

This dissertation has been
microfilmed exactly as received

70-2329

MUNZ, David C., 1943-
AN EVALUATION OF PERCEIVED ITEM-DIFFICULTY
SEQUENCING IN THE ACADEMIC SETTING.

The University of Oklahoma, Ph.D., 1969
Psychology, experimental

University Microfilms, Inc., Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

AN EVALUATION OF PERCEIVED ITEM-DIFFICULTY
SEQUENCING IN THE ACADEMIC SETTING

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
degree of
DOCTOR OF PHILOSOPHY

BY
DAVID C. MUNZ
Norman, Oklahoma
1969

AN EVALUATION OF PERCEIVED ITEM-DIFFICULTY
SEQUENCING IN THE ACADEMIC SETTING

APPROVED BY









DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

An endeavor such as this represents the effort and cooperation of several people. The writer is indebted to Dr. Paul D. Jacobs, committee chairman, for his invaluable advice and encouragement throughout the period of the doctoral program. Special acknowledgment goes to the dissertation committee, Dr. N. Jack Kanak, Dr. Thomas Miller, and Dr. Albert D. Smouse. Their critical comments of the manuscript were deeply appreciated. In addition, special acknowledgment goes to Mr. Terry Libkuman for his cooperation in subject procurement and Mr. John Baker for writing the computer program which greatly facilitated data analysis in the early stages of the investigation.

Special feelings are also reserved for my wife, Pat, whose emotional support and encouragement throughout my graduate training never faltered.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	v
LIST OF FIGURES.....	vii
Chapter	
I. INTRODUCTION.....	1
II. METHOD.....	23
III. RESULTS.....	34
IV. DISCUSSION.....	44
V. SUMMARY.....	51
REFERENCES.....	56
APPENDIX A.....	61
APPENDIX B.....	97
APPENDIX C.....	106
APPENDIX D.....	113
APPENDIX E.....	117
APPENDIX F.....	124
APPENDIX G.....	129

LIST OF TABLES

Table	Page
1. Intraclass Correlation Summary Table of the Average Intercorrelation of the Instructor Perceived Item-Difficulty Ratings.....	35
2. Means and Standard Deviations for the Median Intensity Ratings of the Students and Instructors.....	36
3. Analysis of Variance Summary Table of Achievement Performance Scores as a Function of Perceived Item-Difficulty Sequencing.....	36
4. Analysis of Variance Summary Table of PSI Scores as a Function of Perceived Item-Difficulty Sequencing.....	37
5. Analysis of Variance Summary Table of Test Difficulty Ratings as a Function of Perceived Item-Difficulty Sequencing.....	38
6. Analysis of Variance Summary Table of Test Fairness Ratings as a Function of Perceived Item-Difficulty Sequencing.....	39
7. Analysis of Variance Summary Table of Test Comprehensiveness Ratings as a Function of Perceived Item-Difficulty Sequencing.....	39
8. Analysis of Variance Summary Table of Performance Scores as a Function of Item Sequence and Personality Type.....	41
9. Means and Standard Deviations of All Cells Involved in the Analysis of Variance in Table 3.....	130
10. Means and Standard Deviations of All Cells Involved in the Analysis of Variance in Table 4.....	130

Table	Page
11. Means and Standard Deviations of All Cells Involved in the Analysis of Variance in Table 5.....	130
12. Means and Standard Deviations of All Cells Involved in the Analysis of Variance in Table 6.....	131
13. Means and Standard Deviations of All Cells Involved in the Analysis of Variance in Table 7.....	131
14. Means and Standard Deviations of All Cells Involved in the Analysis of Variance in Table 8.....	132

LIST OF FIGURES

Figure	Page
1. Mean performance scores for achievement-anxiety types (facilitators, high-affecteds, debilitators, and nonaffecteds) on perceived item-difficulty arrangements consisting of random (R), easy-to-hard (E-H), and hard-to-easy (H-E).....	42

AN EVALUATION OF PERCEIVED ITEM-DIFFICULTY
SEQUENCING IN THE ACADEMIC SETTING

CHAPTER I

INTRODUCTION

When a test constructor sets out to develop a measurement instrument he immediately is confronted with numerous technical problems. One of the most serious of these problems "is to arrange conditions so that the thing being measured is disturbed as little as possible in the act of measuring" (Ross, 1947, p. 117). Unfortunately, research has shown that test scores often are influenced by the examinee's mental set to react to test items in a particular manner (Cronbach, 1946). Such a response set is established in the test-taking situation by either task, situational, or organismic variables, or any of the possible interactions of these factors. Because such a response set leads to two examinees receiving different scores on the test when they have equally true scores on the content factor, the basic assumption underlying measurement theory (i.e., the content of the items comprising a test dictates what a test measures) is violated.

More specifically, a response set is "any tendency causing a person consistently to give different responses to test items than he would when the same content is presented in a different form" (Cronbach, 1946, p. 476). There is general agreement among test authorities (e.g., Cronbach, 1946, 1950; Loevinger, 1959) that this noncontent determined variance should be eliminated or controlled, especially in achievement testing situations.

The result of a response set is reflected in "response bias." The distinction between "response set" and "response bias" is an important one. Response sets are internal, organismic states which are inferred from specific response biases, or behavioral manifestations, evident in the outcome of test score statistics. In fact, "While some sets of an examinee may shift from test to test and even from item to item, others represent apparently enduring qualities that can be called personality traits" (Guilford, 1954, p. 453). Jackson and Messick (1958) have preferred to call these persistent qualities response styles and further suggest that "research involving response styles may contribute to a more systematic measurement in personality. . . ."

[p. 250]. The important point is that response sets, or styles, are organismic characteristics, whether they be transient or stable, and therefore are inferred from response biases reflected in total test score statistics

and individual item statistics.

Experimental evidence has been presented to demonstrate the existence of response sets in the testing situation. Cronbach (1946, 1950) and Guilford (1954) have discussed some of the more strongly verified response sets (e.g., toward gambling, impulsion, acquiescence, etc.), along with providing suggestions for counteracting their effects on test scores. The determinants, however, of response sets are not so easily ascertained. Guilford (1954) suggests that response sets tend to operate maximally in ambiguous and unstructured situations. He also suggests that difficult tests are conducive to response sets (see Stricker, 1965, for a recent empirical investigation of these two determinants).

Recent research has suggested a number of specific test construction characteristics and administrative procedures which are determinants of response set. For example, Aiken (1964) investigated the effects of difficulty context and item position on the tendency to leave items blank on multiple-choice vocabulary tests of varying difficulty levels. The results of this study supported a difficulty context effect but not the hypothesized item-position effect. Perloff and Wolins (1965), using a somewhat different procedure, have demonstrated the effects of item-difficulty context on individual item-difficulty values. Wevrick (1962) reported that positional

response sets can be induced in subjects by having the correct response occupy a particular position within a multiple-choice format. For example, if the third alternative is correct on a greater than chance number of occasions, Wevrick's results suggest that a response position bias unbeknownst to the examinee can have an influence on his total test score. Metfessel and Sax (1958) reported that of the fifteen well-known standardized tests of aptitude and achievement which they reviewed, one-third were keyed in such a way as to provide an opportunity for a positional response set. Mollenkopf (1950), investigated the effects of rearrangement of test items on item statistics under power and highly speeded conditions, found that the position of items in a speeded test had an undesirable influence on item statistics. He further reported that speeded tests tend to induce more guessing. Flaughner, Melton, and Myers (1968) found that moderate rearrangement of test items under slightly speeded conditions also influenced item statistics. Through the manipulation of three types of instruction, Yamamoto and Dizney (1965) demonstrated the effect of an instructional set on intelligence test scores.

It is evident from this research that response sets do exist and are induced by numerous test construction practices and administrative procedures. The fact that the score an individual receives on a test is

sometimes a function of the way in which the items are presented rather than the trait being measured must be reckoned with. While it is true that the view presented up to this point has been that response sets lead to sources of errors in the interpretation of test scores, it should be mentioned that in some cases this influence is a meaningful source of variance. In other words, when "response sets do occur, they may represent either an undesirable influence which needs to be eliminated or they may reflect a real and important dimension of human differences which is potentially useful for making predictions about behavior" (Helmstadter, 1964, p. 154). Broen and Wirt (1958) suggest that such effects should be eliminated:

. . . only in those cases where there is no correlation between the response set itself and the criterion for which prediction is being attempted. If there is a correlation between a response set and a criterion, then either the suppression of that set or the neglect to use tests in which that response set can operate may lead to lower validities [p. 239].

However, if a particular content area is to be assessed, as often is the case in achievement testing, it is necessary to eliminate the influence of response sets on the test scores. Methods have been developed for obtaining the separate "set" and "content" components of a test score (Helmstadter, 1957; Messick, 1961; Peabody, 1964; Webster, 1958), therefore allowing the test constructor

to capitalize upon the effect of response set when it is relevant to his needs and to eliminate it when it is undesirable.

Item-Difficulty Sequence as a Determinant
of a Response Set

After individual test items have been selected through the appropriate psychometric procedures, the test constructor must decide upon how to organize the items within the chosen test format. Standard practices in arranging test items in the area of achievement and aptitude testing have recommended that the items be arranged in order of increasing difficulty, that is, the easier items first followed by the progressively more difficult ones (Conrad, 1951; Greene & Jorgensen, 1939; Helmstadter, 1964; Jordan, 1953; Nunnally, 1959; Remmers & Gage, 1943; Ross, 1947; Thorndike & Hagen, 1955). The rationale underlying this prevalent practice provides various logical arguments for its utility. One such typical argument is the following concerning the arrangement of objective test items:

The items in the test should be arranged in ascending order of difficulty. It is especially important to have the easiest items at the beginning and the hardest ones at the end of the test. It will be recalled that one of the problems of measurement is to arrange conditions so that the thing being measured is disturbed as little as possible in the act of measuring. The psychological justification for placing the easiest items first is that such an arrangement has a

wholesome effect upon the morale of the pupils taking the test (Ross & Stanley, 1954, pp. 152-153).

This practice is not limited to objective examinations. In fact, Helmstadter (1964) suggests that this "requirement seems to be of less importance with objective tests than with essay tests. . . ." [p. 173]. Even though there is general agreement among authorities in the field that this test construction procedure is conducive to valid and reliable measurement, regardless of test format employed (i.e., multiple-choice, short-answer, or essay), there is no consistent agreement as to the rationale underlying the practice. Lund (1953) has summarized the various logical lines of reasoning into the following five major arguments:

1. Difficult items early in the test disrupt the intellectual functioning of the subject.
2. Encountering difficult items early in testing is likely to result in an unwise use of time and to thereby lower the performance score.
3. The less able subject is unable to cope with the items and will be discouraged. This reasoning implies that this principle applies only to the low ability subjects.
4. Normal arrangement of items is necessary to encourage naive subjects but sophisticated subjects would not be disconcerted by other arrangements.
5. The overall morale or motivation level of the subjects would be adversely affected by encountering difficult items early in the testing [pp. 7-8].

The proposed value of arranging items in a sequence from easy-to-hard (E-H) is apparent in these arguments.

Test constructors believe that such an arrangement eliminates adverse response sets which other item-difficulty orders, such as hard-to-easy (H-E) or random (R), would provoke. Consequently, a test construction practice which increases examinee motivation while eliminating such inimical mental states which lead to disrupted intellectual functioning, unwise use of time, and test-taking anxiety, seemingly would provide a more valid and reliable measure of the test content factor. From this position the establishment of such a response set by the E-H arrangement may be argued for as advantageous to both the examinee and examiner. While the advocated value of this test construction procedure is apparent, the recent empirical investigations in aptitude and achievement settings suggests a more complex interpretation of its utility.

In the aptitude setting there is some evidence to support the "superiority" (higher test scores) of the E-H arrangement of items over a subtest arrangement (Sax & Carr, 1962) and over an arrangement consisting of the hardest 25% of the items followed by a remaining 75% in an E-H order (Lund, 1953). Sax & Carr (1962) took a spiral omnibus¹ (E-H) aptitude test (Henmon-Nelson Test

¹Difficulty of a test item is based on the "p" index which is derived through item analysis. The "p" index is defined as the per cent of persons answering an item correctly (Nunnally, 1959). All of the following research discussed in this section employs the "p" index to generate the various item-difficulty arrangements.

of Mental Ability) and divided it into three subtests containing vocabulary, mathematics, and spatial relationship items, respectively. The results indicated that subjects gain significantly higher scores and attempted more items on the E-H format as compared with the subtest form. The authors hypothesized from these results that:

. . . the presence of increasingly complex items in a subtest tends to discourage students from responding to the more difficult items, and, conversely, the presence of different types of questions may provide some partial reinforcement and motivation to continue if the subject is able to respond correctly, let us say, to a vocabulary item rather than having to face the prospect of additional mathematics items when he has already had difficulty with a number of them (Sax & Carr, 1962, p. 374).

In light of these results, the authors concluded that:

Inasmuch as subjects attempted significantly more items and obtained significantly higher scores on the spiral-omnibus form than on the subtest form, evidence indicating the presence of a response set dependent upon test format and form of item presentation was indicated (Sax & Carr, 1962, p. 375).

Lund (1953) in a similar study employed two groups of 90 subjects and administered two forms of the Henmon-Nelson Test of Mental Ability. One form contained items arranged in the conventional E-H order. The other form contained the most difficult 25 per cent of the items first with the remaining 75 per cent in the E-H arrangement. The results indicated that the latter form (hard-to-conventional E-H) resulted in a significant lowering of performance for subjects at all levels of intelligence and for all levels of anxiety. The interpretation offered by Lund was that

subjects encountering difficult items early in the test do not have their ability to answer easy items significantly affected, but instead, the experience of failure early in the test resulted in a lowered motivational level that made the subjects reluctant to confront subsequent difficult items and, thereby, performance scores were significantly lowered.

While these studies (Lund, 1953; Sax & Carr, 1962) have found support for the superiority of the E-H arrangement over a subtest and hard-to-conventional E-H order, a recent study (Berger, Munz, Smouse, & Angelino, 1969) using E-H, H-E, and R orders of the Henmon-Nelson Test found no effects of item-difficulty sequence on aptitude test performance of high school seniors under two test settings (laboratory versus field).

In the achievement testing situation the effects of item-difficult sequencing on achievement test performance have been somewhat more consistent. In a series of four experiments, Brenner (1964) investigated the effect of various item-difficulty orders on total achievement test reliability, difficulty, and discrimination. The results obtained under typical college examination conditions suggested that "item-difficulty order on a power test of facts and principles given in the normal college classroom will not significantly affect these three test statistics" [p. 98]. In an attempt to detect and assess

the possible disruptive and/or adaptive effects of item-difficulty arrangements of test items on achievement test performance, Smouse and Munz (1968) assigned Ss final examinations consisting of items arranged either E-H, H-E, or at R. The Ss were tested under conditions of either high or normal anxiety and a situational anxiety measure, the MAACL (Zuckerman, 1960), was obtained on each S at the end of the examination. Consistent with Brenner's study (1964), no item-difficulty order effect was evidenced in total group test score measures under either test-taking condition. Furthermore, no effect of the three item-difficulty sequences on post test-taking anxiety was indicated by the results of the situational anxiety measure. The authors concluded:

. . . this research does not support the hypothesis that there is an item-difficulty sequence effect generally operating on the scores of power achievement tests typically found in the classroom. Further, within the limits of this study, it can be concluded that the two major arguments for arranging test items in an easy-to-hard sequence are not valid in any general way (Smouse & Munz, 1968, p. 183).

Speculating about their findings, the authors suggested that:

Another possible approach to the assessment of sequence effects would be the investigation of individual reactions to test-taking anxiety which would have been masked by the group measures used in this study. Such differential reactions might well interact with item difficulty sequence . . . (Smouse & Munz, 1968, p. 183).

In an attempt to investigate the assumption that

individual reaction to test taking mediates the effect of item-difficulty sequence on performance, Munz and Smouse (1968) administered college students a final examination with items sequenced either E-H, H-E, or at R. Based on prior responses to the Achievement Anxiety Test (Alpert & Haber, 1960), Ss were classified within each sequence as to achievement anxiety type. Analysis of the data confirmed their hypotheses, i.e., there was no significant item sequencing main effect upon performance scores; differential reactions to test-taking anxiety, as measured by the AAT, did significantly affect performance scores; and item difficulty sequence (E-H, H-E, & R) and achievement anxiety reaction type (facilitators, debilitators, nonaffecteds, and high-affecteds) did interact to produce a significant effect on achievement performance scores. On the R form the facilitators and high-affecteds scored significantly higher than the debilitators and non-affecteds, and on the E-H form, facilitators scored significantly higher than the other three anxiety types. There were no significant differences among the four anxiety types on the H-E form. These results were interpreted using the inverted-U hypothesis (Easterbrook, 1959; Fiske & Maddi, 1961; Malmo, 1959) and the assumption that item sequences are progressively more arousing in the order of R, E-H, and H-E. A follow-up analysis of these data and data eliminated by the original design further

suggested the influence of item-difficulty sequence as a determinant of differential response styles (Smouse & Munz, 1969). The re-analysis confirmed the author's hypothesis that organismic factors (as measured by the Achievement Anxiety Test) significantly and increasingly account for achievement test score variance under the following order of test formats: H-E, E-H, R. This finding suggests that the H-E sequence is the ideal format of the three to use when assessing academic achievement since measurement of the content factor of the test is less contaminated by test-taking personality variables. However, as the authors pointed out, further research is needed before such a generalization can be made, inasmuch as the H-E format may be provoking other differential response styles not assessed by their research design.

In summary, research in the area of achievement testing under actual field conditions suggests that the basic arguments espoused by test constructors for arranging items in an E-H order are not justified. There apparently is no group item-difficulty sequence effect on achievement performance scores. However, the fact that recent research suggests that item-difficulty sequencing does interact with test-taking personality variables to produce differential response-bias on achievement tests, which are assumed to measure only content determined factors, suggests the need for further systematic investigation

of the underlying causal mechanisms.

One area of needed exploration centers around the definition of a difficult test item. All of the research reported herein has employed the "p" index, or a similar derivative, of item-difficulty to generate the varying item-difficulty orders. The "p" index is the empirical probability (proportion of persons answering an item correctly in a tryout sample) that a particular population will pass an item (Guilford, 1954; Helmstadter, 1964). Unfortunately, the above definition reflects the psychometrist's concern with difficulty only in a statistical sense, that is, with the proportion of the sample who answered an item correctly (objective difficulty). The test constructor, however, also is "concerned with difficulty in a psychological sense (subjective difficulty)² as it affects the morale or behavior of the test taker" (Myers, 1962, p. 565). When advocating that item-difficulty sequencing affects test performance, regardless of the underlying rationale, it appears that the test constructor is concerned with the psychological sense of item-difficulty as it influences test performance. Can one assume that these two notions of difficulty are perfectly related?

Most test constructors would agree that the

²Writer's inclusion.

psychological meaning of a difficult item is a function of the subjective difficulty an individual experiences in reaching a solution to the particular question. Nevertheless, as Coombs (1950) pointed out, the test constructor still defines the difficulty of an item by the proportion of people passing it. Therefore, the difficulty of an item has been statistically represented with the same number for all the individuals in the sample even though we know that some individuals found the item easy and consequently passed it, while for others it was so difficult that they failed it (Coombs, 1950). Furthermore, the objective index disregards experiential variability within the group who passed the item and the group who failed it. This suggests that two items may have the same "p" value yet differ in average subjective difficulty and in variability of subjective difficulty around that average.

To complicate matters further, it is evident that subjective difficulty "may be affected by many different things such as ambiguity or complexity in the phrasing of the question, the reasonableness of the wrong alternatives, or the examinees' familiarity with the area of knowledge being sampled" (Myers, 1962, p. 566). Reese (1943) has suggested the following list of possible criteria for subjective difficulty:

1. Estimation of how difficult the task would be for a large number of people ("difficult" being defined objectively).

2. Estimate of the correctness of the observer's own answer.

3. Confidence in final answer after the task is completed.

4. Confidence that a correct reply will be obtained during the course of the reproduction.

5. Lack of familiarity with the type of task (the task may be subjectively difficult because the observer has never had any experience with the material).

6. Length of time to solve the problem.

7. Complication of the problem (i.e., the problem may be long and intricate but objectively easy).

8. Feelings of strain and effort.

9. Feelings of indecision [p. 72].

However, Reese has suggested that all of these "criteria" are closely related to each other. It appears from this discussion that more needs to be known about what makes an item difficult and possibly a better index would ensue.

Going back to the initial question, "Can one assume that these two notions of difficulty are perfectly related?", the above rational argument takes the position that the two concepts of difficulty are most likely not compatible. Empirical research, however, tends to support the conclusion that subjective difficulty roughly parallels objective difficulty (Farmer, 1928; Guilford & Cotzin, 1941; Hertzman, 1937; Reese, 1943). Yet, no attempt has been made to relate subjective difficulty and objective difficulty of multiple-choice test items derived under typical

achievement testing situations.

The importance of this question to item-difficulty sequencing as a determinant of a response set is paramount. Possibly the lack of evidence for a general item-difficulty sequence effect on test performance is due to the inability of the test constructor to actually sequence items in such a way that the induced degree of subjective difficulty (or lack of it) early in a test increases morale, increases test motivation, lowers test anxiety, or whatever the argument justifying the arrangement of items by difficulty. In other words, the "p" index may not be the best index of subjective difficulty. If an index were developed which more closely estimated the average subjective difficulty experienced by examinees when encountering an item and also allowed for the exclusion of items from the test which had large subjective-difficulty variability values, then the value of the E-H arrangement may be justified empirically and the causal mechanisms underlying a sequencing effect explored. As yet, no such index has been developed.

Statement of the Problem

The present study was an attempt to (1) develop a method for deriving perceived item-difficulty values of multiple-choice questions typically used in the college classroom, (2) compare student perceived item-difficulty values to instructor perceived item-difficulty values, (3) investigate systematically the effect of

perceived item-difficulty sequencing on academic examination performance, (4) assess possible underlying factors responsible for such an effect, (5) determine if the basic findings of the Munz and Smouse (1968) study are valid in situations where student "perceived difficulty" values are used to arrange the achievement test items, and (6) compare student perceived item-difficulty values to item-difficulty values derived from the typical "p" index.

No specific hypotheses were formulated concerning the development and employment of a method for deriving perceived item-difficulty values of multiple-choice questions. Of course, the argument previously stated concerning the drawbacks of the "p" index, if generally valid, would imply that there would be varying degrees of perceived item-difficulty variability with some items exhibiting more "efficient" estimates of the item-difficulty parameter than other items.

No specific hypotheses were formulated concerning the relationship between student perceived item-difficulty ratings and instructor perceived item-difficulty ratings. However, it would be desirable to find a high degree of relationship between the two sets of ratings. This would suggest that the results of this investigation concerning item-difficulty sequencing based on student perceived item-difficulty ratings could be extended to instructor ratings.

Furthermore, intercorrelations within the instructor ratings would allow for a comparison of the consistency of their ratings. A high degree of consistency among the instructor ratings would have the practical advantage of allowing the results of perceived item-difficulty sequencing to be generalized to sequences developed by the individual instructor. Of course, this generalization would be two steps removed and can be considered only in light of the ideal set of findings.

Concerning the effects of perceived item-difficulty sequencing on academic performance, it was hypothesized that perceived item-difficulty sequencing (E-H, H-E, and R arrangements) does significantly affect achievement testing scores, i.e., Ss receiving the E-H test perform significantly higher than Ss receiving either the R or H-E test. Furthermore, Ss receiving the H-E test perform significantly lower than the R group.

In an attempt to assess the possible underlying factors responsible for a perceived item-difficulty sequencing effect, two measures would be taken after the Ss had handed in their examinations. Measure I--the Perceived Stress Index (PSI: Jacobs & Munz, 1968)--would be employed to determine if item arrangement had an effect on test-taking anxiety. The index, a situational measure of experienced stress, would be used to measure how an individual generally felt while taking the examination.

In order to utilize the PSI, it would be necessary to alter the instructions of the "at this moment" scale so as to be able to assess a past affective state--"generally felt while taking the exam." It was hypothesized that perceived item difficulty sequencing significantly affects test anxiety, i.e., Ss examined with the E-H form score significantly lower on the PSI than the R and H-E groups. Furthermore, the H-E group score significantly higher on the PSI than the R group. The second measure--Test Evaluation Form (alteration of Burgess' Test Evaluation Form, 1968)--would be employed to assess the S's evaluation of the test after the examination. The form assesses (1) the perceived difficulty of the test, (2) the perceived fairness of the test, and (3) the perceived adequacy of coverage of the course. It was hypothesized that perceived item-difficulty sequencing significantly affects the student evaluation of the test, i.e., Ss examined with the E-H form score significantly higher on all three aspects of the evaluation form (high scores reflecting feelings of a fair test, an easy test, and a comprehensive test) as compared with the R and H-E groups. Furthermore, the H-E group score significantly lower on the evaluation form as compared with the R group.

This study also attempted to determine if the basic findings of the Munz-Smouse study (1968) are valid in situations where student-perceived difficulty values

are used to sequence achievement test items. In other words, does student perceived item-difficulty sequencing provoke differential response styles in the examinees which affect total test score performance and, therefore, mask out a general item sequence effort. The specific hypotheses stemming from the Munz-Smouse study (1968) are the following:

Hypothesis 1. Perceived item-difficulty sequencing of test items does not significantly affect achievement performance scores.

Hypothesis 2. Differential reactions to test-taking anxiety, as measured by the AAT [facilitators, those respondents scoring relatively high on the facilitating scale (AAT+) and relatively low on the debilitating scale (AAT-); debilitators, those respondents scoring relatively high on AAT- and relatively low on AAT+; non-affecteds, those respondents scoring relatively low on both AAT+ and AAT-; and high-affecteds, those respondents scoring relatively high on both AAT+ and AAT-], significantly affect performance scores, i.e., facilitators score significantly higher than debilitators and non-affecteds.

Hypothesis 3. Perceived item-difficulty sequencing (E-H, H-E, and R) and achievement anxiety reaction types (facilitators, debilitators, nonaffecteds, and high-affecteds) interact to produce a significant effect on

performance scores. Recent research (Sweeney, Smouse, Munz, & Rupiper, 1969) has suggested that, because of the nature of the underlying mechanisms hypothesized as mediating the interaction effect, specific subhypotheses should be made only about the nonaffecteds and the debilitators. These are: (1) debilitator's performance is not significantly affected by perceived item difficulty sequencing; (2) nonaffected's performance significantly improves across forms in the order R, E-H, and H-E.

No specific hypotheses were formulated regarding the relationship between student perceived item-difficulty values and item-difficulty values derived from the typical "p" index. The best inference, based on previous research, would be that there is a slight relationship ($\underline{r} = .50$; Hertzman, 1937).

CHAPTER II

METHOD

Subjects

Three groups of subjects were used in this study. One group consisted of 142 male and female students enrolled in two sections of a fall semester introductory psychology course taught at the University of Oklahoma by the same instructor. Another group of subjects consisted of nine introductory psychology instructors each of whom had just completed teaching two sections of an introductory psychology course in the fall semester (1968). Included in these nine instructors was the instructor whose two classes were used to comprise the above mentioned group of introductory psychology subjects. The third group of subjects consisted of 133 male and female students enrolled in two sections of a spring semester introductory psychology course also taught by the same instructor who had previously taught the initial group of introductory psychology subjects.

Item-Difficulty Scaling

In an attempt to develop an index of subjective item-difficulty which more closely approximates the

average subjective difficulty experienced by the examinees for each item, perceived item-difficulty values of an item pool of multiple-choice questions typically used in the introductory psychology course examination were derived through the use of Thurstone's method of equal-appearing intervals (Thurstone, 1928; Thurstone & Chave, 1929). This specific technique was chosen because it provides an absolute index of subjective difficulty along with an index of interjudge variability for each item.

An initial item pool was developed through the use of the instructor's manual (Morgan & King, 1967) for Introduction to Psychology (Morgan & King, 1966). This manual contains 20 multiple-choice questions (four alternatives) for each chapter in the textbook. These items closely approximate the typical item generated by the introductory psychology instructor in that they have not been subjected to item-analysis and were initially developed by the authors for use in their introductory psychology course examinations. From this item pool approximately 15 items were selected to represent each of the 14 chapters taught in the one semester introductory course. This provided a total of 210 multiple-choice items. Because of the time factor in scaling 210 multiple-choice questions, two booklets of test items (Form A and Form B) each containing 105 multiple-choice questions were randomly generated (see Appendix A).

A sample of 142 introductory students were employed for scaling purposes. In order to provide a highly valid and reliable index of perceived item-difficulty for each item, inter-teacher variance was held at a minimum by employing Ss from two classes taught by the same instructor. Furthermore, the Ss scaled the multiple-choice questions after having completed their final comprehensive examination in their fall introductory psychology course. This procedure was followed in order to insure that the Ss were maximally sensitized to the task, that is, they had studied their textbook for their final examination and now were prepared to more accurately assess subjective difficulty.

The test booklets were randomly distributed to the 142 Ss after they had completed their final examinations. The instructions attached to each booklet directed the Ss to rate each multiple-choice question on the degree of "subjective difficulty one would experience in reaching a solution to that particular question no matter whether that solution is correct or incorrect." For each question an 11-point continuum ranging from "extremely difficult" through "average" to "extremely easy" was provided. This rating procedure was a modification of Thurstone's prescribed sorting technique; that is, instead of sorting each statement into one of eleven piles, the raters were provided with a continuum for each item consisting

of numbers from one to 11 with a circle under each number and were asked to darken in the circle that indicated their rating (Jacobs & Munz, 1968). Each continuum was labeled appropriately--extremely difficult, 11; average, 6; and extremely easy, 1 (see Appendix B for instructions and answer sheet).

After the ratings for each multiple-choice question were compiled, median intensity scores (Mdn) and semi-interquartile ranges (Q) were computed (see Appendix C). The median intensity score provides an index of the absolute position of the item on the subjective difficulty continuum and the Q statistic is an index of the interjudge variability around the median intensity score.

Nine graduate instructors (all but one beyond the "Masters" degree level), each of whom had just completed teaching two semesters of introductory psychology from the Morgan and King text, also scaled the 210 multiple-choice questions (see Appendix C). The scaling procedure was identical to that described above except for the fact that each instructor (1) scaled all 210 questions and (2) rated each multiple-choice question on the degree of "subjective difficulty an introductory psychology student would experience in reaching a solution to that particular question no matter whether that solution is correct or incorrect." Median intensity ratings and semi-interquartile ranges for each item also were computed from the

instructor ratings. In an attempt to assess instructor consistency (interjudge agreement) in rating subjective difficulty, Ebel's intraclass correlation (Ebel, 1957) was computed for the median ratings. This statistic gives essentially an average intercorrelation or, more specifically, the reliability of mean ratings for k raters, where " k " is the number of raters.

Comparison of student perceived item-difficulty ratings with instructor perceived item-difficulty ratings was made by correlating item median intensity values for both groups. Furthermore, the average subjective difficulty of the group of items was assessed by comparing the mean of the median intensity ratings for the instructor group against the student's average median intensity value.

Construction of Item Arrangements

In order to investigate the effect of item-difficulty sequencing in the academic setting three forms of an achievement test (E-H, H-E, & R) were constructed based on student perceived item-difficulty values. The construction of these three item arrangements followed the major criterion of constructing test forms which maximized the effect of an "item-arrangement" response set on academic performance. The following points emerged as crucial guidelines in generating item arrangements which adhered to this criterion (Munz, 1968):

1. Freedom from Conflicting Response Sets. In order to maximize the effect of an "item-arrangement" response set on academic test performance it was necessary that the achievement test be relatively free from conflicting response sets which might, unbeknownst to the investigator, either mask such an effect, or interact with item sequencing to produce an artificial sequence effect [p. 10].

This condition was satisfied by using a multiple-choice test format, since it "appears to be the only generally useful form that is free from response sets" (Cronbach, 1946, p. 488). Furthermore, other testing procedures were utilized which also have been suggested as producing a relatively "response-set" free examination (Cronbach, 1946). Namely, (1) the use of a power test as contrasted with a speeded test, (2) the use of a heterogeneous group of item-difficulty values, and (3) giving instructions to respond to all items.

2. Item-Difficulty Values. The item-difficulty values needed to be extremely reliable such that, for example, an easy-to-hard arrangement of items would truly consist of items which progressively become more difficult for the student [p. 10-11].

This condition was satisfied by selecting for inclusion in the item arrangements only those items which had relatively small semi-interquartile ranges. Furthermore, teacher contamination of item-difficulty values due to teaching idiosyncrasies was avoided by having the same instructor teach both the scaling group (fall semester) and the experimental group (spring semester) of Ss.

3. Test Construction Characteristics. The chosen items and resulting sequences needed to comply with as many of the requirements of an ideal achievement test as possible. Such test construction characteristics as an adequate content sample, equal number of items at each difficulty level in the possible range, large number of questions, a mean test difficulty level near the middle of the possible score range, etc. (see Nunnally, 1959), were considered as important characteristics for a "good" test [p. 11].

This condition was fulfilled by adhering to the following procedure in choosing the test items for the final three test formats:

1. A large number of questions ($N = 100$) which adequately represented the content of the 14 chapters was selected from the item pool.

2. The 100 items were selected so that there was a relatively even spread of item difficulty values ranging from 1.42 to 9.20.

3. The mean difficulty level of the 100 items was near the middle of the difficulty range 5.86.

4. The 100 items were arranged in the following orders: increasing order of perceived item-difficulty (Form E-H), decreasing order of item difficulty (Form H-E) and randomly mixed (Form R) (see Appendix D).

After completing the examination, the students filled out a questionnaire booklet containing the PSI (Jacobs & Munz, 1968) and a test evaluation form similar to Burgess' Test Evaluation form (1968; see Appendix E). The Ss were informed prior to the examination that a

questionnaire booklet containing "two questionnaires" would be administered on a voluntary basis after the examination. Furthermore, they were informed that if they would fill out this booklet they would earn experimental credit for their effort. The Ss were also informed that the results of these questionnaires would be used for research purposes only and would in no way affect their grades in the course.

Pre-Experimental Procedure

Approximately two weeks prior to the final examination the Achievement Anxiety Test (AAT; Alpert & Haber, 1960) was administered to all Ss (see Appendix F). The AAT was designed to measure the effects of anxiety experienced in test-taking situations. The theory underlying the test is that test-taking anxiety (achievement anxiety) is a bidimensional construct which may have facilitating as well as debilitating effects on academic performance. Therefore, some individuals have their performance facilitated by test anxiety, while for others test anxiety depresses their test performance. According to this view, there are also those individuals whose test performance is not affected by the typical anxiety-provoking situations, either by improving or depressing their scores. "Thus, an individual may possess a large amount of both anxieties or of one but not the other, or of none of either" (Alpert & Haber, 1960, p. 213).

Each type of anxiety is measured by a separate subtest of items (facilitating scale--AAT+, and debilitating scale--AAT-). Four basic achievement anxiety reaction types can be operationalized through use of both scales. These types are the following: (1) facilitators, those respondents scoring relatively high on the facilitating scale (AAT+) and relatively low on the debilitating scale (AAT-); (2) debilitators, those respondents scoring relatively high on the AAT- and relatively low on the AAT+; (3) nonaffecteds, those respondents scoring relatively low on both AAT+ and AAT-; and (4) high-affecteds, those respondents scoring relatively high on both AAT+ and AAT-.

Procedure

Both sections of a spring introductory psychology course (N = 133) were tested in the evening as a group. The instructor randomly distributed the three forms of the examination (Form E-H, Form H-E, and Form R) and instructed the Ss to work directly through the test booklet and only go back to unanswered questions after having attempted all questions once. The rationale given for this directive was that "on previous examinations some students had worked from the back to the front of their exam so that they could copy from their neighbor's exam. Therefore, any student not working progressively from the front of his test booklet will be regarded with suspicion." The Ss were also told that they would have more than adequate time to complete the examination.

In order to test the hypothesis of a general sequencing effect, the test performance data (number of questions answered correctly) for each S (N = 133) were subjected to a one-way, unequal N, analysis of variance. The PSI data were subjected to the same one-way, unequal N, analysis of variance. The evaluation data were subjected to three one-way, unequal N, analyses of variance.

In order to test the hypotheses generated from the Munz-Smouse study (1968), the four anxiety reaction types based on the AAT were identified within each item-sequence. The following typing procedure was employed (Munz & Smouse, 1968):

An AAT+ score and an AAT- were obtained on each S after which the AAT- score was subtracted from the AAT+. A positive difference indicated a relatively high AAT+ and a negative difference indicated a relatively high AAT-. When these differences scores were ranked, the top 10 Ss in the distribution were defined as facilitators and the bottom scoring Ss were defined as debilitators. For all remaining Ss the two scores were summed and ranked. The top 10 Ss in the resulting distribution were defined as high-affecteds while the bottom 10 scores were defined as non-affecteds [p. 371].

This process eliminated approximately 10 Ss from each of the item sequence groups leaving data for a 3 (item sequence) by 4 (achievement anxiety type) analysis of variance with a cell size of N = 8 and a total N = 96.

In order to compare perceived item-difficulty values with "p" index values, a correlation coefficient was computed within the random group between perceived

item-difficulty values and "p" index values derived through item analysis of the final examination.

CHAPTER III

RESULTS

Evaluation of Perceived Item

Difficulty Ratings

As suggested, the perceived item-difficulty ratings of the instructors and the students (see Appendix C) reflected varying degrees of inter-item variability with some items being more "efficient" estimates of the perceived item-difficulty parameter than other items. The instructor inter-item variability range (largest item Q score minus smallest item Q score) was 3.01 with the largest Q score of 3.33 for item number 65 (form A) and the smallest Q score of .32 for item number 11 (form A). The standard deviation of the Q scores was .50. The student inter-item variability range was 1.73 with the largest Q score of 2.64 for item number 32 (form A) and the smallest Q score of .91 for item number 27 (form A). The standard deviation of these Q scores was .30.

Table 1 presents the analysis of the reliability (average intercorrelation) of the instructor ratings (Mdn scores). As computed through use of Ebel's

TABLE 1

Intraclass Correlation Summary Table of the Average
Intercorrelation of the Instructor Perceived
Item-Difficulty Ratings

Source of Variation	Sum of Squares	<u>df</u>	Variance	<u>r</u>
Items	3258.11	209	15.59	.78
Raters	1124.82	8	*	
Remainder	5822.96	1672	3.48	
Total	10205.89	1889		

*Variance not needed in final computation.

intraclass correlational method the average intercorrelation was $\underline{r} = .78$.

The relationship between student perceived item-difficulty ratings (Mdn scores) and instructor perceived item-difficulty ratings (Mdn scores) was reflected by a correlation coefficient of $\underline{r} = .62$. A comparison between the two groups of the subjective average difficulty of the group of items was made by comparing the mean of the median intensity ratings for the instructor group against the mean of the student ratings. Table 2 presents the means and standard deviations for the two groups of ratings. The mean score of the instructors' ratings was not significantly different from the students' mean score ($\underline{t} < 1.00$, $\underline{df} = 418$, $p > .05$).

TABLE 2

Means and Standard Deviations for the Median Intensity Ratings of the Students and Instructors

Group	M	S.D.
Instructor	5.95	1.5
Students	5.98	1.2

Evaluation of Perceived Item-Difficulty Sequencing

Table 3 presents the one-way, unequal N, analysis of variance of the performance scores as a function of perceived item-difficulty sequencing (see Appendix G for the primary data). As evidenced in Table 3, the analysis did not support the hypothesized general sequencing effect ($F < 1.00$, $df = 2/130$, $p > .05$).

TABLE 3

Analysis of Variance Summary Table of Achievement Performance Scores as a Function of Perceived Item-Difficulty Sequencing

Source of Variation	<u>df</u>	MS	<u>F</u>	p
A (Item Sequence)	2	140.60	< 1.00	N.S.
Error	130	163.23		
Total	132			

NS--Not significant at the .05 level.

As shown in Table 4 (see Appendix G for the primary data), the hypothesized effect of item sequencing upon test-taking anxiety also was not supported ($F < 1.00$, $df = 2/129$, $p > .05$).

TABLE 4

Analysis of Variance Summary Table of PSI Scores as a Function of Perceived Item-Difficulty Sequencing

Source of Variation	<u>df</u>	MS	<u>F</u>	p
A (Item Sequence)	2	3.16	< 1.00	N.S.
Error	129	4.45		
Total	131			

NS--Not significant at the .05 level.

A separate statistical analysis was performed on the data derived from each of the three questions comprising the Test Evaluation Form. Table 5 presents the one-way, unequal N, analysis of variance of the "perceived test difficulty ratings" as a function of item sequence (see Appendix G for the primary data). As hypothesized, there was a statistically significant item-difficulty sequence effect upon the Ss ratings of perceived test difficulty ($F = 4.14$, $df = 2/126$, $p < .05$); however, the specific predictions were only partially supported. An a priori comparison of means was made through use of the t-test statistic (Bruning & Kintz, 1968, p. 112-114).

The t-test analysis revealed that: (1) as predicted, Ss rated the E-H form an easier examination as compared with the ratings of Ss having the R form ($p < .05$) and Ss having the H-E form ($p < .05$); and (2) inconsistent with the predicted effect, there was no statistically significant difference between the ratings of Ss having the H-E form and those having the R form ($p > .05$).

TABLE 5

Analysis of Variance Summary Table of Test Difficulty Ratings as a Function of Perceived Item-Difficulty Sequencing

Source of Variation	<u>df</u>	MS	<u>F</u>	p
A (Item Sequence)	2	2.36	4.14	< .05
Error	126	.57		
Total	128			

As hypothesized, there was a statistically significant effect of item sequence upon the ratings of "test fairness" ($F = 4.84$, $df = 2/126$, $p < .01$) (see Table 6 and Appendix G for the primary data). However, a priori t-test comparison of means revealed only partial support for the hypothesized effect. That is, Ss receiving the R and E-H sequences significantly rated their test "fairer" as compared with the ratings of the H-E group of Ss ($p < .01$). Contrary to prediction, there was no statistical

difference between R and E-H "fairness" ratings ($p > .05$).

TABLE 6

Analysis of Variance Summary Table of Test Fairness Ratings as a Function of Perceived Item-Difficulty Sequencing

Source of Variation	<u>df</u>	MS	<u>F</u>	p
A (Item Sequences)	2	4.40	4.84	< .01
Error	126	.91		
Total	128			

As evidenced in Table 7 (see Appendix G for the primary data), the hypothesized effect of item sequence upon test comprehensiveness did not materialize ($\underline{F} = 1.16$, $\underline{df} = 2/125$, $p > .05$).

TABLE 7

Analysis of Variance Summary Table of Test Comprehensiveness Ratings as a Function of Perceived Item-Difficulty Sequencing

Source of Variation	<u>df</u>	MS	<u>F</u>	p
A (Item Sequence)	2	1.34	1.16	N.S.
Error	125	1.16		
Total	127			

NS--Not significant at the .05 level.

Table 8 presents the 3 (item sequence) by 4 (anxiety reaction type) analysis of variance results along with the simple main effects analysis (see Appendix G for the primary data). Consistent with the Munz-Smouse findings (Hypothesis 1), the analysis revealed no statistically significant differences among the group performance scores of the three item sequences ($F = 1.48$, $df = 2/84$, $p > .05$). Moreover, consistent with Hypothesis 2, there was a significant effect of personality type on performance scores ($F = 7.84$, $df = 3/84$, $p < .01$). Probing with the Newman-Keuls Test (NKT) (see Winer, 1962, pp. 77-85) revealed support for the specific predictions, that is, facilitators scored significantly higher than debilitators ($p < .01$) and nonaffecteds ($p < .01$). However, inconsistent with Hypothesis 2 was the finding that facilitators scored higher than high-affecteds ($p < .01$).

Table 8 shows that Hypothesis 3 was supported in that there was a statistically significant interaction among the three item-difficulty sequences and the four achievement-anxiety types ($F = 2.43$, $df = 6/84$, $p < .05$). A simple main effects analysis indicated that within the achievement anxiety factor the R sequence and E-H sequence levels were significant ($F = 9.54$, $df = 3/84$, $p < .01$; $F = 3.36$, $df = 3/84$, $p < .05$, respectively). Probing these two levels with the NKT revealed that (1) on the random form, facilitators and nonaffecteds scored significantly

TABLE 8

Analysis of Variance Summary Table of Performance Scores as a Function of Item Sequence and Personality Type

Source of Variation	<u>df</u>	MS	<u>F</u>	p
A (Item Sequence)	2	190.20	1.48	N.S.
B (Achievement-Anxiety Type)	3	1004.62	7.84	<.01
B for A ₁ (Random)	3	1221.80	9.54	<.01
B for A ₂ (Easy-to-Hard)	3	429.78	3.36	<.05
B for A ₃ (Hard-to-Easy)	3	261.58	2.04	N.S.
A X B	6	310.89	2.43	<.05
Error	84	128.07		
Total	95			

NS--Not significant at the .05 level.

higher than the debilitators ($p < .05$) and the facilitators scored significantly higher than the high-affecteds ($p < .05$), and (2) on the E-H form, facilitators and high-affecteds scored significantly higher than the nonaffecteds ($p < .05$; see Figure 1). It is evident from inspection of Figure 1 that the two specific subhypotheses formulated under Hypothesis 3 were contradicted, that is, the non-affecteds' performance did not improve systematically across forms in the order R, E-H, and H-E as predicted. Furthermore, there was a statistically significant

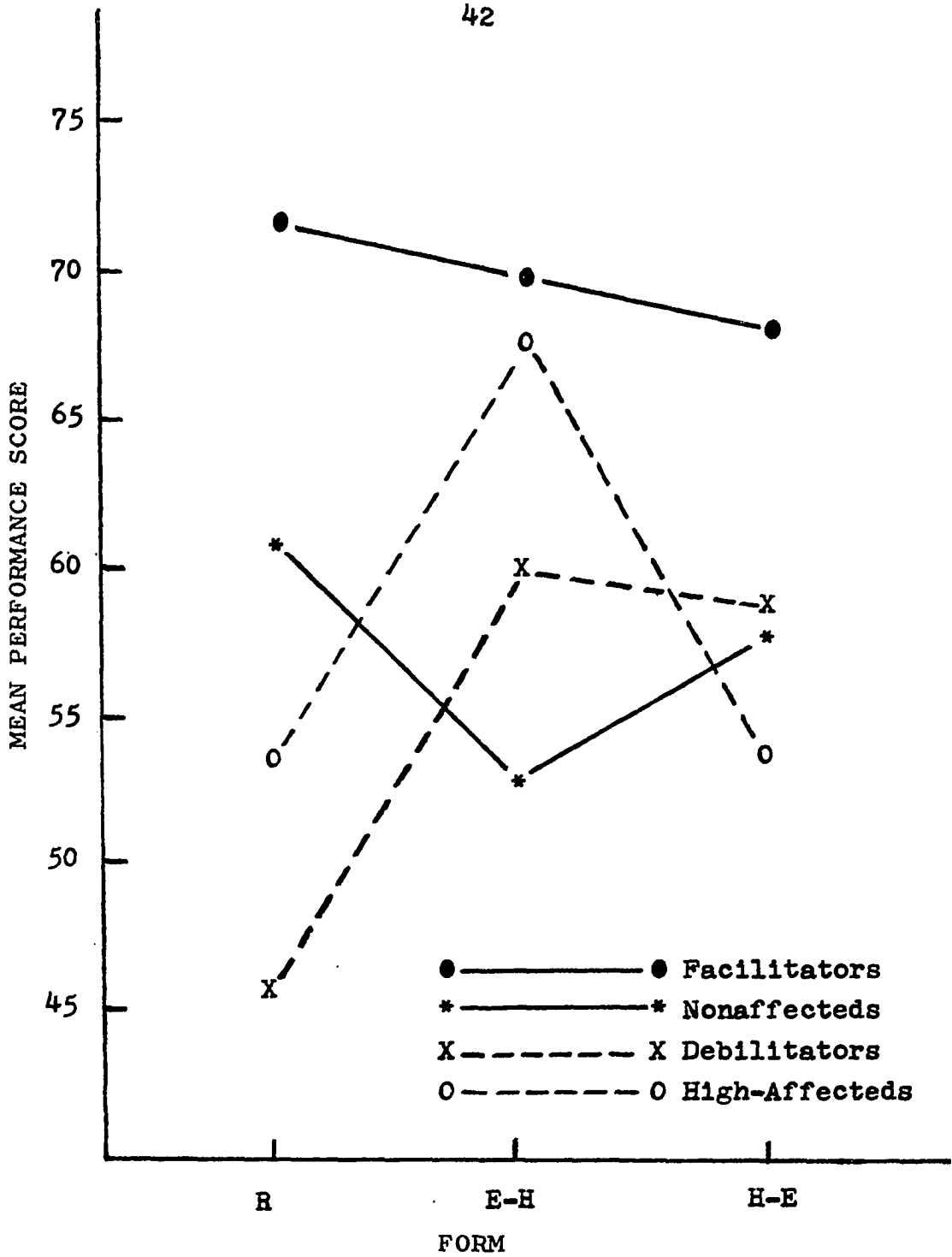


Figure 1. Mean performance scores for achievement-anxiety types (facilitators, high-affecteds, debilitators, and nonaffecteds) on perceived item-difficulty arrangements consisting of random (R), easy-to-hard (E-H), and hard-to-easy (H-E).

improvement in the debilitators' performance on forms E-H and H-E as compared with form R performance ($p < .05$).

Item analysis of the final examination performance data for the R form yielded "p" index values (objective difficulty values) for each question of the 100 question examination. A correlation coefficient was computed between the objective "p" index values of item-difficulty and the perceived item difficulty values derived through psychological scaling procedures. In accordance with previous research (e.g., Hertzman, 1937), a slight relationship was found ($r = .52$).

CHAPTER IV

DISCUSSION

The present study had two general aims: first, to develop and evaluate a method for deriving perceived item-difficulty values of multiple-choice examination questions, and second, to assess the effects of perceived item-difficulty sequencing in the academic setting.

The results of the various analyses of the perceived item-difficulty Thurstone ratings were very encouraging. Both the instructor ratings and student ratings supported the contention that an adequate numerical representation of the difficulty of a test item is more complex than reflected by the typical "p" index. It appears that if the test constructor desires to represent statistically the difficulty of an item with the same number for all individuals of a defined sample, then because of the varying degrees of inter-item subjective-difficulty variability (as evidenced in the student and instructor ratings), it would behoove the test constructor to procure for inclusion in his test only those items which are the more efficient estimators of the item-difficulty parameter.

The moderately high reliability (inter-observer agreement) of the instructor ratings lends support for the value of the scaling technique by suggesting that something, whether it be item-difficulty or not, is being scaled consistently by individuals with varying teaching idiosyncrasies. The validity of this index receives some support from the finding that there is a moderate relationship between student perceived item-difficulty ratings and the empirically derived "p" index values. Furthermore, the above results when combined with the findings that there is a moderate relationship between student and instructor perceived item-difficulty ratings and no difference between the two groups on the average subjective difficulty of the group of scaled items, lends credence to the supposition that an absolute index of subjective item-difficulty is meaningful.

An interesting practical aspect has emerged from these results. Because of the moderate relationship between student and instructor item-difficulty ratings and the high reliability of the instructor ratings, it seems plausible to suggest that the classroom instructor, interested in obtaining rough item-difficulty estimates for classroom test construction use, need not go through the laborious task of having students scale his test items. He may, instead, scale the item himself and feel somewhat confident that his ratings will roughly approximate those

of the students. However, while the subjective ratings of the individual instructor are easier to derive than those of the students, and for that matter, the "p" index values, they are open to one of the major criticisms of the "p" index--no index of variability. Consequently, chosen questions based on the individual instructor ratings will not be the best estimates of subjective item-difficulty experienced by his examinees. For practical test construction purposes, however, this approach appears to have merit. For research purposes, where a more sensitive item-difficulty index is desired, the subjective index based on student ratings may provide the needed refinement. Nevertheless, whether the subjective index is a more meaningful index of item-difficulty, as compared with the "p" index, is an empirical question. The true utility of the index and its underlying methodology can only be adequately judged by further psychometric research.

The effects of perceived item-difficulty sequencing in the academic setting, as assessed by this study, generally supported previous research findings which employed the objective index. However, a number of new and interesting findings also emerged which now need to be incorporated into previous interpretations of the utility of item-difficulty sequencing in test construction.

In accordance with recent investigations of objective item-difficulty sequencing in the achievement setting

(Brenner, 1964; Munz & Smouse, 1968; Smouse & Munz, 1968), the results of this study supported the contention that there is no empirical justification for the advocated test construction practice of arranging test items in an easy to hard order. No justification, that is, from the argument that an E-H item-difficulty sequence produces, via some provoked response set, higher performances scores than an R or H-E arrangement; or from the argument that an E-H item-difficulty sequence significantly reduces test-taking anxiety as compared with an R or H-E arrangement. However, the results derived from the test evaluation form suggest that, while sequencing does not appear to affect group performance scores or test-taking anxiety, examinees do leave the examination with different evaluative feelings about the examination depending on which sequence form they had received. It appears that examinees receiving the E-H form have more positive post-examination feelings about their examination (easier and fairer) than do examinees receiving the H-E and R form. The R group, while having similar feelings as the H-E group concerning test difficulty, has feelings as positive as the E-H group about the fairness of their test.

Test constructors have not suggested post-examination feelings about such facts as test fairness and test difficulty as important considerations for sequencing test items in an E-H fashion. The value of such post-examination

feelings about the course examination seems quite apparent in view of the voluminous amount of research demonstrating the influence of acquired attitudes and beliefs on subsequent behavior. Further research, instead of attempting to demonstrate a direct sequencing effect on performance scores, may investigate this new dimension of utility of item difficulty sequencing. For example, such research questions as the following may yield interesting results:

1. Does an E-H test given as the first examination in a course affect subsequent course examination performance as compared with H-E and R exams?

2. Do E-H course examinations given throughout the duration of the course affect a student's over-all course evaluation as compared with R and H-E exams?

3. Does the objective "p" index sequence yield results similar to those derived from the subjective sequencing of test items?

Until future research investigates the above questions and the numerous others which are apparent from this discussion, it may be concluded from this study that, while the E-H arrangement of test items does not appear to improve test performance or reduce test-taking anxiety, its more subtle value lies in the fact that students leave the examination with a more positive set of attitudes toward the test.

Further, the results of this study suggest the

existence of a complex relationship between test-taking personality types, perceived item-difficulty sequencing and academic performance. While the results generally supported the Munz-Smouse (1968) findings, i.e., there was no significant perceived item-difficulty sequencing effect upon performance scores; differential reactions to test-taking anxiety did significantly affect performance scores; and perceived item-difficulty sequencing and achievement anxiety reaction type did interact to produce a significant effect on achievement performance scores, the specific hypothesized fluctuations of means for the item-sequence by personality-type groups were not in agreement with their results. One possible explanation for this inconsistency of results is that the underlying mechanisms causing the one interaction are not the same as those which caused the other interaction. The plausibility of this explanation is attested to by the fact that a subjective index of item-difficulty was employed in this study to generate the item sequences as compared with the objective "p" index in the earlier study. Nevertheless, there appears to be one consistent line of evidence. As shown in Figure 1, as one progresses from the R sequence to the H-E sequence it appears that less variance is attributable to the personality type variable. This finding is consistent with the Smouse-Munz (1969) re-analysis of their original data. It appears, then, that while this study

does not ideally support the Munz-Smouse study, it can be suggested from inspection of Figure 1 that subjective, as well as objective, item-difficulty sequencing will interact with personality test-taking types to provoke differential response styles. Furthermore, these differential response styles are more influential in determining test score performance under the R and H-E sequence than under the H-E sequence.

It is evident from this study that further research is needed before a conclusive statement can be made about the nature of these differential response styles and the role that item-difficulty sequencing plays in their instigation. While the Munz-Smouse (1968) study hypothesized a specific set of underlying causal factors which were responsible for the interaction between sequencing and personality type, the results of this study, if generally valid, clearly suggest the need for a new set of hypothesized mechanisms to account for such an interaction. However, in view of the fact that a specially developed index of subjective difficulty was employed in this study and thus lacks the needed research to suggest its influence on other variables, it seems premature at this point to make such speculation. Future research in both the aptitude and achievement setting surely will lead to such theorizing.

CHAPTER V

SUMMARY

A review of the literature suggested that the lack of evidence for item-difficulty sequencing as a determinant of a response set in the academic setting may be due to the inability of the test constructor to actually sequence items in the prescribed order of easy-to-hard. In other words, the objective "p" index may not be the best index of psychological difficulty. The present study was an attempt to (1) develop a method for deriving perceived item-difficulty values of multiple-choice questions, (2) assess the method by comparing student perceived item-difficulty values to instructor perceived item-difficulty values, by determining the degree of inter-observer agreement among the instructor ratings and by determining if the average subjective difficulty of the group of items scaled by the two groups differed significantly, (3) investigate systematically the effect of perceived item-difficulty sequencing on academic examination performance, (4) assess possible underlying factors responsible for such an effect, (5) determine if the basic findings of the Munz and Smouse (1968) study, e.g., the interaction found

between objective item-difficulty sequencing and achievement-anxiety reaction type as evidence in achievement test score performance, are valid in situations where student "perceived item-difficulty" values are used to arrange the achievement test items, and (6) compare student perceived item-difficulty values to item-difficulty values derived from the typical "p" index.

Utilizing Thurstone's method of equal-appearing intervals as the scaling technique, 142 introductory psychology students and nine introductory psychology instructors scaled 210 multiple-choice examination questions on the "subjective difficulty an introductory psychology student would experience in reaching a solution to that particular question no matter whether that solution was correct or incorrect." Median intensity scores (Mdn) and semi-interquartile ranges (Q) were computed for each group of ratings. The construction of the item arrangements, hard-to-easy (H-E), easy-to-hard (E-H), and random (R) were based on student perceived item-difficulty values (Mdn) and followed the major criteria for constructing test forms which maximized the effect of an "item-arrangement" response set on academic performance. One hundred questions with Mdn values ranging from 1.42 (very easy) to 9.20 (very difficult) and low Q scores were chosen to comprise the final three test forms.

The following semester the three arrangements were

randomly distributed as a final examination to 133 introductory psychology students. To assess the effects of sequencing in the academic setting a situational measure of test-taking anxiety was taken along with test evaluation information. Furthermore, prior to the final examination the Achievement Anxiety Test (AAT; Alpert & Haber, 1960) was administered to all Ss allowing for the classification of Ss into achievement-anxiety reaction types following the typing procedure employed by Munz and Smouse (1968).

The following is a summary of the specific findings of this investigation:

1. The reliability (average intercorrelation) of the instructor ratings (Mdn scores) computed by Ebel's intraclass correlational method was $\underline{r} = .78$.

2. The relationship between student perceived item-difficulty ratings (Mdn scores) and instructor perceived item-difficulty ratings (Mdn scores) was $\underline{r} = .62$.

3. No significant difference between the two groups on the average subjective difficulty of the group of scaled items was found.

4. The hypothesized effect of perceived item-difficulty sequencing influencing achievement performance scores was not supported.

5. The hypothesized effect of perceived item-difficulty sequencing influencing test-taking anxiety scores was not supported.

6. Partial support for the hypothesized effect of item-sequencing affecting ratings of perceived test difficulty was found, i.e., Ss rated the E-H form as an easier test when compared with the ratings of Ss having the R form and Ss having the H-E form.

7. Partial support for the hypothesis of an item sequence effect upon the ratings of "test fairness" was found, i.e., Ss receiving the R and E-H sequences significantly rated their test "fairer" as compared with the ratings of the H-E group of Ss.

8. No support was found for the hypothesized effect of item sequence upon ratings of "test comprehensiveness."

9. The Munz-Smouse (1968) findings were partially supported. That is, no statistically significant differences among the performance scores of the three item sequence groups were found, there was a significant effect of personality test-taking type on achievement performance, and the interaction between item-sequence and personality-type was evidenced. However, the specific fluctuations of the means were not congruent with those found in the Munz-Smouse study.

10. There was a relationship between "p" index values derived from the R form examination data and perceived item-difficulty values of $r = .52$.

These results were interpreted as suggesting that

(1) the subjective index, derived through the use of Thurstone's scaling procedure, may have practical as well as theoretical implications for test constructors, (2) test constructors should consider the effects of subjective item-difficulty sequencing on examinees' post-examination test evaluation as justification for sequencing items in an easy to hard order, and (3) there are different causal mechanisms underlying the interaction between subjective item-difficulty sequencing (as compared with objective item-difficulty sequencing) and achievement-anxiety reaction type upon achievement test score performance.

REFERENCES

- Aiken, L. R. Jr. Item content and position effects on multiple choice tests. Journal of Psychology, 1964, 58, 369-373.
- Alpert, R., & Haber, R. N. Anxiety in academic achievement situations. Journal of Abnormal and Social Psychology, 1960, 61, 207-215.
- Berger, V. F., Munz, D. C., Smouse, A. D., & Angelino, H. The effects of item difficulty sequencing and anxiety reaction type on aptitude test performance. Journal of Psychology, 1969, 71, 253-258.
- Brenner, M. H. Test difficulty, reliability, and discrimination as functions of item difficulty order. Journal of Applied Psychology, 1964, 48, 98-100.
- Broen, W. & Wirt, R. Varieties of response sets. Journal of Consulting Psychology, 1958, 22, 237-240.
- Bruning, J. L. & Kintz, B. L. Computational Handbook of Statistics. Illinois: Scot, Foresman, & Co., 1968.
- Burgess, T. C. Item arrangement and student evaluation of examinations. Psychology, 1968, 5, 69-73.
- Conrad, H. S. The experimental tryout of test materials. In E. F. Lindquist (Ed.), Educational Measurement. Washington, D. C.: American Council on Education, 1951, pp. 250-265.
- Coombs, C. H. The concepts of reliability and homogeneity, Educational and Psychological Measurement, 1950, 10, 43-56.
- Cronbach, L. J. Response sets and test validity. Educational and Psychological Measurement, 1946, 6, 475-494.

- Cronbach, L. J. Further evidence on response sets and test design. Educational and Psychological Measurement, 1950, 10, 3-31.
- Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 1959, 66, 186-201.
- Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424.
- Farmer, E. Concerning subjective judgment of difficulty. British Journal of Psychology, 1928, 18, 438-442.
- Fiske, D. W., & Maddi, S. R. (Eds.). Functions of Varied Experience. Homewood, Ill.: Dorsey Press, 1961.
- Flaugher, R. L., Melton, R. S., & Myers, C. T. Item rearrangement under typical test conditions. Educational and Psychological Measurement, 1968, 28, 813-824.
- Greene, H. A., & Jorgensen, A. N. The Use and Interpretation of High School Tests. New York: Longmans, Green, and Co., 1939.
- Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1954.
- Guilford, J. P., & Cotzin, M. Judgment of difficulty of simple tasks. American Journal of Psychology, 1941, 54, 38-52.
- Helmstadter, G. Procedures for obtaining separate set and content components of a test score. Psychometrika, 1957, 22, 381-393.
- Helmstadter, G. Principles of Psychological Measurement. New York: Appleton-Century-Crofts, 1964.
- Hertzman, M. Confidence ratings as an index of difficulty. Journal of Experimental Psychology, 1937, 21, 113-119.
- Jackson, D. N., & Messick, S. Content and style in personality assessment. Psychological Bulletin, 1958, 55, 243-252.
- Jacobs, P. D., & Munz, D. C. An index for measuring perceived stress in the college population. Journal of Psychology, 1968, 70, 9-15.

- Jordan, A. M. Measurement in Education. New York: McGraw-Hill, 1953.
- Loevinger, J. Theory and techniques of assessment. Annual Review of Psychology, 1959, 287-318.
- Lund, K. Test performance as related to order of item difficulty, anxiety, and intelligence. Unpublished Doctoral dissertation, Northwestern University, Evanston, Illinois, 1953.
- Malmo, R. B. Activation: A neurophysiological dimension. Psychological Review, 1959, 66, 367-386.
- Messick, S. Separate set and content scores for personality and attitude scales. Educational and Psychological Measurement, 1961, 21, 915-923.
- Metfessel, N., & Sax, G. Systematic bias in the keying of correct responses on certain standardized tests. Educational and Psychological Measurement, 1958, 18, 787-790.
- Mollenkopf, W. G. An experimental study of the effects on item analysis data of changing item placement and test time limit. Psychometrika, 1950, 15, 291-315.
- Morgan, C. T., & King, R. A. Introduction to Psychology. New York: McGraw-Hill, 1966.
- Morgan, C. T., & King, R. A. Supplementary Questions to Accompany Introduction to Psychology. New York: McGraw-Hill, 1967.
- Munz, D. C. The effects of item difficulty sequences on achievement testing scores under two conditions of anxiety. Unpublished Master's thesis, University of Oklahoma, Norman, Oklahoma, 1968.
- Munz, D. C., & Smouse, A. D. The interaction effects of item difficulty sequence and achievement anxiety reaction on academic performance. Journal of Educational Psychology, 1968, 59, 370-374.
- Myers, C. T. The relationship between item difficulty and test validity and reliability. Educational and Psychological Measurement, 1962, 22, 565-571.
- Nunnally, J. C. Tests and Measurements. New York: McGraw-Hill, 1959.

- Peabody, D. Models for estimating content and set components in attitude and personality scales. Educational and Psychological Measurement, 1964, 24, 255-269.
- Perloff, R., & Wolins, L. Item difficulty as a function of perceived item direction. Educational and Psychological Measurement, 1965, 25, 79-85.
- Reese, T. W. The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. Psychological Monographs, 1943, 55, No. 251.
- Remmers, H. H., & Gage, N. L. Educational Measurement and Evaluation. New York: Harper, 1955.
- Ross, C. C. Measurement in Today's Schools (2nd ed.). New York: Prentice-Hall, 1947.
- Ross, C. C., & Stanley, J. C. Measurement in Today's Schools. New Jersey: Prentice-Hall, 1954.
- Sax, G., & Carr, A. An investigation of response sets on altered parallel forms. Educational and Psychological Measurement, 1962, 22, 371-376.
- Smouse, A. D., & Munz, D. C. The effects of anxiety and item difficulty sequence on achievement testing scores. Journal of Psychology, 1968, 68, 181-184.
- Smouse, A. D., & Munz, D. C. Item difficulty sequencing and response style: A follow-up analysis. Educational and Psychological Measurement, 1969, in press.
- Stricker, L. J. Difficulty and other correlates of criticalness response style at the item level. Educational and Psychological Measurement, 1965, 25, 683-706.
- Sweeney, C., Smouse, A. D., Munz, D. C., & Rupiper, O. A test of the inverted U hypothesis relating achievement anxiety and academic test performance. Manuscript in preparation, 1969.
- Thorndike, R. L., & Hagen, E. Measurement and Evaluation in Psychology and Education. New York: Wiley, 1955.

- Thurstone, L. L. Attitudes can be measured. American Journal of Sociology, 1928, 33, 526-554.
- Thurstone, L. L., & Chave, E. J. The Measurement of Attitude. Chicago: University of Chicago Press, 1929.
- Webster, H. Correcting personality scales for response sets or suppression effects. Psychological Bulletin, 1958, 55, 62-64.
- Wevrick, L. Response set in a multiple-choice test. Educational and Psychological Measurement, 1962, 22, 533-538.
- Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.
- Yamamoto, K., & Dizney, H. F. Effects of three sets of test instructions on scores on an intelligence scale. Educational and Psychological Measurement, 1965, 25, 87-94.
- Zucherman, M. The development of an affect adjective check list for the measurement of anxiety. Journal of Consulting Psychology, 1960, 24, 457-462.

APPENDIX A

TEST ITEM BOOKLETS

FORM A

QUESTIONS

1. Which is not a stage in the general-adaptation syndrome?
 - a. expectancy
 - b. alarm reaction
 - c. resistance to stress
 - d. exhaustion

2. Which of the following may not be true of an item on the Minnesota Multiphasic Personality Inventory (MMPI)?
 - a. It may be on a validity scale.
 - b. It may be on a scale other than one of the ten personality scales.
 - c. The content of the item is a good indication of the scale it will be on--e.g., face validity.
 - d. It may be on a hypomania scale.

3. "The trend toward making mental hospitals therapeutic, rather than custodial, institutions is bound to be accelerated by these agents." What agents are being considered?
 - a. reserpine and chlorpromazine
 - b. LSD-25 and similar agents
 - c. psychosurgical operations
 - d. narcotics

4. Projection is best defined as
 - a. an eagerness to discuss personal problems.
 - b. adoption of the values and taboos of older persons.
 - c. assigning traits we ourselves possess to others.
 - d. a tendency to react to others according to our motivations.

5. The qualitative change theory of memory maintains that
 - a. a memory gradually loses all "shape" or "form."
 - b. a memory is interfered with because of neural mechanisms located in the temporal lobe.
 - c. memories are distorted by the process of repression.
 - d. a memory goes through a series of systematic "distortions" as it is being forgotten.

6. The Minnesota Paper Form Board Test is a
 - a. scholastic aptitude test.
 - b. projective test.
 - c. psychomotor test.
 - d. mechanical aptitude test.

7. The difference between a compulsion and an obsession is like the difference between
 - a. words and phrases.
 - b. starting and stopping.
 - c. acts and ideas.
 - d. normal and abnormal.

8. Considering all patients, the trend over the last few years in the number of hospital admissions and resident patients per 100,000 has been a/an
 - a. increase in resident patients and an increase in admissions.
 - b. increase in resident patients and a decrease in admissions.
 - c. decrease in resident patients and an increase in admissions.
 - d. decrease in resident patients and a decrease in admissions.

9. In situational personality tests
 - a. the individual is presented with pictures which are somewhat ambiguous and asked to describe the situation pictured.
 - b. the individual is observed in the home.
 - c. a "real-life" problem situation is set up and the person's reactions observed.
 - d. the change in personality is observed as a person is changed from one environment to another.

10. Modern studies of dreaming indicate that there is/are
 - a. usually only one long dream during a night's sleep.
 - b. several dreams during the night and that each apparently goes on in "real time."
 - c. several dreams during the night and that the last dream of the night is typically completed in a few seconds.
 - d. several dreams during the night and that each is completed in a few seconds.

11. The founder of psychoanalysis was
 - a. Alfred Adler.
 - b. Sigmund Freud.
 - c. Carl Jung.
 - d. Carl Freud.

12. As psychology developed, many different schools of thought came to be emphasized and then faded into the background. Today
 - a. the primary emphasis is on structuralism.
 - b. Gestalt psychology is in the limelight.
 - c. primary emphasis is on the phenomenological method.
 - d. psychology has become more eclectic.

13. The control over responding after a go no-go discrimination has been learned is sometimes referred to as
 - a. response control of behavior.
 - b. simultaneous control.
 - c. stimulus control of behavior.
 - d. successive control.

14. The "lie detector" relies upon the notion that
 - a. autonomically controlled responses are not under voluntary control.
 - b. patterns of autonomic responses are different for each emotional state.
 - c. there are no neutral questions.
 - d. blood pressure is not highly related to the GSR.

15. The response pattern seen in the startle response is largely
 - a. acquired.
 - b. inborn.
 - c. inherited, but differs from person to person.
 - d. dependent upon cultural training.

16. An instinctive behavior must be species wide and
 - a. dependent on learning.
 - b. independent of maturation.
 - c. uncaused.
 - d. unlearned.

17. Noises
 - a. are twice the SPL of musical tones.
 - b. are caused by aperiodic waves.
 - c. do not contain any frequencies at all.
 - d. are caused by a regular single sine wave.

18. Needs tend to have their greatest effect on the perception of
 - a. objects.
 - b. ambiguous situations.
 - c. strongly structured situations.
 - d. colors.

19. Which is not an example of adaptation?
 - a. The smell of food cooking fades after a long stay in the kitchen.
 - b. We are not normally aware of the pressure of clothing.
 - c. The felt temperature depends upon the temperature to which we were previously exposed.
 - d. The brightness of an object depends upon the brightness of surrounding objects.

20. The Rorschach Test is different from the Thematic Apperception Test (TAT) in that
- the pictures used in the TAT are less ambiguous--most depict scenes or persons.
 - the TAT is an objective test.
 - the individual taking the TAT is merely asked to name the individuals in the pictures.
 - checks for test faking are built into the TAT.
21. In the history of psychology the nature versus nurture question was argued for many years. Today psychologists hold that
- men are more or less equal in heredity and that the environment in which one is nurtured determines what he will become.
 - heredity determines pretty much what kind of person one will be.
 - both environment and heredity play important parts in determining what kind of person one will be.
 - personality traits cannot be thought of as determined.
22. Of those admitted to public mental hospitals, one relatively large group is composed of people who are characterized by "minimal subjective anxiety, and little or no sense of distress. In most instances, the disorder is manifested by a lifelong pattern of action and behavior, rather than by mental or emotional symptoms." This characterizes the class of
- chronic brain syndromes.
 - personality disorders.
 - mild adjustment problems.
 - psychoneurotic reactions.
23. Which of the following developed a theory which emphasized that anxiety arises from social tensions in childhood rather than from conflicts between id and ego in childhood?
- Alfred Adler
 - Abraham Maslow
 - Karen Horney
 - Henry Murray
24. Indian children, who are tightly bound for long periods, show
- retarded motor development.
 - normal motor development.
 - accelerated motor development.
 - retarded motor development and accelerated language development.

25. The taste receptors are termed
 - a. papillae.
 - b. taste knobs.
 - c. taste epithelial surfaces.
 - d. taste buds.

26. The problem which is particularly acute in personality tests is
 - a. reliability.
 - b. practicality.
 - c. validity.
 - d. standardization.

27. The XY combination of chromosomes is characteristic of
 - a. females.
 - b. males.
 - c. either males or females.
 - d. neither males nor females.

28. The forgetting of traumatic (unpleasant, threatening) childhood experiences would most probably come under the heading of
 - a. motivated forgetting.
 - b. short-term forgetting.
 - c. qualitative-change forgetting.
 - d. interference forgetting.

29. The type of therapy, done especially with children, which utilizes toys, drawings, modeling, and other activities is sometimes called
 - a. group therapy.
 - b. psychodrama.
 - c. release therapy.
 - d. evaluation therapy.

30. Under intense illumination, some retinene is converted, reversibly, into
 - a. red cone pigments.
 - b. iodopsin.
 - c. vitamin A.
 - d. orange intermediates.

31. A hebephrenic schizophrenic is one who is
 - a. unmoving.
 - b. suspicious.
 - c. childish and silly.
 - d. depressed.

32. An anecdote about an introductory psychology class tells of a student who one day complained to his teacher that his parents were not spending their hard-earned money to have him come and learn about some &\$_#s* Russian and his slobbering dogs. He was probably referring to a lecture on
- B. F. Skinner's experiments on instrumental learning.
 - K. R. Roskolnikov's experiments on partial reinforcement.
 - studies by Bekhterev on avoidance learning.
 - I. P. Pavlov's classical conditioning studies.
33. Here is a list of traits: (1) impulsive behavior; (2) tendency toward a "multiple-personality;" (3) eccentric thinking; (4) withdrawal from others. Which of these traits characterize the schizoid personality?
- 2,3
 - 1,4
 - 2,4
 - 3,4
34. The theory proposed by Maslow has needs arranged in a hierarchy. The need highest in this hierarchy is the need for
- self-actualization.
 - functional autonomy.
 - belongingness.
 - esteem.
35. Animal A is trained on a VI schedule for a long period of time; animal B is trained on a continuous-reinforcement schedule for the same length of time. Both animals are then placed on extinction. You would expect
- animal B to make more responses during extinction than animal A.
 - animal A to make more responses during extinction than animal B.
 - animal A to stop responding almost immediately.
 - both animals to make approximately the same number of responses during extinction.
36. Normal curves are best characterized by
- the absolute heights of the histogram bars for each score interval.
 - positive and negative skewing.
 - a dip in the middle of the curve.
 - their smooth, symmetrical shape.

37. An apparatus for presenting perceptual materials for a brief time period is called a/an
- anomalouscope.
 - tachistoscope.
 - pseudoscope.
 - velociscope.
38. Genes are carried on which parts of a cell?
- enzymes
 - chromosomes
 - zygotes
 - alleles
39. A test is reliable to the extent that it
- correlates well with other measures of a particular trait.
 - can be given on a second occasion without fear of interference between the first and second scores.
 - yields the same measure on retest.
 - is based on a sample which adequately represents the intended population.
40. Infants allowed to choose their own food
- develop strong preferences for sweets.
 - develop preferences for meats and carbohydrates.
 - eat a balanced diet at each meal.
 - eat a balanced diet over the long run.
41. Any aspect of personality that is reasonably characteristic and distinctive is usually called a
- mode.
 - factor.
 - trait.
 - type.
42. One of the major aims of psychoanalytic therapy is to
- increase the resistance to anxiety-laden ideas.
 - make the patient aware of the sources of his anxiety.
 - treat schizophrenics by psychosurgery.
 - eliminate transference at all stages of the psychotherapeutic relationship.
43. The process in which one person gets another to accept an idea or attitude without citing proof or using coercion is called
- reeducation.
 - nondirective therapy.
 - suggestion.
 - behavior therapy.

44. Gwa the chimp excelled Donald in which behaviors early in development?
- Vocalizations
 - Rhythmicity of feeding cycle
 - Motor abilities
 - Emotional stability.
45. As a person with normal vision grows older, he will probably become more
- nearsighted.
 - farsighted.
 - photosensitive.
 - sensitive to extraspectral hues.
46. Sometimes a subject reports that a small hot stimulus feels cold. This is called
- anomalous warmth.
 - paradoxical cold.
 - anamalous cold.
 - punctate warmth.
47. One of the problems in the interpretation of dreams is
- the occurrence of rapid eye movements (REMs).
 - secondary elaboration of the dream.
 - paradoxical sleep.
 - positive transference from the dream.
48. Complex psychological characteristics which have a genetic cause
- can always be traced to a single pair of genes.
 - always involve a large number of genes.
 - sometimes involve primarily a single pair of genes and sometimes involve many genes.
 - can be traced to a single pair of genes 83 percent of the time.
49. The definition of the term "personality" is a long one. Part of the definition of personality refers to the
- characteristic and instinctive traits of an individual.
 - strength of character.
 - depth of the unconscious.
 - nomothetic traits which make a person act differently from time to time in the same situation.
50. Adolph stopped water from passing down the esophagus of a dog into its stomach. He then put enough water into the stomach to satisfy biological needs and allowed the dog to drink. If the dog was given an opportunity to drink immediately after the water

- had been put into the stomach, the dog would
- a. not drink.
 - b. drink about as much as he would have drunk had water not been put into the stomach.
 - c. drink until the source of water was withdrawn.
 - d. take only two or three sips of water to wet his mouth and tongue before stopping.
51. One advantage of the Wechsler Intelligence Scale over the Stanford-Binet test is that it
- a. does not involve merely a test of ongoing behavior.
 - b. is more useful for adults.
 - c. has a smaller standard deviation.
 - d. is a group test and not an individual test.
52. Negative transfer of training occurs most strongly when two paired-associates lists of nonsense syllables are related as which of the following?
- a. stimuli similar, responses similar
 - b. stimuli similar, responses dissimilar
 - c. stimuli dissimilar, responses similar
 - d. stimuli dissimilar, responses dissimilar
53. The process of adjusting focal length through changes in the shape of the lens is known as
- a. appropriation.
 - b. activation.
 - c. accommodation.
 - d. amblyopiation.
54. An aim of one type of therapy might be stated in the following way: the therapist tries to free the patient "to explore his life and experience anew, frees him to perceive in that experience new meaning and new goals." What type of therapy is being described?
- a. behavior therapy
 - b. directive therapy
 - c. psychoanalysis
 - d. client-centered therapy
55. The usual point of view with respect to defense mechanisms is that, in general, they defend a person against
- a. repression.
 - b. motivation.
 - c. anxiety.
 - d. the consequences of aggression.

56. Up to a point at least, one theoretical statement of the relationship between drive and behavior is that
- drive
habit = performance strength.
 - drive X habit = performance strength.
 - drive X habit = motivation.
 - drive
habit = motivation.
57. The necessity of sensory experience for the development of various types of perception is shown by the experiment with a chimpanzee who failed to
- hear after his ears were stopped up for two years.
 - learn to respond to touch after touch experience was restricted.
 - show visual figural after-effects.
 - show visual transposition.
58. What label is given by the text to the senses of the internal organs within the body cavity?
- vestibular senses
 - bodily senses
 - visceral senses
 - organic senses
59. The word that best describes the relation of heredity and environment is
- interaction.
 - indeterminancy.
 - dominance.
 - primacy.
60. The functionalists were especially influenced by the writings of
- Sigmund Freud.
 - Charles Darwin.
 - Rene' Descartes.
 - Max Wertheimer.
61. As the eye adapts to darkness, the earlier adaptation effects are produced by changes in the
- rods.
 - cones.
 - iris.
 - lens.

62. People with their stomachs removed
- still get hungry.
 - do not experience hunger.
 - have a constant mild hunger.
 - will not learn a task reinforced by food presentation.
63. Most important binocular cue for depth perception is
- convergence.
 - movement parallax.
 - interposition.
 - retinal disparity.
64. Human engineering concerns the
- selection of marriage pairs.
 - treatment of deviant behavior patterns.
 - design of equipment to match human capabilities.
 - survey of consumer attitudes.
65. The arithmetic mean is
- an interval-scale statistic.
 - the best measure of central tendency for skewed distributions of scores.
 - an ordinal scale statistic.
 - the most frequent score in a frequency distribution.
66. If the cues to depth or distance perception are eliminated, our perception of the size of an unfamiliar object tends to
- depend on the transposition of the figure-ground relationship.
 - follow the principles of size constancy.
 - depend on the size of the retinal image.
 - follow the brightness-gradient, or ratio of intensities, principle.
67. Two groups of water-and food-satiated rats are allowed to explore a T maze for several hours. For Group A the maze is empty; for Group B a water bottle is always present in the left arm of the maze. Group B animals are water-satiated and never drink.
- The animals are then made thirsty and are reinforced with water for left turns in the T maze. If Group B learns to go left faster than Group A, this fact might argue most directly for
- changed-response learning.
 - latent learning.
 - classical conditioning.
 - blocked-response learning.

68. High scores on which of the following Differential Aptitude Tests (DAT) might be especially compatible with the goal of a scientific major in college?
- clerical and mechanical
 - numerical and abstract
 - mechanical and psychomotor
 - theoretical and social
69. For lights, complementary colors are hues which
- mix to create another hue.
 - are perceived as gray or white when mixed.
 - go well together.
 - are entirely extraspectral.
70. Studies of the development of fear reactions in chimpanzees have indicated that
- adult chimpanzees would not escape a very intense noise.
 - certain stimuli, e.g., a disembodied chimp head, frightened chimps of all ages.
 - these stimuli (see above) frightened only infant chimps.
 - these stimuli (see above) frightened older chimps, but not infants.
71. Grouping principles in perception include
- proximity.
 - kurtosis.
 - scatter.
 - set, or expectancy.
72. Which one of the following men could not be credited with an important part in the founding of experimental psychology?
- Wilhelm Wundt
 - William James
 - Gustav Fechner
 - Alfred Adler.
73. Visual acuity is
- best when the retinal image falls 35 degrees from the fovea.
 - enhanced by contrast between the viewed object and its background.
 - better for an individual with 20/100 vision than for a person with 20/10 vision.
 - enhanced by the negative aftereffect.

74. Some basic methods of measuring long-term memory include three of the following. Which is not such a method?
- recall
 - reintegration
 - recognition
 - savings
75. The scores on the Kuder Preference Record are in terms of
- general interest categories--artistic and literary, for instance.
 - preferences for specific occupations.
 - a profile of specific aptitudes.
 - preferences for certain types of people.
76. Freudians contend that Oedipal anxiety (the Oedipus complex) is typically defended against by which defense mechanism?
- fixation
 - identification
 - projection
 - regression.
77. The person with a neurasthenic reaction is likely to
- be easily aroused to anger.
 - feel compelled to act without being able to control himself.
 - be generally nervous and fatigued.
 - suffer from delusions.
78. The psychoanalytic explanation of phobic reactions stresses
- stimulus generalization.
 - classical conditioning.
 - displacement and symbolism.
 - rationalization and compensation.
79. A centile score is the percentage of cases falling at or below a particular person's score in a distribution. The z score equal to a centile score of 50 is
- +1.65.
 - +1.00.
 - 0.00.
 - 1.00.
80. Learning without explicit instructions or intent to learn is called
- accidental learning.
 - implicit learning.
 - incidental learning.
 - nonpurposive learning.

81. The method of magnitude estimation
- is a psychophysiological method.
 - is a psychophysical method.
 - usually produces a logarithmic relation of physical intensity to perceived magnitude.
 - usually produces a logarithmic relation of physiological and psychological processes.
82. One goal of factor analysis is to
- see which questions on a test are really testing the same thing.
 - help the factor solve its personal problems.
 - eliminate intercorrelation of the test items.
 - minimize "faking" by test takers.
83. Place-learning experiments, blocked-response experiments, and changed-response experiments all indicate that
- S-S learning by animals is possible.
 - learning cannot take place in the absence of reinforcement.
 - S-S learning by animals is not possible.
 - need-reduction reinforcement is necessary for learning.
84. Dark-reared cats
- never develop a strong preference for the shallow side of a visual cliff.
 - require several days to develop a strong preference for the shallow side of a visual cliff.
 - show an immediate strong preference for the shallow side of a visual cliff.
 - tend to fall off a visual cliff.
85. Monkeys will learn a discrimination when pressing on one of two panels leads to the opportunity to look around, while pressing on the other panel does not lead to this. This experiment, as described above, shows that curiosity
- can be used as a reinforcer.
 - is a learned drive.
 - can be satiated.
 - is increased by secondary reinforcement.
86. The difference between sociology and social psychology is that
- social psychology is a behavioral science but sociology is not.
 - sociology is a behavioral science but social psychology is not.

- c. social psychology is primarily interested in the effect of a group on an individual's behavior, whereas sociology is primarily interested in the structure and formal characteristics of the group.
 - d. sociology is primarily interested in the effect of a group on an individual's behavior, whereas social psychology is primarily interested in the structure and formal characteristics of the group.
87. A reinforcer which is effective without prior reinforcement is known as a
- a. classical reinforcer.
 - b. primary reinforcer.
 - c. conditioned reinforcer.
 - d. secondary reinforcer.
88. An elephant is trained on a Lashley jumping stand to jump to a medium-gray square and not to a white square. In the test situation he is faced with the medium-gray square and a black square. He will (assuming elephants behave like other creatures) choose the
- a. gray square--because it is the same color as his aged mother.
 - b. black square--an example of transposition.
 - c. gray square--an example of transposition.
 - d. black square--an example of perceptual learning.
89. When a counselor advises someone, he generally does so on the basis of
- a. the single test which has been found most reliable in the past.
 - b. the outcome of a number of different tests and facts which guide impressions obtained in interviews, etc.
 - c. years of experience in interviewing--he does not use test results.
 - d. special aptitude tests only--intelligence tests are never used.
90. The absolute amount of transfer, whether positive or negative, is a function of
- a. stimulus similarity.
 - b. response substitution.
 - c. stimulus substitution.
 - d. response similarity.

91. Which of the following cues to depth cannot be used by a person blinded in one eye?
- interposition
 - linear perspective
 - retinal disparity
 - gradient of texture
92. The gradual weakening of response resulting from presentation of the CS without the UCS is called
- disconnection.
 - erasing.
 - extinction.
 - elimination.
93. Secondary goals may be
- acquired by classical conditioning.
 - acquired by instrumental learning.
 - involved in what the text calls "complex motivation."
 - all the above.
94. In the long history of nonscientific attempts to deal with behavior disorders, demon possession of the afflicted was one of the favorite explanations. The ritual treatment for demon possession was called
- exorcism.
 - phrenology.
 - animistic ejection.
 - tautology.
95. Weber's law states that
- the absolute size of the differential threshold depends on the location of the stimulus along its continuum.
 - the point of subjective equality depends on the location of the stimulus along its continuum.
 - the JND is not obtainable for stimuli below the absolute threshold.
 - reported experience is equal to $k \log I$.
96. Which of the below is not an advantage of programmed instruction?
- immediate knowledge of results.
 - forces the students to respond to the material.
 - forces the students to work at the same fast pace.
 - divides the material into small, easy steps.

97. Electroshock therapy has been found most useful in alleviating
- depression.
 - sexual problems.
 - disturbed thought patterns.
 - paranoia.
98. In general, behavior therapy may be said to aim at
- alleviating anxiety due to unconscious conflicts.
 - the more or less direct alteration of disordered behavior.
 - fostering a situation in which a person can express himself freely and experience "emotional growth."
 - the control of disordered behavior through the use of tranquilizers.
99. The scale of the Minnesota Multiphasic Personality Inventory (MMPI) measuring tendencies toward anti-social and amoral conduct is the _____ scale.
- paranoia (Pa)
 - psychopathic deviation (Pd)
 - psychasthenia (Pt)
 - poikilothermia (Pk)
100. The weeding out of test questions which do not discriminate between criterion groups is accomplished by
- face analysis.
 - item analysis.
 - factor analysis.
 - validity scales.
101. If we were studying errors in telephone dialing under the new all-digit system of dialing, a likely dependent variable would be the
- number of digits in the telephone number.
 - speed of all-numeral dialing.
 - number of errors made in dialing.
 - cost of converting to the all-numeral dialing system.
102. Suppose it takes 200 trials to learn a certain amount of material the first time; the second time it takes only 100 trials to relearn the same amount of material. Savings in this case is
- 50 percent.
 - 66 $\frac{2}{3}$ percent.
 - 100 percent.
 - 200 percent.

103. Analysis of typical long-term, laboratory-derived memory curves has indicated that much of the drop in the first day can be ascribed to
- reminiscence.
 - repression.
 - proactive interference.
 - assimilation.
104. The receptive field is
- the area of the receptor which will influence the activity of a particular sensory cell.
 - the surface area which encompasses the entire visual field of a normal human.
 - the sensitive area of the receptor surface.
 - a subfield within the area of psychology concerned with the study of sensory processes.
105. In the interpretation of an IQ score, it is important to keep many factors in mind. Which of the following is an incorrect statement about intelligence tests?
- Intelligence tests measure present ability, not native capacity.
 - Intelligence tests are essentially free of cultural bias.
 - Most general intelligence tests are heavily weighted with verbal material.
 - An IQ score from one intelligence test does not mean the same thing as an IQ score from another test of intelligence.

FORM B

QUESTIONS

1. Homeostasis is
 - a. illustrated by taking off your clothes when you're hot.
 - b. a form of sexual perversion practiced by chimpanzees.
 - c. illustrated by a person's hitting his head against a wall harder and harder.
 - d. the third stage of the motivational cycle.

2. Nonsense syllables are frequently used in experiments in verbal learning because
 - a. short-term forgetting for nonsense syllables is less than for words.
 - b. their association value can be evaluated more easily than that of words.
 - c. they lend themselves especially easily to serial presentation.
 - d. they cause less retroactive inhibition than words.

3. Product-moment correlation is to r as rank-difference correlation is to
 - a. ρ .
 - b. z .
 - c. X .
 - d. C .

4. The fovea is
 - a. the region of maximum visual acuity.
 - b. a raised place in the retina.
 - c. a region where there is much rhodopsin and no iodopsin.
 - d. another name for the blind spot.

5. Different patterns of visceral (internal) response have been shown in some studies of certain emotional states. The emotions in which different visceral patterns have been shown are
 - a. anxiety and fear.
 - b. resentment and anger.
 - c. fear and anger.
 - d. jealousy and hate.

6. A psychologist wishes to know what the effect of factor A is upon factor B. In order to find out, the

- psychologist varies factor A, and sees what happens to factor B. In this experiment
- a. factor A is the independent variable.
 - b. factor A is the dependent variable.
 - c. factor A is a single-blind control.
 - d. factor A is the theoretical construct.
7. If the distribution of scores in the diagram representing the relationship between two sets of variables is quite spread out, the correlation is likely to be
- a. perfectly negative.
 - b. perfectly positive.
 - c. low, and positive or negative.
 - d. normally distributed.
8. Learning curves
- a. always rise and never fall.
 - b. rise at first, but then fall back almost to the baseline.
 - c. might better be called performance curves.
 - d. plot errors on the horizontal axis.
9. The statement "If I am (like) the aggressor, I cannot be aggressed against" exemplifies best a variety of which defense mechanism?
- a. rationalization
 - b. reaction formation
 - c. identification
 - d. compensation
10. Which of the following is a true statement about theory?
- a. Sciences more advanced than psychology have developed to the point where they no longer need theory.
 - b. Theory always corresponds to reality.
 - c. Theories are useful for making predictions about things which have not yet been observed.
 - d. Theories always follow and summarize, rather than precede and guide, research.
11. The inner ear contains the
- a. ossicles.
 - b. auditory meatus.
 - c. cochlea.
 - d. both b and c above.
12. In general, psychiatrists differ from clinical psychologists in that they
- a. only diagnose behavioral disorders.
 - b. are concerned only with the id and ego.
 - c. hold medical degrees.
 - d. treat mental illness.

13. Experiments with inverting lenses demonstrate that
 - a. hallucinations depend upon sensory input.
 - b. visual perception is plastic and adaptable.
 - c. perception is plastic in early development, but becomes fixed later and cannot be changed.
 - d. prolonged distortion is harmful to the vision of the lens wearer--it produces retinal degeneration.
14. Maternal behavior in rats is increased most by injection of
 - a. estrogen.
 - b. prolactin.
 - c. thyroxin.
 - d. adrenalin.
15. A person who is autistic and who reports hallucinations would probably be classified as
 - a. suffering from a phobic reaction.
 - b. psychotic.
 - c. sociopathic.
 - d. suffering from a dissociative reaction.
16. The "achievement motive" is generally strongest in those who were reared in a home where great emphasis was placed on
 - a. dependence.
 - b. independence.
 - c. toilet training.
 - d. social interaction.
17. The development of the responses, interests, and behaviors appropriate to one's sex is called
 - a. sexual differentiation.
 - b. the masculine (or feminine) response.
 - c. syndrome development.
 - d. sex-typing.
18. A general rule, good control of variables in an experiment
 - a. makes nominal measurements possible.
 - b. allows one to obtain significant information from relatively small samples.
 - c. makes the use of large samples necessary.
 - d. results in greater variability of measures.
19. The incremental theory of learning maintains that in an experiment on paired-associate learning
 - a. associations are either at full strength or are of zero strength.
 - b. associations are gradually strengthened from trial to trial.

- c. associations are merely connections between sensory and motor neurons.
 - d. associations never reach maximum strength.
20. In perception, it would be said that the print on this page is _____; the paper (the page itself) is _____
- a. contoured; continuous.
 - b. figure; ground.
 - c. object; figure.
 - d. a group; a unity.
21. A recessive gene can express itself only when it is
- a. paired with a dominant gene.
 - b. paired with a recessive gene.
 - c. genotypic.
 - d. composed of DNA rather than RNA.
22. Which classes of words are learned first?
- a. verbs and pronouns
 - b. verbs and nouns
 - c. nouns and pronouns
 - d. adjectives and nouns
23. Sounds can "mask" each other--that is, a sound can prevent another sound from being heard. Which of the following is not one of the observed relations in masking?
- a. Sound A masks sound B more effectively if it is close to sound B in frequency.
 - b. Sound A masks sound B more effectively if it is of lower pitch than sound B.
 - c. Sound A masks sound B more effectively if it is of comparable timbre.
 - d. If tones of lower and higher pitch are sounded together, the lower pitched tone will be heard more easily--other things being equal.
24. If a person fears enclosed places because of a childhood trauma involving near suffocation, the person is displaying motive
- a. discrimination.
 - b. fixation.
 - c. generalization.
 - d. substitution.
25. The blood substances in sleep have
- a. been identified in transfusion experiments with dogs.
 - b. not been identified.
 - c. been shown to be crucial for sleep, although not identified, in observations on Siamese twins.

- d. been shown to come from the pituitary gland--
"the master gland"--although they have not been
identified.
26. The receptors for pain seem to be
- Meissner's corpuscles.
 - free-nerve endings.
 - Pacinian corpuscles.
 - bulbs of Krause.
27. As we move from lower animals to man, heredity seems
to have an influence on behavior through its effect
on
- aptitudes rather than complete instinctive
behaviors.
 - complete instinctive behaviors rather than apti-
tudes.
 - recessive determination rather than dominant deter-
mination.
 - dominant determination rather than recessive deter-
mination.
28. The establishment of a conditioned response to a neu-
tral stimulus through pairing of this stimulus with a
conditioned stimulus is called
- Pavlovian conditioning.
 - higher-order conditioning.
 - operant conditioning.
 - trace conditioning.
29. A child gradually learns that two equal volumes of
water remain equal no matter what the shape of the
vessels containing them. The principle the child uses
is known as
- egocentrism.
 - conservation.
 - formal operations.
 - concrete operations.
30. Ganglion cells in the retina
- contain retinene and opsin.
 - produce generator potentials.
 - produce "on," "off," and "on-off" responses.
 - are not influenced by the activity of cones.
31. Which two of the following characteristics of the
stimulus play a role in determining the perceived loud-
ness of a sound? (1) intensity (2) frequency (3) pitch
(4) timbre
- 1,2
 - 1,3
 - 1,4
 - 2,3

32. For a certain experiment a dog is placed in a harness and electrodes are attached to its leg. At random intervals a buzzer is sounded. The response to be conditioned is lifting the leg to which the electrodes are attached. If the dog lifts its leg within five seconds after the buzzer sounds it will receive no shock. If it does not lift its leg within the five-second period, it will receive a mild shock. This is an example of
- classical conditioning.
 - instrumental conditioning.
 - both classical and instrumental conditioning.
 - learning set.
33. The detection of the direction of a sound depends
- upon differences at the two ears.
 - only upon input into one ear--either one.
 - only upon input into the ear farthest from the sound source.
 - only upon input into the ear nearest the sound source.
34. In a laboratory experiment in a dark room, a fixed luminous dot in a luminous frame may be perceived as moving when the frame is physically moved. This is an example of
- beta movement.
 - autokinetic movement.
 - stationary-pattern movement.
 - induced movement.
35. In order to compute a z score, one needs to know the _____ of a distribution.
- median and mode
 - standard deviation and variability
 - mean and mode
 - standard deviation and mean
36. In Freud's theory of personality the instincts of the id are called the
- superego.
 - ego.
 - ego ideal.
 - libido.
37. The type of temporary amnesia which is accompanied by flight to another place is known as
- fugue.
 - psychasthenia.
 - the catastrophic reaction.
 - schizotaxia.

38. The prognosis for schizophrenic patients is
- generally hopeless.
 - related to the patient's adjustment before entering the hospital.
 - best for schizotaxic patients.
 - dependent upon the interaction of the effects of double-bind therapy with psychoanalytic therapy.
39. Schachter found the effect of an epinephrine injection depended upon
- the expected side effects.
 - the activities of a "stooge" partner to the subject.
 - only the dose level injected.
 - both a and b.
40. Alfred Adler split with Freud and developed a theory strongly emphasizing a drive for
- reduction of basic anxiety.
 - superiority.
 - self-actualization.
 - escape from freedom.
41. The sympathetic and parasympathetic systems are divisions of the
- primary nervous system.
 - autonomic nervous system.
 - epinephrine-norepinephrine system.
 - general nervous system.
42. The defense mechanism of repression is related most generally to
- reverting to infantile behavior.
 - motivated forgetting.
 - asserting a motivation opposite to the actual motivation.
 - shifting to a higher motivation than the actual motivation.
43. The Allport-Vernon-Lindzey scale measures
- aggression.
 - intelligence.
 - values.
 - mechanical aptitudes.
44. The same person may exhibit different emotional reactions to drugs in different situations. This is most easily explained by which theory of emotion?
- norepinephrine-epinephrine theory
 - cognitive theory
 - emergency theory
 - James-Lange theory

45. Experimental group:
 Control group:
 Learn 1 Learn 2 Measure retention of 1
 Learn 1 Rest Measure retention of 1
- This is one design for experiments on
- retroactive inhibition.
 - proactive inhibition.
 - negative transfer.
 - reorganization of the memory trace.
46. The items in a series which are hardest to learn are
- at the end of the list.
 - at the beginning of the list.
 - in the middle of the list.
 - none of the above; the serial position does not have an effect on difficulty of learning.
47. A rat is trained to go to the white arm rather than the black arm of a T maze in order to escape shock. It takes the rat 20 trials to learn this response. The maze is then filled with cold water and the rat must learn to swim to the white rather than the black arm in order to escape from the cold water. This second task, of course, requires a set of responses not required in the first task. You would expect the rat to
- take about 20 trials to learn this second task.
 - require a good many more than 20 trials to learn this second task.
 - take fewer than 20 trials to learn this second task.
 - drown.
48. Psychology is the science of
- human behavior.
 - behavior.
 - mental life.
 - thoughts and feelings.
49. The frequency of the auditory stimulus determines the
- amplitude of the sound wave.
 - kind of deafness.
 - position of maximum displacement along the basilar membrane.
 - location of the endocochlear potential.
50. Sampling error is always present in statistical studies. But it is usually less troublesome than error due to sample bias because it
- can be estimated if the sample size is known.
 - is smaller in magnitude.
 - is not relevant to the conclusions drawn.
 - is always normally distributed.

51. The dimension of visual experience which refers to the purity of a hue--the degree to which a hue is diluted or not diluted by grayness or whiteness--is known as
- brightness.
 - simultaneous contrast.
 - successive contrast.
 - saturation.
52. Which of the following was not characteristic of behaviorism?
- It rejected the introspective method.
 - It denied the existence of instinct and of inborn tendencies.
 - It denied the importance of the conditioned reflex as the basic unit of behavior.
 - It claimed that observable behavior was the thing to be studied by psychologists.
53. A person who develops an ulcer after being exposed to prolonged emotional stress is often said to be suffering from a/an
- anxiety reaction.
 - chronic epinephrine-release reaction.
 - psychosomatic reaction.
 - activation reaction.
54. In a normal distribution, the difference between the 25th and 50th centiles is
- less than that between the 50th and 75th centiles.
 - more than that between the 50th and 75th centiles.
 - equal to that between the 50th and 75th centiles.
 - zero.
55. The standard deviation of this set of numbers (5, 5, 5, 5, 5) is
- 5.
 - 25.
 - zero.
 - infinite.
56. According to one group of psychologists, certain learning experiments demonstrate perceptual reorganization and insight in animals. Which group is this?
- Skinnerians
 - Gestalt psychologists
 - Pavlovians
 - Instrumentalists
57. Harlow's "surrogate mother" study showed that
- the most important role of mother was as the source of food.

- b. "contact comfort" with the surrogate mother was important.
 - c. a food-giving mother was preferred over a soft mother.
 - d. the maternal drive of higher primates is learned.
58. Coefficients of contingency are usually used when the measures are
- a. interval measures.
 - b. nominal measures.
 - c. ratio measures.
 - d. ordinal measures.
59. The argument in the text about extinction maintains that extinction should be considered to be
- a. simply a weakening of the association formed during conditioning.
 - b. learning not to respond.
 - c. forgetting when to respond.
 - d. the same thing as spontaneous recovery.
60. Motives are
- a. directly observed.
 - b. independent of behavior.
 - c. unlearned; needs are learned.
 - d. inferred from behavior.
61. An experiment mentioned in the text showed that neural activity in the auditory system of a cat may be attenuated by
- a. contrast.
 - b. presentation of a jar of mice.
 - c. confusing figure and ground.
 - d. patterning the stimulation after a recent Beattle's recording.
62. The age of acquisition of the conservation principle is about
- a. 2 to 4 years.
 - b. 4 to 7 years.
 - c. 8 to 9 years.
 - d. 11 to 14 years.
63. In avoidance learning, the term latency refers to the elapsed time between the onset of the _____ and the beginning of the _____.
- a. conditioned stimulus, unconditioned stimulus.
 - b. unconditioned stimulus, conditioned stimulus.
 - c. conditioned stimulus, response.
 - d. response, conditioned stimulus.

64. A frequency of 10,000 cps is five times one of 2,000 cps. As the frequency increases five times between 2,000 cps and 10,000 cps, the perceived pitch increases by
- more than five times.
 - more than ten times.
 - less than two times.
 - between five and ten times.
65. The characteristic of Piaget's second stage of thought (preconceptual thought) is
- beginning of foresight.
 - conservation.
 - trial-and-error learning.
 - representational thought.
66. A normal one-month-old infant
- is practically blind.
 - is unable to focus well on objects which are close up.
 - can see about as well as a seven-year-old child.
 - is too young to be given visual acuity tests.
67. The term "emotion" is a complex one and has many facets. Which of the following is not included in the definition of emotion given by the text?
- outward expression
 - experience
 - situations producing emotion
 - bodily change.
68. Which are the two "chemical senses"? (1) hearing (2) smell (3) temperature (4) taste
- 1, 2
 - 2, 3
 - 2, 4
 - 1, 4
69. The type of afterimage which is the complement of the original stimulus color is known as the
- negative afterimage.
 - positive afterimage.
 - complementary afterimage.
 - opponent-process afterimage.
70. The perception of movement from A to B, which is seen by subjects shown light A following light B with certain time characteristics, illustrates
- induced movement.
 - zeta movement.
 - stroboscopic movement.
 - autokinetic movement.

71. The type of test most appropriate for giving an accurate evaluation of a foreign-born, non-English-speaking subject would most likely be a
- group test.
 - verbal test.
 - speed test.
 - performance test.
72. The sense of muscular movement and position of the limbs is called
- olfaction.
 - kinesthesia.
 - cathegogition.
 - the vestibular sense.
73. The tendency to value certain goals when alternative ones might do just as well is sometimes called motive
- generalization.
 - fixation.
 - displacement.
 - substitution.
74. According to the work of Funkenstein, which hormone is produced primarily during anger rather than during fear?
- epinephrine
 - norepinephrine
 - ACTH
 - mecholy
75. The gland which seems especially important in the general-adaptation syndrome is the
- adrenal gland.
 - parathyroid gland.
 - pineal gland.
 - thyroid gland.
76. The intelligence test was first developed by Alfred Binet in order to
- predict who could become capable government employees.
 - predict which children would not profit from attending school.
 - help distinguish between the effects of certain cultural enrichments.
 - show the superiority of the French school system.
77. Which of the following is not one of the ways of reducing conflict and frustration mentioned in the text?
- adherence to the "pleasure principle"
 - postponing satisfaction

- c. expressing emotion
 - d. increasing frustration tolerance
78. Fixation of response was observed in rats who were presented with
- a. an approach-approach conflict.
 - b. an insoluble discrimination.
 - c. a plexiglas barrier in a straight-alley runway.
 - d. extinction of a bar-press response.
79. Treatment arranged either to help a person through a crisis, or in a chronic case where the resources for health are poor, is often called
- a. insight therapy.
 - b. behavior therapy.
 - c. situation-change therapy.
 - d. supportive therapy.
80. An educational psychologist differs from a school psychologist in which of the following ways?
- a. The educational psychologist is concerned with general problems, while the school psychologist is concerned with problems of individual cases.
 - b. The educational psychologist deals with college students, while the school psychologist deals with high school students.
 - c. The school psychologist is likely to be an experimenter, while the educational psychologist is likely to be a counselor.
 - d. None of the above.
81. Schizophrenia usually appears
- a. in early childhood.
 - b. in late teens and early twenties.
 - c. in forties or fifties.
 - d. with about equal frequency for people of all ages.
82. A paranoid reaction is best characterized by the word(s)
- a. withdrawal.
 - b. anhedonia.
 - c. delusions.
 - d. affective psychosis.
83. In judging which emotion is being expressed by a person, we are most likely to be accurate when we
- a. know only the situation in which the emotion occurs.
 - b. see only the facial expression of the emotional person.
 - c. see only the posture of the emotional person.
 - d. have a record of the GSR of the emotional person.

84. Here is a partial list of nonsense syllables for a verbal learning experiment.

S	R
ZEG	
ZEG	XAR
DAQ	
DAQ	ZUY

What kind of list is this?

- serial-anticipation list
 - consonant-diagram list
 - noncontinuous list
 - paired-associate list
85. The coefficient which represents a moderate degree of correlation is
- 1.00.
 - 0.50.
 - +1.00.
 - +2.50.
86. It has been proposed that the psychological causes, or psycho-dynamics, of antisocial behavior are to be found in
- excessive repression.
 - failure of superego development.
 - oral fixation.
 - unresolved anxiety.
87. Psychoneuroses are best described as
- antisocial reactions.
 - anxiety produced.
 - brain disorders.
 - loss of contact with reality.
88. Transduction is the
- conversion of physical measurement to psychological measurement.
 - explanation of psychological laws in terms of physiological laws.
 - change imposed on neural activity as it proceeds up the sensory system toward the brain.
 - conversion of physical energy into energy which will generate neural activity.
89. One theory of emotion maintains that the emotional bodily changes and the felt emotion occur simultaneously. Another theory maintains that the felt emotion depends upon the perception of the bodily state. This theory stressing perception of the bodily state is the
- James-Lange theory.
 - emergency theory.
 - psychosomatic theory.

90. According to the two-factor theory, avoidance learning is hard to extinguish because
- the animal seldom stays in the situation long enough to experience the lack of pairing of CS and US.
 - the animal is never reinforced for avoiding.
 - partially reinforced avoidance responses are less resistant to extinction than continuously avoidance-reinforced responses.
 - punishment during extinction suppresses spontaneous recovery.
91. One advantage of standard scores is that they
- are negative as well as positive.
 - make comparisons between scores on different tests possible.
 - are symmetrical.
 - never go over +1.96.
92. Studies of family influences on the development of personality in children have shown that
- imitation and modeling are not important.
 - children will not imitate specific aggressive actions, but will become generally aggressive.
 - children from controlled home atmospheres tend to be more socially aggressive than children from democratic home atmospheres.
 - Children from controlled home atmospheres tend to be more conforming than children from democratic home atmospheres.
93. Anger is most commonly caused by
- fear of failure.
 - strangeness of a situation.
 - frustration of an activity.
 - an impending aversive event.
94. People with "hysterical personalities" often
- seem unusually naive.
 - are easily disturbed emotionally.
 - display la belle indifference.
 - all of the above.
95. Complex motives are
- like physiological drives in that they are primarily unlearned.
 - always unconscious.
 - often the result of instrumental learning.
 - dependent upon fixation for their development.

96. A question is scientifically meaningless when it cannot be answered by
- logic.
 - observation.
 - authority.
 - intuition.
97. According to the definitions given in the text, lack of satisfaction of a need is called _____; blocking of behavior directed toward a goal is called _____.
- approach-approach conflict; avoidance-avoidance conflict
 - motivation; conflict
 - frustration; conflict
 - deprivation; frustration
98. The term "psychoanalyst"
- is a general one covering all psychotherapeutic endeavors.
 - is synonymous with the term "clinical psychologist."
 - should be reserved for those who apply therapeutic techniques derived from psychoanalysis.
 - is synonymous with the term "behavior therapist."
99. During free association, a patient often is unable to remember significant events in his past or to talk freely about anxiety-charged subjects. A part of the psychoanalyst's task is the overcoming of such
- manifest contents.
 - unfree associations.
 - resistances.
 - blockages in extinction.
100. One theory of the mechanism of the sense of smell is called the
- corpuscle theory.
 - "lock-and-key" theory.
 - telephone theory.
 - opponent-process theory.
101. Operational definitions
- reduce misunderstandings.
 - are required in medical work.
 - are imprecise.
 - restrict affective communication between scientists.
102. In determining a threshold, a number of methods are available. When one presents a stimulus, asks for a report, and presents the next stimulus independently

of what this report is, the method of _____ is being used.

- a. adjustment
 - b. constant stimuli
 - c. limits
 - d. approximation
103. The "critical," or immediate, event in starting activity in the auditory nerve is
- a. resonance of the ossicles.
 - b. movement of the round window.
 - c. shearing movements of the hair cells.
 - d. tension on rods and cones.
104. Discrimination learning involves
- a. extinguishing responses which occur because of stimulus generalization.
 - b. continuous reinforcement of responses made to the S.
 - c. simultaneous, but not successive, presentation of the positive and negative stimuli.
 - d. all of the above.
105. In the single-blind technique
- a. the experimenter does not know the experimental group to which the subject belongs.
 - b. the subject does not know the experimental group to which he belongs.
 - c. both of the above
 - d. neither of the above

APPENDIX B

INSTRUCTIONS AND ANSWER SHEET

INSTRUCTIONS

Perceived Difficulty Ratings of Psychology I
Multiple-choice QuestionsInstructions

On the following pages you will find a number of Psychology I multiple-choice questions. The investigator is interested in the subjective difficulty level of typical multiple-choice questions used in Psychology I examinations; i.e., the degree of difficulty a question provokes in the student when he encounters it in the examination setting. Your job is to rate each of the following multiple-choice questions on an 11-point difficulty scale ranging from "extremely easy" (scale value - 1) through "average difficulty" (scale value - 6) to "extremely difficult" (scale value - 11). In other words, you are rating each multiple-choice question on the degree of subjective difficulty one would experience in reaching a solution to that particular question no matter whether that solution is correct or incorrect. For example, the following question:

1. Psychology is a(an) _____ .
 - a. Science
 - b. Art
 - c. Natural Science
 - d. Medical Science

would probably get a difficulty rating somewhere toward the lower end of the scale (see below) since this question appears to be rather "easy."

Difficulty Scale Continuum

Extremely Easy			Average					Extremely Difficult		
1	2	3	4	5	6	7	8	9	10	11
0	0	0	0	0	0	0	0	0	0	0

Indicate the following multiple-choice question scale scores by blackening in a circle under the desired scale number. For example, if you had believed that

the above multiple-choice question provokes an extremely easy subjective feeling when encountered by students on a test you would have possibly blackened in the circle under number 2.

Before you begin to rate the multiple-choice questions spend a few minutes scanning them over so as to become familiar with the general range of difficulty that they cover and represent.

It is extremely important that you rate the question according to the amount of subjective difficulty you think it provokes when encountered by students in the examination setting. The investigator appreciates your cooperation in performing this task. If you decide (for some unknown reason) to randomly place marks on the answer sheet, etc., the investigator would rather have you turn in your answer sheet blank. You still will be given credit. Of course, it is hoped that you will cooperate.

If you are not sure where to rate a question try hard to make the best judgment that you possibly can. If you have any questions raise your hand.

Thank you for your cooperation.

APPENDIX C

**STUDENT AND INSTRUCTOR RATINGS OF ITEM -
DIFFICULTY: MEDIAN INTENSITY SCORES
AND SEMI-INTERQUARTILE RANGES**

Question Number Form A	Student Ratings (N = 71)		Instructor Ratings (N = 9)	
	Mdn	Q	Mdn	Q
1	6.10	1.97	5.00	1.67
2	6.95	1.54	8.25	1.31
3	6.43	1.56	6.67	.90
4	3.56	1.34	4.75	1.31
5	7.00	1.46	8.25	1.06
6	4.41	1.91	5.00	2.62
7	3.91	1.71	5.25	1.31
8	4.75	1.95	5.00	1.88
9	5.57	1.78	5.00	2.17
10	4.00	1.54	6.33	.90
11	1.50	0.96	1.14	.32
12	6.65	1.19	3.67	1.14
13	6.92	1.46	8.33	.90
14	5.12	1.58	6.00	2.56
15	4.75	1.67	4.00	2.38
16	4.38	2.17	3.33	2.14
17	6.28	1.59	7.25	1.31
18	6.44	1.39	6.25	1.12
19	5.70	1.48	7.00	2.06
20	6.32	2.03	3.12	.84
21	6.10	1.25	4.00	2.56
22	6.21	1.03	7.75	1.31
23	6.10	2.10	7.60	1.07
24	5.70	1.80	5.00	1.31
25	3.41	2.13	3.33	1.33
26	5.17	1.88	4.00	2.50
27	1.42	.91	1.75	1.09
28	5.60	1.43	5.25	1.31
29	4.77	1.68	3.88	1.09
30	7.58	1.67	8.00	1.75
31	4.62	2.05	4.00	1.81
32	4.50	2.64	2.00	.81
33	6.14	1.75	7.75	1.31
34	6.14	1.87	5.75	2.06
35	7.07	1.74	7.00	1.48
36	5.00	1.93	3.75	1.88
37	7.50	1.80	6.00	2.56
38	2.75	2.07	3.25	2.06
39	5.38	1.48	5.75	2.06
40	4.78	1.77	5.00	1.75
41	3.68	1.22	6.12	1.16
42	5.78	1.34	6.12	.66
43	5.97	1.42	6.33	1.08
44	6.07	1.62	7.00	1.25
45	5.71	1.89	6.00	2.06
46	7.20	1.60	7.67	.90
47	6.25	1.72	8.12	.59
48	7.03	1.45	8.75	1.06
49	6.35	1.07	7.67	1.40
50	6.08	1.39	7.00	1.25

Question Number Form A	Student Ratings (N = 71)		Instructor Ratings (N = 9)	
	Mdn	Q	Mdn	Q
51	5.42	1.35	4.33	1.40
52	7.60	1.33	7.00	1.48
53	6.06	1.75	4.33	1.33
54	6.38	1.26	6.67	1.64
55	5.33	1.53	4.00	1.06
56	7.87	1.16	7.25	1.48
57	7.14	1.06	8.00	1.23
58	5.74	1.71	5.33	1.14
59	5.85	1.40	4.25	2.25
60	6.57	1.62	7.00	2.67
61	4.61	1.79	6.25	1.12
62	3.78	1.52	4.33	1.75
63	6.23	1.74	6.00	1.25
64	6.95	1.48	5.67	1.33
65	5.73	2.00	7.67	3.33
66	7.10	1.17	7.67	.90
67	7.21	1.33	9.25	.84
68	6.50	1.36	6.00	1.91
69	6.18	1.21	7.00	1.75
70	7.00	1.29	7.75	1.38
71	6.93	1.67	5.75	2.17
72	6.30	1.56	5.33	1.33
73	6.30	1.55	8.12	.66
74	5.41	1.31	4.00	1.48
75	6.50	1.31	7.25	1.31
76	5.61	2.11	5.75	1.38
77	6.12	1.76	6.38	1.03
78	6.57	1.10	6.12	.84
79	7.80	1.92	7.75	1.81
80	5.68	1.39	5.67	1.64
81	7.37	1.49	8.00	1.00
82	6.50	1.14	7.00	1.00
83	7.25	1.33	7.38	.78
84	7.38	1.28	7.62	.56
85	6.05	.92	6.00	1.08
86	6.36	1.49	4.33	2.08
87	5.66	1.28	4.25	1.38
88	6.35	1.21	7.67	.90
89	5.25	1.32	4.00	1.62
90	6.58	1.24	8.00	1.81
91	6.11	1.70	5.00	1.62
92	6.06	1.42	3.00	1.06
93	6.00	1.24	6.33	.68
94	5.95	2.38	5.75	1.42
95	8.14	1.33	8.33	.75

Question Number Form A	Student Ratings (N = 71)		Instructor Ratings (N = 9)	
	Mdn	Q	Mdn	Q
96	6.68	1.30	4.00	1.81
97	5.94	1.24	5.00	1.62
98	6.36	1.09	5.75	1.38
99	7.38	1.34	7.67	1.90
100	6.41	1.50	6.12	.66
101	5.92	1.48	6.25	1.31
102	5.12	1.76	6.00	1.50
103	6.78	1.11	6.33	.90
104	7.14	1.78	6.75	1.17
105	6.06	1.36	6.25	1.50

Question Number Form B	Student Ratings (N = 71)		Instructor Ratings (N = 9)	
	Mdn	Q	Mdn	Q
1	3.13	1.72	3.12	1.34
2	6.81	1.23	5.75	1.50
3	9.20	1.90	7.00	1.42
4	4.62	2.08	5.62	1.56
5	6.90	1.38	6.00	1.81
6	4.83	2.04	5.75	1.42
7	5.94	1.56	6.00	3.06
8	5.73	1.11	7.00	1.23
9	5.36	1.67	7.00	1.23
10	5.59	1.35	6.88	.66
11	4.32	1.69	5.00	2.00
12	4.33	1.72	2.33	1.33
13	6.77	1.36	6.00	1.41
14	5.93	1.86	6.00	1.62
15	6.12	1.70	5.67	1.90
16	4.28	1.59	5.67	1.40
17	5.50	1.64	6.25	1.31
18	6.27	1.82	6.00	1.38
19	7.88	1.33	6.67	1.14
20	5.50	1.81	4.00	1.88
21	3.00	1.62	4.67	1.40
22	4.12	1.97	7.12	.59
23	8.12	1.38	7.67	.90
24	5.76	1.61	6.67	1.08
25	7.50	1.77	7.00	.81
26	5.59	2.32	6.25	1.31
27	7.08	1.50	7.67	1.14
28	6.37	1.26	4.33	1.83
29	6.37	1.31	7.25	.98
30	7.17	1.25	7.88	.84
31	6.17	1.50	6.25	1.67
32	5.50	1.55	6.25	1.31
33	6.12	1.08	5.75	1.62
34	7.12	1.34	5.25	1.12
35	6.83	1.45	6.25	2.31
36	4.09	1.58	5.00	1.88
37	4.38	1.88	5.33	.90
38	6.00	1.54	5.75	1.56
39	7.85	1.15	8.25	1.67
40	7.04	1.67	4.33	1.33
41	3.70	1.90	3.25	1.56
42	4.94	1.66	4.12	1.16
43	6.50	1.80	6.60	.71
44	7.56	1.38	6.33	1.14
45	8.25	1.38	7.00	1.67

Question Number Form B	Student Ratings (N = 71)		Instructor Ratings (N = 9)	
	Mdn	Q	Mdn	Q
46	5.93	1.35	5.00	1.25
47	7.06	1.14	6.00	1.31
48	2.06	1.33	1.40	.96
49	6.83	1.53	6.12	.84
50	6.69	1.04	7.33	.75
51	6.25	1.27	6.00	1.38
52	6.96	1.23	6.67	1.33
53	5.71	1.21	5.67	2.14
54	6.95	1.71	6.67	1.08
55	3.07	2.28	6.00	1.31
56	6.68	1.34	5.00	1.23
57	5.30	1.44	5.33	1.18
58	7.62	1.27	8.38	.78
59	6.45	1.04	7.20	.75
60	5.66	1.38	6.67	.75
61	6.57	1.52	5.75	1.38
62	7.17	1.76	8.00	1.16
63	7.23	1.43	6.33	1.14
64	8.50	1.28	9.67	1.18
65	7.62	1.37	8.62	.97
66	5.50	1.51	6.75	1.17
67	5.79	1.30	6.75	1.75
68	4.04	1.54	4.00	1.62
69	6.41	1.62	4.33	1.75
70	7.38	1.48	7.25	1.17
71	4.88	1.77	5.67	1.08
72	4.94	1.70	4.25	1.56
73	5.64	1.17	5.88	.84
74	6.72	1.58	7.25	1.48
75	7.03	1.27	6.00	2.56
76	4.50	1.70	4.33	1.40
77	6.12	1.25	7.00	1.56
78	6.50	1.08	6.67	.90
79	6.30	1.21	5.88	.66
80	6.11	1.32	6.25	1.75
81	4.95	1.49	5.75	1.48
82	5.23	1.29	4.00	2.67
83	6.10	1.55	7.00	.45
84	8.00	1.75	7.00	1.98
85	6.39	1.63	6.33	.90
86	6.69	1.18	6.33	1.08
87	6.00	1.48	4.75	1.56
88	6.09	1.37	5.00	1.81
89	6.57	1.18	5.00	1.23
90	7.14	1.37	6.12	.84

Question Number
Form B

Student Ratings
(N = 71)

Instructor Ratings
(N = 9)

	Mdn	Q	Mdn	Q
91	6.29	1.09	6.33	1.33
92	5.85	1.32	7.00	1.38
93	4.75	1.50	5.00	1.38
94	5.50	1.00	6.00	.81
95	6.05	1.26	7.00	1.16
96	5.58	1.50	5.25	1.75
97	5.50	1.40	6.00	1.88
98	5.75	1.56	5.00	1.06
99	5.56	1.41	4.88	.84
100	6.22	1.43	5.67	1.40
101	5.85	1.34	4.25	1.25
102	6.91	1.32	6.67	.75
103	7.61	1.14	6.75	1.56
104	6.75	1.18	7.12	.84
105	4.17	1.74	5.75	2.06

APPENDIX D

**FINAL EXAMINATION QUESTIONS, STUDENT PERCEIVED
ITEM-DIFFICULTY VALUES (MDN), AND
ITEM FORMAT POSITIONS**

Question (See Appendix A)	Student Perceived Item-Difficulty Value	Question Number on Form		
	Mdn	H-E	E-H	R
Form B - 3	9.20	1	100	54
Form B - 64	8.50	2	99	69
Form B - 45	8.25	3	98	83
Form A - 95	8.14	4	97	32
Form B - 23	8.12	5	96	95
Form B - 84	8.00	6	95	51
Form B - 19	7.88	7	94	45
Form A - 56	7.87	8	93	28
Form B - 39	7.85	9	92	17
Form B - 58	7.62	10	91	80
Form B -103	7.61	11	90	4
Form A - 52	7.60	12	89	36
Form A - 30	7.58	13	88	25
Form B - 44	7.56	14	87	34
Form B - 25	7.50	15	86	21
Form A - 84	7.38	16	85	71
Form A - 99	7.38	17	84	31
Form B - 70	7.38	18	83	87
Form A - 83	7.25	19	82	33
Form A - 67	7.21	20	81	62
Form B - 30	7.17	21	80	76
Form B - 90	7.14	22	79	52
Form A - 57	7.14	23	78	9
Form B - 34	7.12	24	77	55
Form A - 66	7.10	25	76	37
Form B - 47	7.06	26	75	40
Form A - 70	7.00	27	74	60
Form A - 5	7.00	28	73	20
Form B - 52	6.96	29	72	78
Form B -102	6.91	30	71	22
Form B - 2	6.81	31	70	3
Form A -103	6.78	32	69	35
Form B -104	6.75	33	68	58
Form B - 50	6.69	34	67	91
Form B - 86	6.69	35	66	99
Form A - 78	6.57	36	65	14
Form B - 78	6.50	37	64	39
Form B - 59	6.45	38	63	38
Form A -100	6.41	39	62	57
Form A - 54	6.38	40	61	26
Form A - 98	6.36	41	60	85
Form A - 49	6.35	42	59	42
Form B - 79	6.30	43	58	66
Form B - 91	6.29	44	57	97
Form A - 22	6.21	45	56	41

Question (See Appendix A)	Student Perceived Item-Difficulty Value	Question Number on Form		
	Mdn	H-E	E-H	R
Form A - 33	6.14	46	55	74
Form B - 33	6.12	47	54	81
Form B - 77	6.12	48	53	29
Form A -105	6.06	49	52	61
Form A - 85	6.05	50	51	56
Form A - 93	6.00	51	50	92
Form A - 97	5.94	52	49	8
Form B - 46	5.93	53	48	30
Form B - 92	5.85	54	47	7
Form A - 42	5.78	55	46	49
Form B - 8	5.73	56	45	82
Form B - 53	5.71	57	44	6
Form B - 60	5.66	58	43	48
Form A - 87	5.66	59	42	44
Form B - 99	5.56	60	41	72
Form B - 94	5.50	61	40	89
Form B - 97	5.50	62	39	5
Form A - 51	5.42	63	38	65
Form A - 74	5.41	64	37	96
Form A - 39	5.38	65	36	98
Form A - 55	5.33	66	35	1
Form A - 89	5.25	67	34	100
Form B - 82	5.23	68	33	94
Form A - 26	5.17	69	32	70
Form A -102	5.12	70	31	68
Form A - 14	5.12	71	30	19
Form B - 81	4.95	72	29	10
Form B - 72	4.94	73	28	53
Form B - 42	4.94	74	27	16
Form B - 71	4.88	75	26	23
Form A - 40	4.78	76	25	24
Form A - 29	4.77	77	24	46
Form B - 93	4.75	78	23	13
Form A - 15	4.75	79	22	75
Form A - 61	4.61	80	21	84
Form B - 76	4.50	81	20	88
Form A - 6	4.41	82	19	50
Form B - 37	4.38	83	18	11
Form B - 12	4.33	84	17	67
Form B - 11	4.32	85	16	86
Form B - 16	4.28	86	15	15
Form B -105	4.17	87	14	2
Form B - 36	4.09	88	13	90
Form B - 68	4.04	89	12	63
Form A - 10	4.00	90	11	93

Question (See Appendix A)	Student Perceived Item-Difficulty Value	Question Number on Form		
	Mdn	H-E	E-H	R
Form A - 7	3.91	91	10	64
Form A - 62	3.78	92	9	47
Form B - 31	3.70	93	8	27
Form A - 41	3.68	94	7	43
Form A - 4	3.56	95	6	73
Form B - 1	3.13	96	5	77
Form B - 21	3.00	97	4	59
Form B - 48	2.06	98	3	12
Form A - 11	1.50	99	2	18
Form A - 27	1.42	100	1	79

APPENDIX E

QUESTIONNAIRE BOOKLET

INSTRUCTIONS

On the following pages are two questionnaires which we would like you to fill out. The first questionnaire is the General Affective State Form and the second questionnaire is the Test Evaluation Form. We would appreciate your honest response to these forms. The results will be used for research purposes and in no way will affect your grade in the course. You need not sign your name to this booklet. Just raise your hand when you are finished and a proctor will collect the forms. Thank you.

GENERAL AFFECTIVE STATE
QUESTIONNAIRE

INSTRUCTIONS

On the following page is a list of words and phrases which can be used to describe your feelings. Please check the word or phrase which best describes the way you NORMALLY FEEL. So that you will become familiar with the general range of feeling that they cover or represent, read the entire list before making your selection. Check only one word or phrase.

- _____ DISTRESSED
- _____ UNRUFFLED
- _____ THREATENED
- _____ AT EASE
- _____ TIMID
- _____ EXTREMELY TERRIFIED
- _____ FEARFUL
- _____ UNEASY
- _____ MARVELOUS
- _____ ALRIGHT
- _____ NOT MATTERING
- _____ THRILLED
- _____ FEELING GOOD
- _____ SCARED STIFF
- _____ KEEN

INSTRUCTIONS

On the following page is a list of words and phrases which can be used to describe your feelings. Please check the word or phrase which best describes the way you generally felt while taking your final examination. So that you will become familiar with the general range of feeling that they cover or represent, read the entire list before making your selection. Check only one word or phrase.

_____ DISTRESSED
_____ UNRUFFLED
_____ THREATENED
_____ AT EASE
_____ TIMID
_____ EXTREMELY TERRIFIED
_____ FEARFUL
_____ UNEASY
_____ MARVELOUS
_____ ALRIGHT
_____ NOT MATTERING
_____ THRILLED
_____ FEELING GOOD
_____ SCARED STIFF
_____ KEEN

TEST EVALUATION FORM

1. Which of the following choices best describes your feelings concerning the "difficulty level" of your final examination. Check one.
- One of the hardest I've ever taken.
 - Hard, but not so bad.
 - About average, like most tests I've taken.
 - Easier than most tests, but no snap.
 - One of the easiest classroom tests I've had.
2. Which of the following choices best describes your feelings concerning the "fairness level" of your final examination. Check one.
- Straightforward - what information was desired by the questions was clearly indicated.
 - Usually clear in asking the questions - tricky and ambiguous rarely.
 - About as fair as the average college test.
 - More ambiguous and tricky than most tests.
 - Very tricky - you had to be on guard all the time.
3. Which of the following choices best describes your feelings concerning the "adequacy of coverage of the course material" by your final examination. Check one.
- Thoroughly - gave a good chance to show what you know.
 - Better than average - few obvious omissions in coverage.
 - About average.
 - Poorer than many - major areas were never touched.
 - Was this test made for this course?

APPENDIX F

ACHIEVEMENT ANXIETY TEST

INSTRUCTIONS

Indicate the degree to which each item applies to you by circling the desired number.

For example: I like animals.

1	2	3	4	5
Not at all			Very much	

1. Nervousness while taking an exam or test hinders me from doing well.

5	4	3	2	1
Always			Never	

2. I work most effectively under pressure, as when the task is very important.

5	4	3	2	1
Always			Never	

3. In a course where I have been doing poorly, my fear of a bad grade cuts down my efficiency.

1	2	3	4	5
Never			Always	

4. When I am poorly prepared for an exam or test, I get upset, and do less well than even my restricted knowledge should allow.

1	2	3	4	5
This never happens to me			This prac- tically always hap- pens to me	

5. The more important the examination, the less well I seem to do.

5	4	3	2	1
Always			Never	

6. While I may (or may not) be nervous before taking an exam, once I start, I seem to forget to be nervous.

5	4	3	2	1
I always forget			I am always nervous during an exam	

7. During exams or tests, I block on questions to which I know the answers, even though I might remember them as soon as the exam is over.

5	4	3	2	1
This always happens to me			I never block on questions to which I know the answers	

8. Nervousness while taking a test helps me do better.

1	2	3	4	5
It never helps			It often helps	

9. When I start a test, nothing is able to distract me.

5	4	3	2	1
This is always true of me			This is not true of me	

10. In courses in which the total grade is based mainly on one exam, I seem to do better than other people.

1	2	3	4	5
Never			Almost always	

11. I find that my mind goes blank at the beginning of an exam, and it takes me a few minutes before I can function.

5	4	3	2	1
I almost always blank out at first			I never blank out at first	

12. I look forward to exams.

1	2	3	4	5
Never			Always	

13. I am so tired from worrying about an exam, that I find I almost don't care how well I do by the time I start the test.

1	2	3	4	5
I never feel this way			I almost always feel this way	

14. Time pressure on an exam causes me to do worse than the rest of the group under similar conditions.

5	4	3	2	1
Time pres- sure always seems to make me do worse on an exam than others			Time pres- sure never seems to make me do worse on an exam than others	

15. Although "cramming" under pre-examination tension is not effective for most people, I find that if the need arises, I can learn material immediately before an exam, even under considerable pressure, and successfully retain it to use on the exam.

5	4	3	2	1
I am always able to use the "crammed" material successfully			I am never able to use the "crammed" material successfully	

16. I enjoy taking a difficult exam more than an easy one.

5	4	3	2	1
Always			Never	

17. I find myself reading exam questions without understanding them and I must go back over them so that they will make sense.

1	2	3	4	5
Never			Almost always	

18. The more important the exam or test, the better I seem to do.

5	4	3	2	1
This is true of me			This is not true of me	

19. When I don't do well on a difficult item at the beginning of an exam, it tends to upset me so that I block on even easy questions later on.

1	2	3	4	5
This never happens to me			This almost always happens to me	

APPENDIX G

**TABLES OF MEANS, STANDARD DEVIATIONS
AND SAMPLE SIZES**

TABLE 9

Means and Standard Deviations of All Cells
Involved in the Analysis of
Variance in Table 3

Group	N	M	S.D.
H-E Sequence	44	57.73	13.23
E-H Sequence	44	60.73	10.82
R Sequence	45	57.56	14.03

TABLE 10

Means and Standard Deviations of All Cells
Involved in the Analysis of
Variance in Table 4

Group	N	M	S.D.
H-E Sequence	43	7.40	1.81
E-H Sequence	44	8.18	2.16
R Sequence	5	8.10	2.27

TABLE 11

Means and Standard Deviations of All Cells
Involved in the Analysis of
Variance in Table 5

Group	N	M	S.D.
H-E Sequence	42	1.74	.73
E-H Sequence	42	2.19	.86
R Sequence	45	1.84	.67

TABLE 12

Means and Standard Deviations of All Cells
Involved in the Analysis of
Variance in Table 6

Group	N	M	S.D.
H-E Sequence	42	3.02	.95
E-H Sequence	42	2.43	.83
R Sequence	45	2.51	1.06

TABLE 13

Means and Standard Deviations of All Cells
Involved in the Analysis of
Variance in Table 7

Group	N	M	S.D.
H-E Sequence	42	2.74	1.27
E-H Sequence	42	2.43	.83
R Sequence	44	2.43	1.09

TABLE 14
Means and Standard Deviations of All Cells
Involved in the Analysis of
Variance in Table 8*

Group	Achievement Anxiety Type			
	<u>Facil.</u>	<u>Debil.</u>	<u>High-affect.</u>	<u>Nonaffect.</u>
H-E Sequence	67.25 9.94 8	58.88 10.96 8	53.88 10.29 8	57.00 14.79 8
E-H Sequence	69.50 9.87 8	59.88 9.70 8	67.12 12.06 8	53.38 8.59 8
R Sequence	71.00 7.52 8	45.38 13.32 8	53.88 15.37 8	60.50 10.60 8

*The three numbers in each cell are the mean, standard deviation, and sample size, respectively.