1972). Introduction to volume I. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), Multidimensional scaling: Theory and applications in the behavioral sciences, volume I: Theory. New York: Seminar Press.
- Silvey, H. M. (1951). Student reaction to the objective and essay test. School and Society, 73, 377-378.
- Stanners, R. F., & Brown, L. T. (1982). Conceptual interrelationships based on learning in introductory psychology. Teaching of Psychology, 9, 74-77.
- Stanners, R. F., Brown, L. T., Price, J. M., & Holmes, M. (1983). Concept comparisons, essay examinations, and conceptual knowledge. Journal of Educational Psychology, 75, 857-864.
- Stanners, R. F., Price, J. M., & Painton, S. (1982). Interrelationships among text elements in fictional prose. Applied Psycholinguistics, 3, 95-107.
- Terry, P. W. (1934). How students study for three types of objective tests. Journal of Educational Research, 27, 333-343.

- Vallance, T. R. (1947). A comparison of essay and objective examinations as learning experiences. Journal of Educational Research, 41, 279-287.
- Voss, J. F. (1974). Acquisition and nonspecific transfer effects in prose learning as a function of question form. Journal of Educational Psychology, 66, 736-740.
- Wainer, H. & Kaye, K. (1974). Multidimensional scaling of concept learning in an introductory course. Journal of Educational Psychology, 66, 591-598.
- White, H. B. (1932). Testing as an aid to learning. Educational Administration and Supervision, 18, 41-46.

APPENDIXES

APPENDIX A

EXPERIMENTAL MULTIPLE-CHOICE TEST

Psychology 1113 - Experimental Multiple-Choice Test

Instructions: Choose the "best" or most correct answer for each of the following questions. Even though this test will not constitute any part of your course grade, you should provide a valid effort as it is required in order to receive the extra-credit points. You will have 15 minutes to answer these questions. If you finish early, remain seated and at the end of the time you will be asked to hand in this test.

1. According to the psychoanalytic theory, the _____ acts as a watchdog that screens unacceptable impulses.
 - a. homunculus
 - b. superego
 - c. id
 - d. ego
2. According to psychoanalytic theory, the id follows the
 - a. pleasure principle.
 - b. reality principle.
 - c. total objective.
 - d. unconscious mandate.
3. According to psychoanalytic theory, the Oedipus complex is characterized by
 - a. resentment toward the mother and attachment to the father.
 - b. resentment toward the father and attachment to the mother.
 - c. resentment toward both parents.
 - d. feelings of attachment to both parents.

4. Inadequate or excessive gratification in any stage can lead to _____ in that stage, according to psychoanalytic theory.
- a. fixation
 - b. resistance
 - c. identification
 - d. repression
5. Entry into the _____ stage allows children to divert their attention from sexual pursuits to schoolwork, according to psychoanalytic theory.
- a. phallic
 - b. latency
 - c. adult
 - d. genital
6. The psychoanalyst who noted that people tend to experience different sorts of crisis or challenges at different times during the adult years was
- a. Erik Erikson.
 - b. Erich Fromm.
 - c. Carl Jung.
 - d. Alfred Adler.
7. According to Sheldon's constitutional theory, the muscular, strong type is
- a. ectomorphic.
 - b. mesomorphic.
 - c. endomorphic.
 - d. allomorphic.

8. The ectomorphic type of individual, according to Sheldon's constitutional theory, is likely to be
- energetic.
 - fearful.
 - relaxed.
 - assertive.
9. Social learning theorists emphasize an ongoing interaction between _____ variables and _____ variables in the explanation and prediction of human behavior.
- situational; biological
 - cognitive; person
 - biological; cognitive
 - person; situational
10. Social learning theory may be criticized for its failure to
- pay sufficient attention to situational variables.
 - emphasize the importance of learning in human behavior.
 - attend to genetic differences among people.
 - engage in research in order to test its concepts.
11. A phrenologist assumes that traits, abilities, and mental functions are
- the result of unconscious processes.
 - located in specific places in the head.
 - the sum total of the activity of neural impulses.
 - determined by the environment as well as by person variables.

12. In order to determine the validity of a personality test, psychologists compare test scores with
 - a. scores earned on another occasion.
 - b. an external criterion.
 - c. scores earned by other individuals.
 - d. grades achieved in school.
13. The reliability of a personality measure is its
 - a. consistency.
 - b. ability to predict an external criterion.
 - c. fairness.
 - d. objectivity.
14. The _____ consists of several hundred items presented in a true-false format and is widely used to help diagnose abnormal behavior.
 - a. CPI
 - b. EPPS
 - c. MMPI
 - d. TAT
15. Hermann Rorschach is the originator of a personality measure that presents subjects with
 - a. vague drawings.
 - b. true-false items.
 - c. a forced-choice format.
 - d. inkblots.

APPENDIX B

EXPERIMENTAL ESSAY TEST

Psychology 1113 - Experimental Essay Test

Name: _____ Section: _____

Instructions: Answer the questions on the following pages. Even though this test will not constitute any part of your course grade, you should provide your best answers (a valid effort on your part is required in order for you to receive the extra-credit points). Limit the length of your answer to the amount of space provided. You will have 15 minutes to answer these questions. If you finish early, remain seated and at the end of the time you will be asked to hand in this test. The numbers in parentheses are the maximum number of points that can be earned on a particular question. Please write neatly and use complete sentences for your answers.

1. Contrast the id and the ego (name two differences). (3)

2. Describe the Oedipus complex and explain its relationship to identification. (3)

3. In what way are Erikson, Jung, and Adler alike? (2)

4. Name one of Sheldon's three basic types of people and describe the body build and temperament of the one you name. (3)

5. What is the one major difference between person variables and situational variables? (2)

6. Give the name of one of the projective personality tests discussed in your text. (2)

APPENDIX C

CONCEPT-COMPARISON RATING TASK

PERSONALITY RATING TASK

NAME:

SECTION:

Instructions: On the following pages of this booklet is a list of word pairs. All words in the list are related to the area of personality theory. Your task will be to make a judgment about the "closeness of connection" of the words in each pair and to give a rating based on your judgment. You are to use the numbers 1, 2, 3, 4, 5, 6, or 7 for your ratings. If you feel that the two aspects of personality theory are "very closely connected," then use a "1" to indicate this. If you feel that the two aspects of personality theory are "very unconnected," then use a "7" to indicate this. You should use the numbers "2," "3," "4," "5," and "6" to indicate degrees of connection that fall between the two extremes.

In many cases the two items of a pair may not have been treated in your textbook as being related to each other. It may be, however, that such items are closely connected. It is up to you to make this judgment based on what you know about a given pair of items.

You will be allowed 3 minutes to work on each set of 12 pairs. Do not start working on the next set until you are told to do so. If you finish a set before being told to go on, you may look over any previous items and make changes, if you wish. Do not worry if you have not finished a set at the time you are told to go on. It is not important that you keep up, only that you not work ahead.

You will be allowed a few minutes to scan the entire list of pairs before the task actually begins. The purpose of this is to give you an idea of how to use the scale.

(1 = "very connected" — 7 = "very unconnected")

- _____ 1. Learning and Early Experiences
 - _____ 2. Self and Early Experiences
 - _____ 3. Teastable and Learning
 - _____ 4. Skinner and Unconscious
 - _____ 5. Instinct and Unconscious
 - _____ 6. Instinct and Skinner
 - _____ 7. Rogers and Self
 - _____ 8. Free Will and Freud
 - _____ 9. Freud and Traits
 - _____ 10. Learning and Free Will
 - _____ 11. Testable and Instinct
 - _____ 12. Self and Free Will
-
- _____ 13. Rogers and Instinct
 - _____ 14. Learning and External Influences
 - _____ 15. Unconscious and Freud
 - _____ 16. Early Experiences and External Influences
 - _____ 17. Early Experiences and Unconscious
 - _____ 18. Traits and Skinner
 - _____ 19. Unconscious and Rogers
 - _____ 20. Freud and Testable
 - _____ 21. Self and External Influences
 - _____ 22. Early Experiences and Freud
 - _____ 23. Freud and Learning
 - _____ 24. Free Will and External Influences

(1 = "very connected" --- 7 = "very unconnected")

- _____ 25. Self and Freud
- _____ 26. Traits and Testable
- _____ 27. Free Will and Skinner
- _____ 28. Rogers and External Influences
- _____ 29. Skinner and Early Experiences
- _____ 30. External Influences and Instinct
- _____ 31. Self and Traits
- _____ 32. Instinct and Self
- _____ 33. Free Will and Testable
- _____ 34. Learning and Instinct
- _____ 35. Learning and Traits
- _____ 36. Early Experiences and Rogers

-
- _____ 37. Testable and Unconscious
 - _____ 38. Early Experiences and Traits
 - _____ 39. Testable and Self
 - _____ 40. Early Experiences and Free Will
 - _____ 41. Unconscious and Free Will
 - _____ 42. Unconscious and Self
 - _____ 43. Learning and Skinner
 - _____ 44. Unconscious and Learning
 - _____ 45. Skinner and Rogers
 - _____ 46. Testable and Rogers
 - _____ 47. Skinner and Testable
 - _____ 48. External Influences and Skinner

(1 = "very connected" --- 7 = "very unconnected")

- _____ 49. Traits and Unconscious
 - _____ 50. Self and Learning
 - _____ 51. Traits and Free Will
 - _____ 52. External Influences and Testable
 - _____ 53. Free Will and Instinct
 - _____ 54. Rogers and Learning
 - _____ 55. Freud and Instinct
 - _____ 56. Freud and Rogers
 - _____ 57. Testable and Early Experiences
 - _____ 58. Instinct and Traits
 - _____ 59. External Influences and Traits
 - _____ 60. Instinct and Early Experiences
-
- _____ 61. Traits and Rogers
 - _____ 62. Rogers and Free Will
 - _____ 63. Skinner and Freud
 - _____ 64. Skinner and Self
 - _____ 65. External Influences and Freud
 - _____ 66. Unconscious and External Influences
 - _____ 67. Rogers and Unconscious
 - _____ 68. Testable and Freud
 - _____ 69. External Influences and Self
 - _____ 70. Free Will and Early Experiences
 - _____ 71. Free Will and Unconscious
 - _____ 72. Self and Unconscious

(1 = "very connected" --- 7 = "very unconnected")

- _____ 73. Self and Rogers
- _____ 74. Freud and Free Will
- _____ 75. Traits and Freud
- _____ 76. Testable and Skinner
- _____ 77. Skinner and External Influences
- _____ 78. Unconscious and Traits
- _____ 79. Learning and Freud
- _____ 80. External Influences and Free Will
- _____ 81. Freud and Self

APPENDIX D

SYLLABUS FOR CLASS M

Syllabus for Psychology 1113

INTRODUCTORY PSYCHOLOGY

CLASS: Section 09, 12:30 MWF, AGH 275
INSTRUCTOR: P. M. Holmes
OFFICE: 010 South Murray Hall
HOURS: 11 - 12 & 1:30 - 2:30 MWF, or by appointment.

Introductory Psychology (Psych 1113) is designed to introduce the student to the basic principles, concepts, and technical language of the science of psychology. More specifically, the purposes of Psychology 1113 are:

- (a) to familiarize the student with the major concepts of psychology so that these may be used effectively as a part of his or her everyday vocabulary;
- (b) to motivate and prepare the student to continue the study of psychology, formally or informally, throughout his or her life;
- (c) to introduce the student to the various ways in which psychological facts and principles may be applied in everyday life;
- (d) to enrich the student's life by making him or her more alert to the complexity and diversity of psychological phenomena;
- (e) to show the student that psychology, rather than being a collection of common-sense opinions, mystical speculations, and far-out therapies, is in reality a rigorous but fascinating area of scientific inquiry; and
- (f) to directly expose the student to psychological phenomena through participation in classroom demonstrations.

STRUCTURE OF THE COURSE

The course is divided into six units. Each unit except the first will begin with a 30-minute discussion of the results of the previous unit's quiz; will follow with approximately 7 hours of lecture material, demonstrations, and discussion; and will end with a 30-point quiz (except for unit 6). The schedule is as follows:

Unit 1	Week 1	Chapter 1	Psychology and Human Behavior
	" 2	" 2	Biology and Behavior
	" 3	" 3	Sensation and Perception
			QUIZ 1 (January 27, Friday)
Unit 2	Week 4	Chapter 4	States of Consciousness
	" 5	" 5	Learning and Memory
	" 6	" 5	Learning and Memory
			QUIZ 2 (February 17, Friday)
Unit 3	Week 7	Chapter 9	Personality
	" 8	" 9	Personality
			QUIZ 3 (February 29, Wednesday)
Unit 4	Week 9	Chapter 6	Language and Intelligence
	" 10	" 7	Motivation and Emotion
	" 11	" 13	Sexual Behavior
			QUIZ 4 (March 30, Friday)
Unit 5	Week 12	Chapter 8	Developmental Psychology
	" 13	" 10	Stress and Adjustment
	" 14	" 11	Abnormal Behavior
	" 14	" 12	Psychotherapy
			QUIZ 5 (April 20, Friday)
Unit 6	Week 15	Chapter 14	Social Psychology
	" 16		***FINAL EXAM*** (May 2, 7 p.m.)

NOTE: Material from Chapter 14 will make up the first ten questions of the final exam.

READING ASSIGNMENTS

All reading assignments are in Psychology by S. A. Rathus (1981). The Study Guide accompanying the text is not required, but it is strongly recommended!

EXAM STRUCTURE

Five Quizzes and the final exam constitutes the course's only examinations. Each of the five quizzes will count 30 points and will consist of 20 multiple-choice items based on material in the textbook and 10 multiple-choice items based on lecture or textbook material. The final exam will contain 50 multiple-choice items based on reading assignments from all six units.

Recently there has been concern among a number of educators regarding the ability of the multiple-choice test to fully assess what students learn. For many years the essay test has been the exception rather than the rule; however, due to the recent concern about multiple-choice tests, this trend may reverse itself. For this reason, the first two unit tests will be in two parts. The first part will be the usual 30 multiple-choice items: the second part will be 5 essay items. The second part of the test will not count toward your course grade and will simply serve to provide you with essay-test experience. To maximize the benefit of the essay-test experience, the items will be "scored" and class time will allotted for a discussion of the correct answers. The essay-test experience is for your benefit and you should, therefore, do your best on these items.

Feedback on quiz performance will be provided at the second class meeting following each quiz. If you miss a question but feel your answer is in fact better than the "correct" answer, or if you feel a question is unfair, prepare a statement (about a paragraph in length) and submit it to me. If I am convinced that you have a good case, you will be awarded an extra point.

The numbers of points required for the different grade categories are given below. The numbers are based on a maximum total of 200 points (five 30-point quizzes plus a 50-point final exam).

<u>Final Grade</u>	<u>Points</u>
A	176 or more (88% correct)
B	156 - 175 (78% correct)
C	136 - 155 (68% correct)
D	116 - 135 (58% correct)
F	115 or less (below 58%)

The numbers of points on a 30-point quiz corresponding to the different grade categories are as follows:

A	--- 26.4
B	--- 23.4
C	--- 20.4
D	--- 17.4

Five (5) extra-credit points will be awarded to those students who do not miss any of the quizzes. Those who do have to miss quizzes must provide a statement signed by a medical officer, parent, or guardian in order to do make-up work for the quiz. These students will be allowed to earn the five extra-credit points by writing reports.

APPENDIX E
SYLLABUS FOR CLASS E

Syllabus for Psychology 1113

INTRODUCTORY PSYCHOLOGY

CLASS: Section 02, 2:30 MWF, HEW 236
INSTRUCTOR: P. M. Holmes
OFFICE: 010 South Murray Hall
HOURS: 11 - 12 & 1:30 - 2:30 MWF, or by appointment

Introductory Psychology (Psych 1113) is designed to introduce the student to the basic principles, concepts, and technical language of the science of psychology. More specifically, the purposes of Psychology 1113 are:

- (a) to familiarize the student with the major concepts of psychology so that these may be used effectively as a part of his or her everyday vocabulary;
- (b) to motivate and prepare the student to continue the study of psychology, formally or informally, throughout his or her life;
- (c) to introduce the student to the various ways in which psychological facts and principles may be applied in everyday life;
- (d) to enrich the student's life by making him or her more alert to the complexity and diversity of psychological phenomena;
- (e) to show the student that psychology, rather than being a collection of common-sense opinions, mystical speculations, and far-out therapies, is in reality a rigorous but fascinating area of scientific inquiry; and
- (f) to directly expose the student to psychological phenomena through participation in classroom demonstrations.

STRUCTURE OF THE COURSE

The course is divided into six units. Each unit except the first will begin with a 30-minute discussion of the results of the previous unit's quiz; will follow with approximately 7 hours of lecture material, demonstrations, and discussion; and will end with a 30-point quiz (except for unit 6). The schedule is as follows:

Unit 1	Week 1	Chapter 1	Psychology and Human Behavior
	" 2	" 2	Biology and Behavior
	" 3	" 3	Sensation and Perception
			QUIZ 1 (January 27, Friday)
Unit 2	Week 4	Chapter 4	States of Consciousness
	" 5	" 5	Learning and Memory
	" 6	" 5	Learning and Memory
			QUIZ 2 (February 17, Friday)
Unit 3	Week 7	Chapter 9	Personality
	" 8	" 9	Personality
			QUIZ 3 (February 29, Wednesday)
Unit 4	Week 9	Chapter 6	Language and Intelligence
	" 10	" 7	Motivation and Emotion
	" 11	" 13	Sexual Behavior
			QUIZ 4 (March 30, Friday)
Unit 5	Week 12	Chapter 8	Developmental Psychology
	" 13	" 10	Stress and Adjustment
	" 14	" 11	Abnormal Behavior
	" 14	" 12	Psychotherapy
			QUIZ 5 (April 20, Friday)
Unit 6	Week 15	Chapter 14	Social Psychology
	" 16		***FINAL EXAM*** (May 2, 7 p.m.)

NOTE: Material from Chapter 14 will make up the first ten questions of the final exam.

READING ASSIGNMENTS

All reading assignments are in Psychology by S. A. Rathus (1981). The Study Guide accompanying the text is not required, but it is strongly recommended!

EXAM STRUCTURE

Five quizzes and the final exam constitutes the course's only examination. Each of the five quizzes will count 30 points and four of the quizzes will consist of 20 multiple-choice items based on materials in the textbook and 10 multiple-choice items based on lecture or textbook material. The final exam will contain 50 multiple-choice items based on reading assignments for all six units.

Recently there has been concern among a number of educators regarding the ability of the multiple-choice test to fully assess what students learn. For many years the essay test has been the exception rather than the rule; however, due to the recent concern about multiple-choice tests, this trend may reverse itself. For this reason, the third unit test (Feb. 29) will be a 30-point essay quiz. To help prepare you for this essay test, the first two unit tests will be in two parts. The first part will be the usual 30-point multiple-choice quiz; the second part will be 5 essay items. The second part of the first two unit tests will not count toward your course grade and will simply serve to provide you with essay-test experience. To maximize the benefit of the essay-test experience, the items will be "scored" and class time will be allotted for a discussion of the correct answers. The essay items on

the first two tests are for your benefit and you should, therefore, do your best on these items. The 30-point essay test (Unit 3 Quiz) will count toward your grade.

Feedback on quiz performance will be provided at the second class meeting following each quiz. If you miss a question but feel your answer is in fact better than the "correct" answer, or if you feel a question is unfair, prepare a statement (about a paragraph in length) and submit it to me. If I am convinced that you have a good case, you will be awarded an extra point.

The numbers of points required for the different grade categories are given below. The numbers are based on a maximum total of 200 points (five 30-point quizzes plus a 50-point final exam).

<u>Final Grade</u>	<u>Points</u>
A	176 or more (88% correct)
B	156 - 175 (78% correct)
C	136 - 155 (68% correct)
D	116 - 135 (58% correct)
F	115 or less (below 58%)

The numbers of points on a 30-point quiz corresponding to the different grade categories are as follows:

- A — 26.4
- B — 23.4
- C — 20.4
- D — 17.4

Five (5) extra-credit points will be awarded to those students who do not miss any of the quizzes. Those who do have to miss quizzes must provide a statement signed by a medical officer, parent, or guardian in order to do make-up work for the quiz. These students will be allowed to earn the five extra-credit points by writing reports.

APPENDIX F

RESULTS OF THE ANALYSES OF THE
THREE-DIMENSIONAL SOLUTIONS

The test of Horan's common-space assumption was made for both the immediate and post conditions of Class E and Class M. The test was significant in all cases, $p < .01$. The ratio of students to the total in each group whose χ^2 -statistics met or exceeded the .10 level were 10/19, 13/19, 6/19, and 10/19 for the Class E immediate-rating group, the Class M immediate-rating group, the Class E post-rating group, and the Class M post-rating group, respectively.

A comparison of the three-dimensional maps of Class E and Class M immediate-rating groups and of the TA and faculty reference groups (see Figures 6, 7, 8, and 9) reveals some of the same basic information as provided by the two-dimensional maps. It should be noted that in the three-dimensional maps the plotted points are on the X, Z plane and that the vertical dimension (Y) is indicated by the length and direction of the line from the points on the X, Z plane. For example, in Figure 9, "Freud" is located in the negative X and negative Z quadrant of the plane, it is in the positive direction of the vertical dimension, and slightly lower than both "Unconscious" and "Instinct."

The three-dimensional maps are more difficult to interpret because of the greater dispersion of the concept labels. As previously noted, the same lower dimensions are retained by the COSPA procedure when a higher dimensional solution is produced. The added dimension will result in a greater dispersion of the concepts and this result may or may not provide additional information that is meaningful. A cluster type of approach was used to interpret the two-dimensional maps because labeling dimensions might be overly speculative at this stage of research. Only the faculty reference map appeared to retain the same clusters as were found in the two-dimensional maps. It may be that the faculty ref-

errence group, because of its possibly greater knowledge of the inter-relationships of the concepts, was able to judge individual pairwise presentations of the concept labels in such a manner as to yield a three-dimensional map that is a reasonable (and possibly more informative) extension of the two dimensional map.

Insert Figures 6, 7, 8, and 9

about here

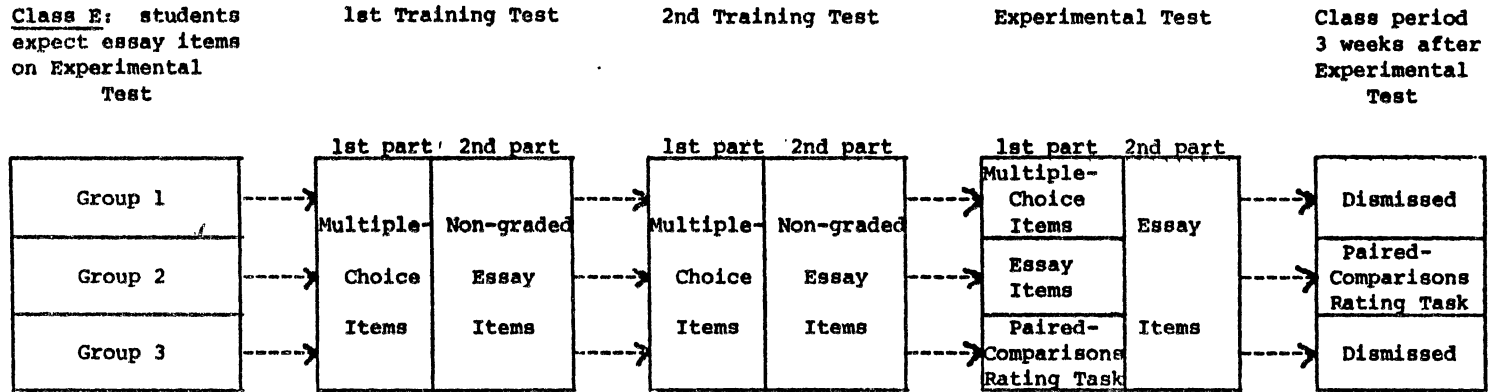
The results of the analysis of variance of the three-dimensional solutions were comparable to those of the two-dimensional solutions. As was the case for the two-dimensional analysis of variance, eight new sets of \underline{v} -statistics were created from a rescaling of the rating data of Class E and Class M (immediate and post conditions) using the two reference coordinate systems as the external coordinate systems. The expected-test-mode main effect was significant, $F(1,72) = 5.85$, $p < .05$. A comparison of the \underline{v} -statistics of Class E ($\underline{M} = .329$) with those of Class M ($M = .376$) show that the expected-test-mode effect was in the opposite direction of that which was hypothesized. It was hypothesized that students expecting an essay test (Class E) would show a closer correspondence in conceptual knowledge with the reference groups than would students expecting a multiple-choice test (Class M).

Neither of the other two main effects (reference groups and time condition) was significant and there were no significant interactions. The mean \underline{v} -statistic for the TA reference group was .358 and .347 for the faculty reference condition. The mean \underline{v} -statistics for the time

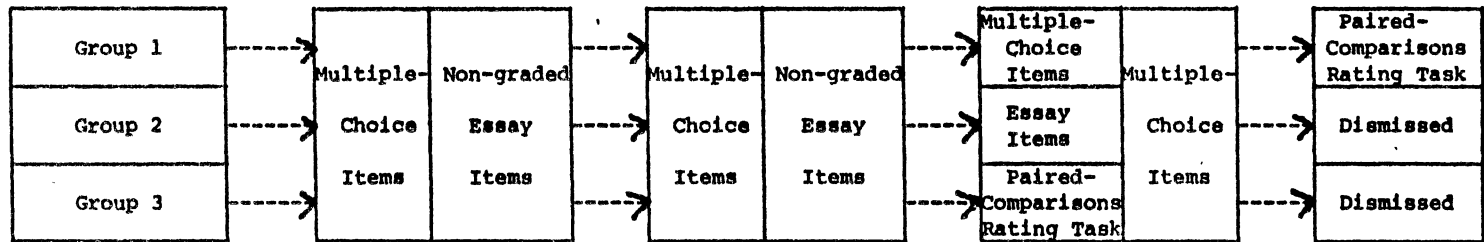
conditions were .358 and .347—immediate and post conditions, respectively. The mean \bar{y} -statistics for Class E immediate and post conditions and Class M immediate and post conditions were .333, .324, .382, and .371, respectively.

APPENDIXES

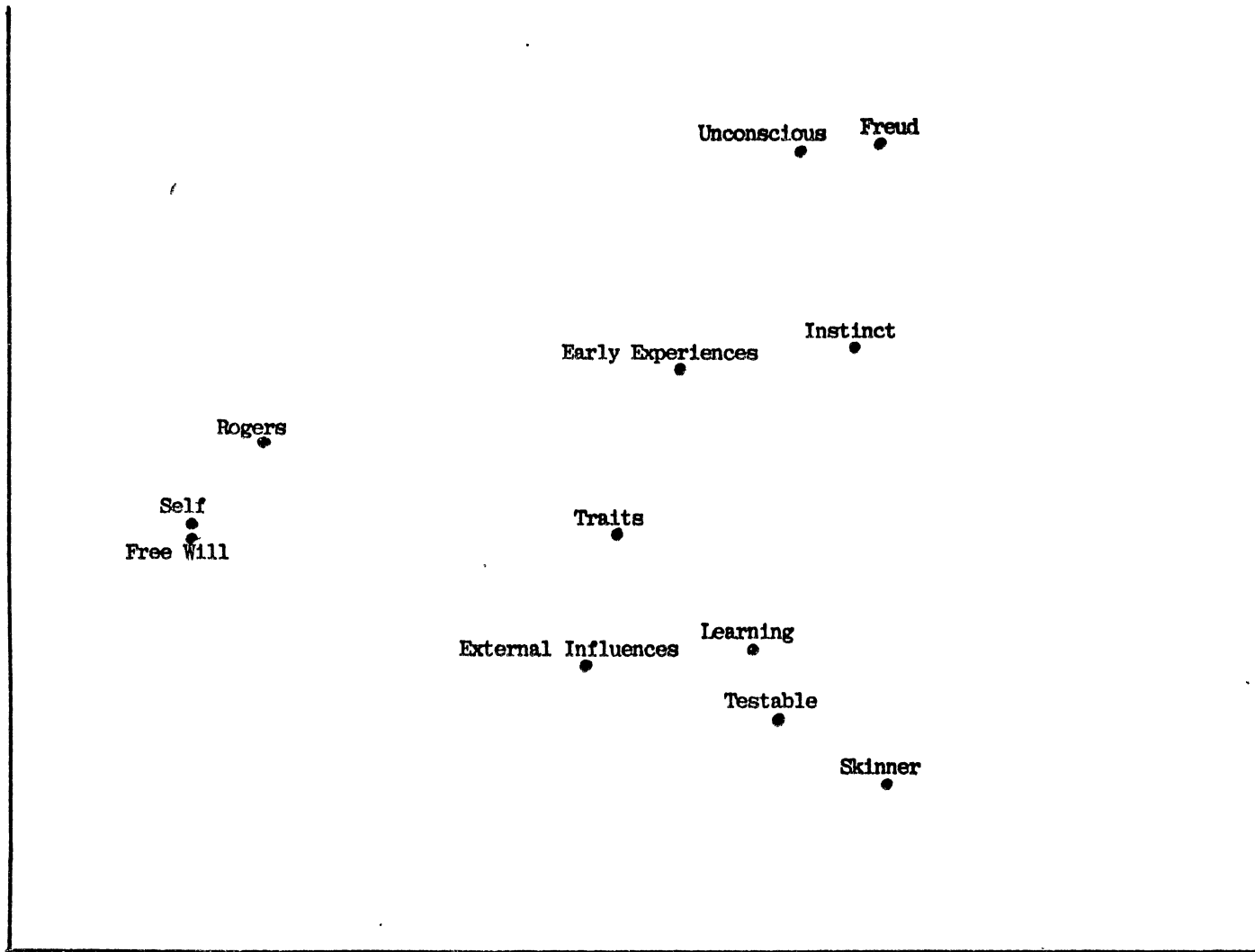
Class E: students expect essay items on Experimental Test

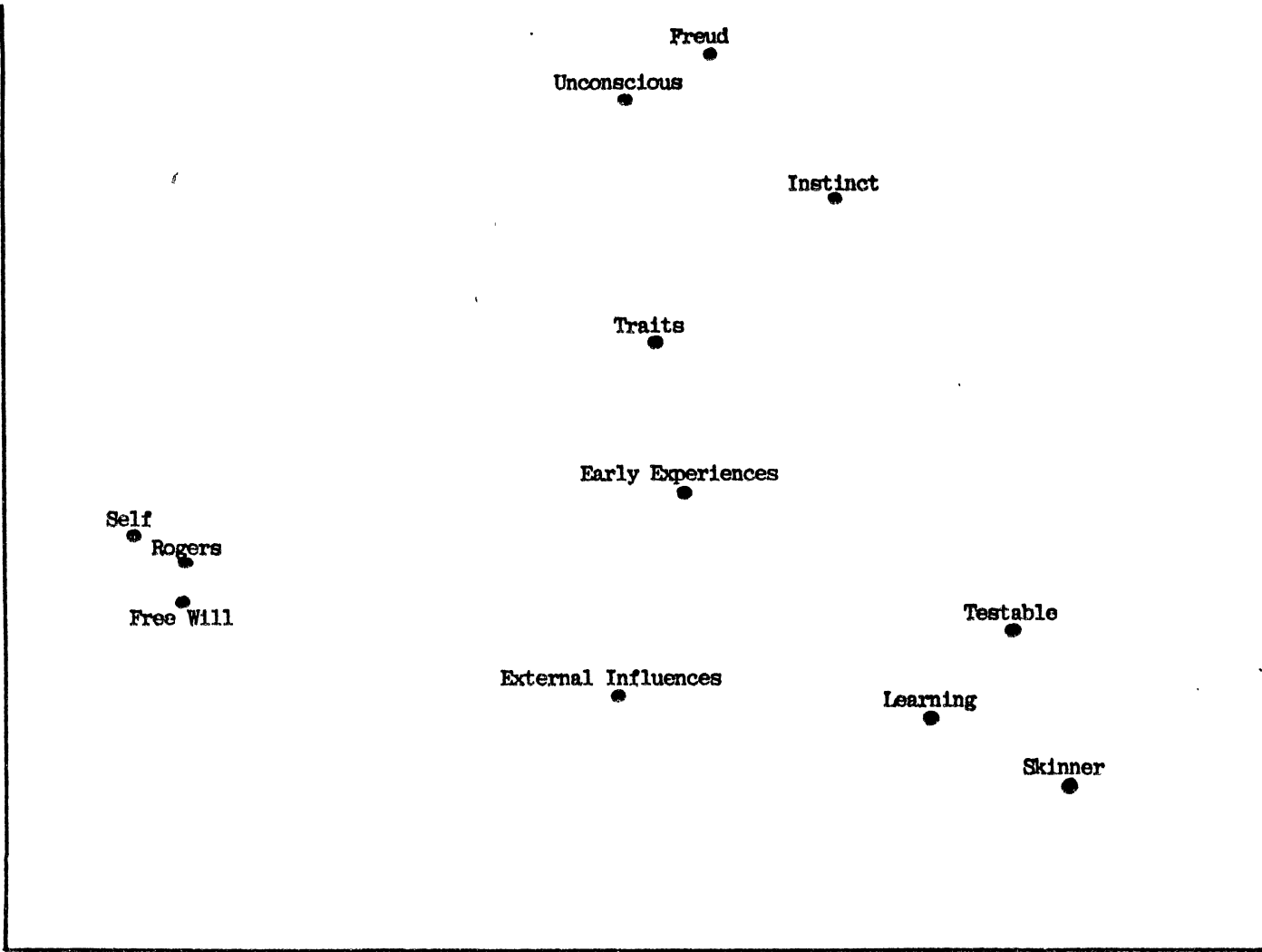


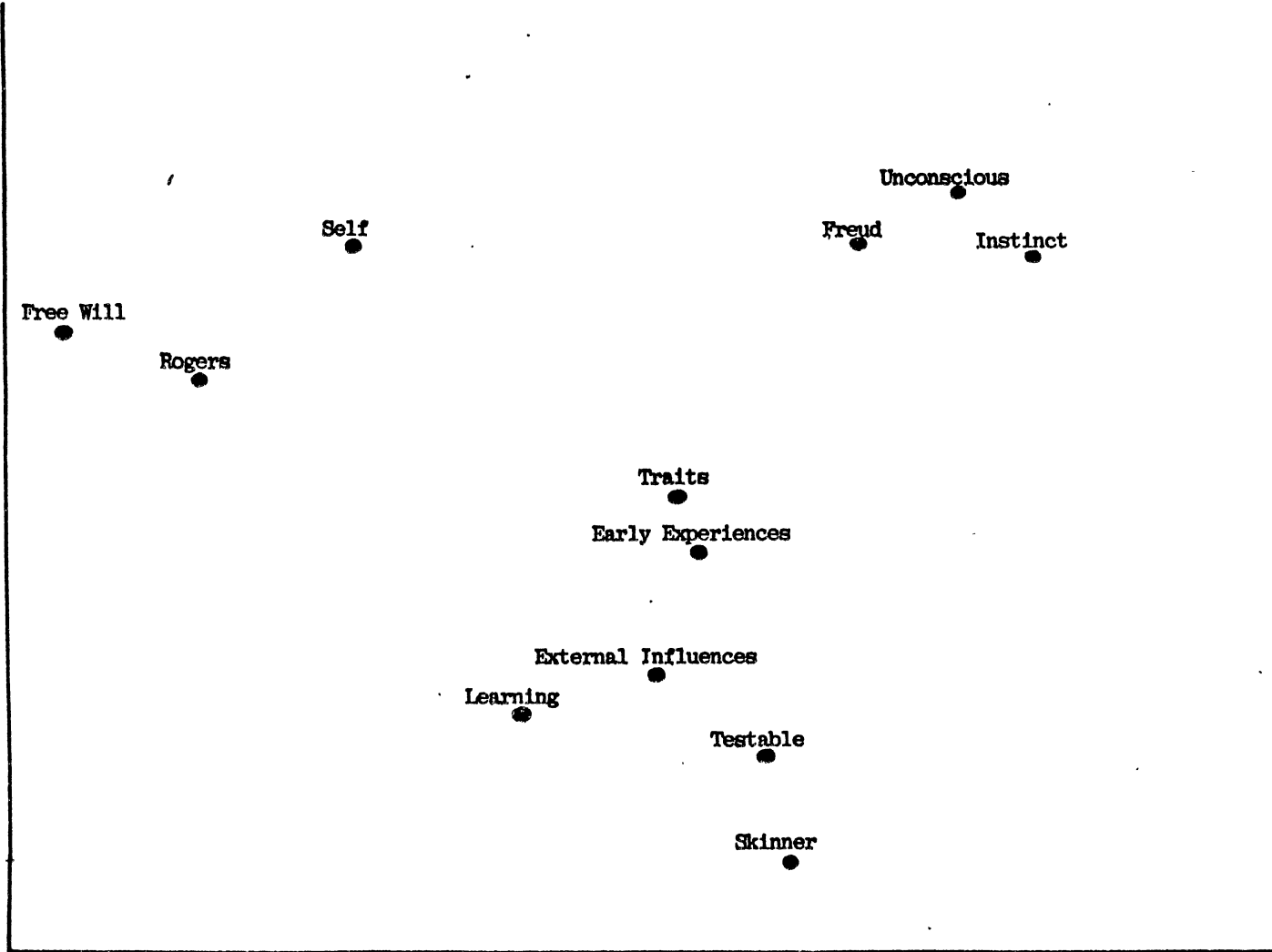
Class M: students expect multiple-choice items on Experimental Test

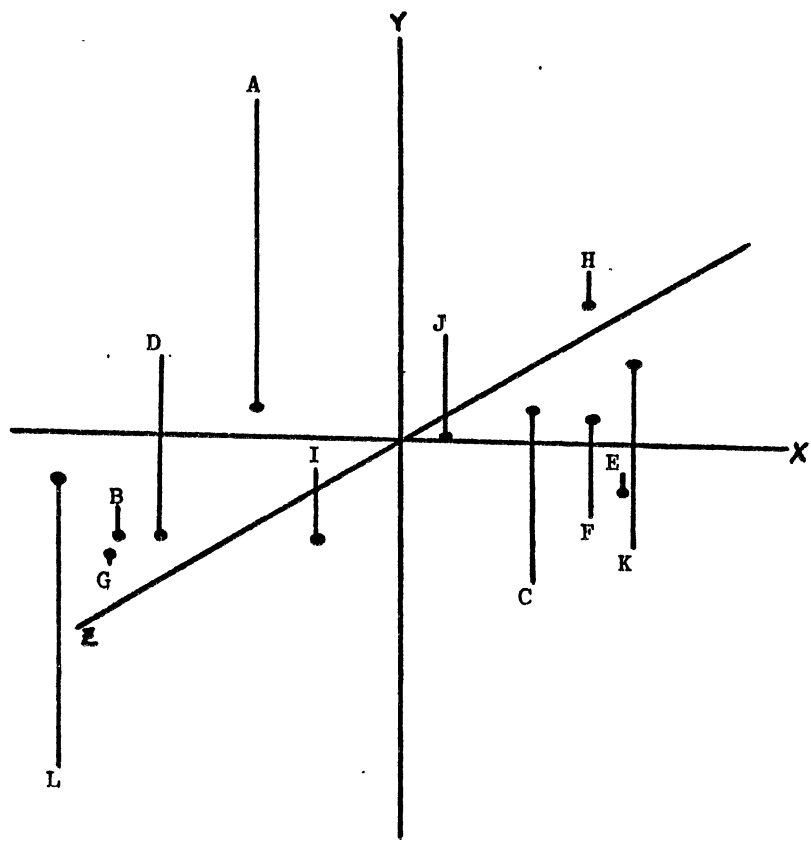




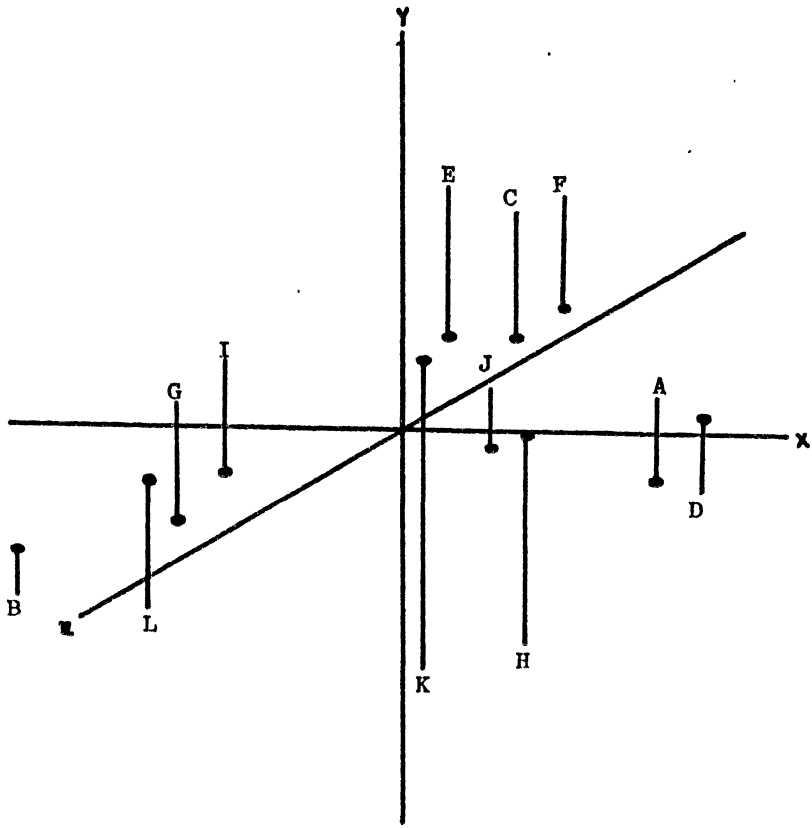




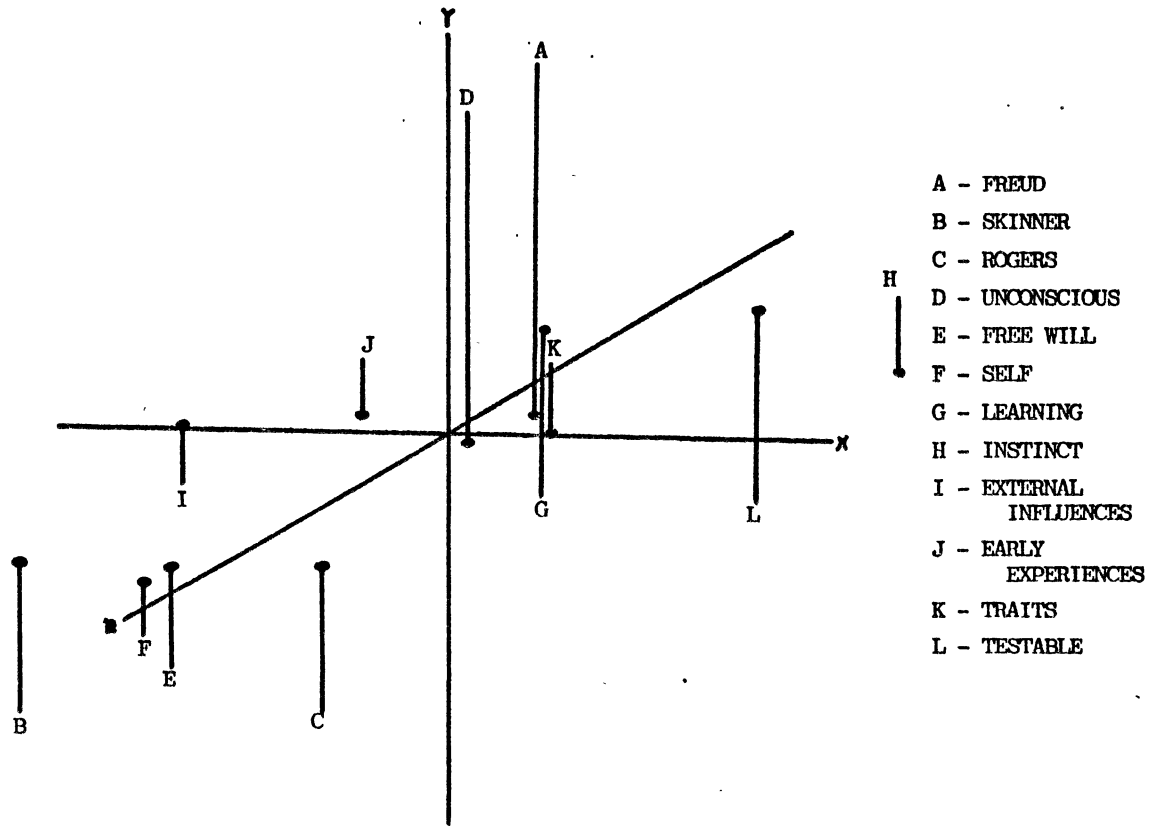


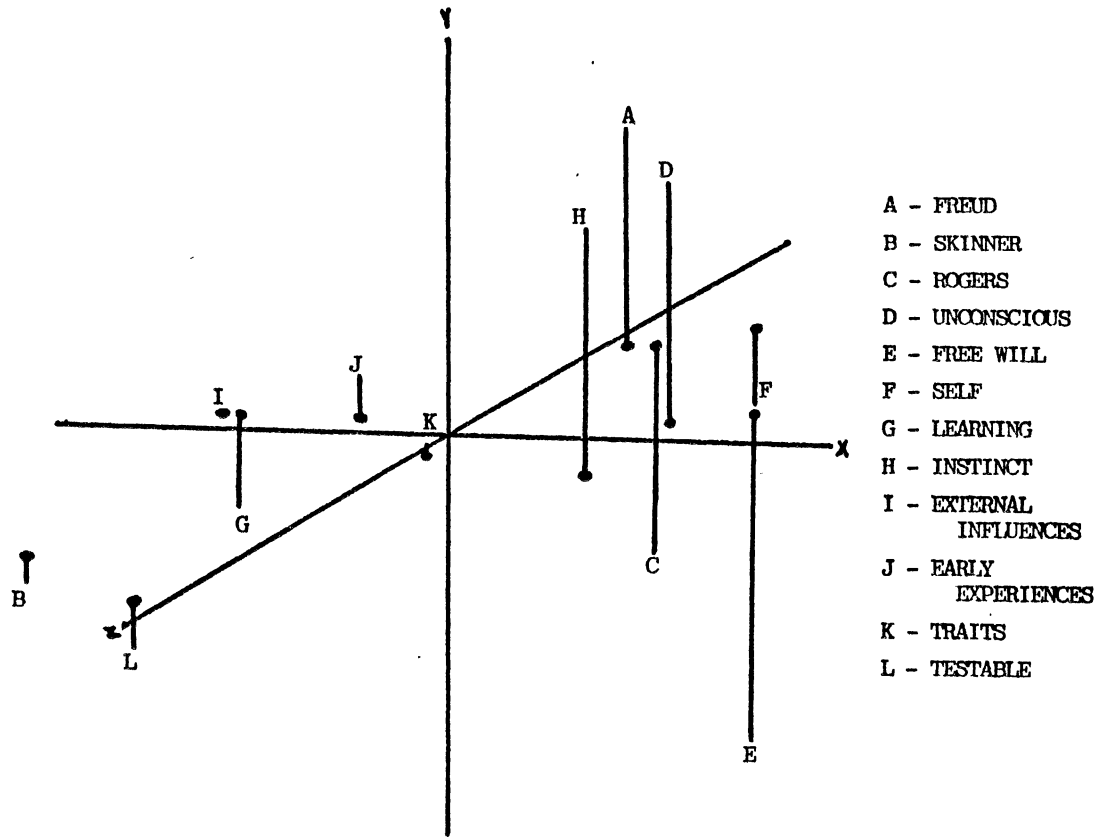


- A - FREUD
- B - SKINNER
- C - ROGERS
- D - UNCONSCIOUS
- E - FREE WILL
- F - SELF
- G - LEARNING
- H - INSTINCT
- I - EXTERNAL INFLUENCES
- J - EARLY EXPERIENCES
- K - TRAITS
- L - TESTABLE



- A - FREUD
- B - SKINNER
- C - ROGERS
- D - UNCONSCIOUS
- E - FREE WILL
- F - SELF
- G - LEARNING
- H - INSTINCT
- I - EXTERNAL INFLUENCES
- J - EARLY EXPERIENCES
- K - TRAITS
- L - TESTABLE





leaf 5

VITA 2

Philip M. Holmes

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE EFFECT OF TEST-MODE EXPECTANCY ON COGNITIVE STRUCTURE

Major Field: Psychology

Biographical:

Personal Data: Born in Durant, Oklahoma, October, 9, 1953, the son of Phil and Marjorie Holmes. Married to Sally J. Crane on March 24, 1978.

Education: Graduated from Broken Arrow High School, Broken Arrow, Oklahoma, in May, 1971; received Bachelor of Arts Degree in Psychology from Northeastern State University in May, 1975; received Master of Education degree from Northeastern State University in May, 1978; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1985.

Professional Experience: Teaching Assistant, Department of Psychology, Northeastern State University, August, 1976, to December, 1977; Lecturer, Department of Psychology, Northeastern State University, January, 1978 to May, 1978; Psychology Instructor, North Arkansas Community College, August, 1978 to July, 1981; Teaching Assistant, Department of Psychology, Oklahoma State University, August, 1981 to May, 1984; Assistant Professor, Department of Psychology, Doane College, August, 1984 to present.