1971

# Teacher-centered vs. student-centered mode of college classroom instruction as related to individual differences.

Fred J. Dowaliby
*University of Massachusetts Amherst*

# TEACHER-CENTERED vs. STUDENT-CENTERED MODE OF COLLEGE CLASSROOM INSTRUCTION AS RELATED TO INDIVIDUAL DIFFERENCES

A thesis Presented

By

Fred J. Dowaliby

Submitted to the Graduate School of the
University of Massachusetts in
partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 1971

Major Subject Psychology

# TEACHER-CENTERED vs. STUDENT-CENTERED MODE OF COLLEGE CLASSROOM INSTRUCTION AS RELATED TO INDIVIDUAL DIFFERENCES

A Thesis

By

Fred J. Dowaliby

Approved as to style and content by:

Harry Schuman
(Chairman of Committee)

Rudolph L. Pettit
(Head of Department)

Ronald K Hambleton
(Member)

[illegible] Roger
(Member)

May 1971

Acknowledgements

The writer wishes to express his appreciation to the members of his thesis committee: Dr. Ronald Hambleton, Dr. James Royer, and Dr. Harry Schumer, for their thoughtful criticisms and suggestions throughout the entire study.

Thanks is due Dr. David Berliner for his invaluable assistance in the early conceptualization of the study.

Additionally, the writer would like to thank his wife Anna for her devoted assistance throughout.

# TABLE OF CONTENTS

# TABLE OF TABLES

# TABLE OF FIGURES

# INTRODUCTION

Studies comparing different methods of college classroom instruction have typically resulted in the conclusion that learning, as measured by exam performance, is not differentially effected by different classroom instructional techniques. Although different studies have compared different methods with each other (e.g., lecture vs. group centered, didactic vs. discussion) the consistent treatment-group distinction has been in terms of teacher-centeredness vs. student-centeredness.

## Teacher-centered vs. Student-centered Instruction

The results of studies investigating the relative merits of student-centered (SC) vs. teacher-centered (TC) methods of college classroom instruction have typically indicated no difference in terms of examination scores. In a recent survey of the research reported in this area, Dubin and Taveggia (1968) concluded that the "data demonstrate clearly and unequivocally that there is no difference among truly distinctive methods of college instruction when evaluated by student performance on final examinations." The general design common to studies that have examined the effects of different modes of instruction has been the imposition of the different teaching techniques to different sections of an ongoing college course. The criterion measure by which the methods of instruction have been evaluated is course examinations.

In a very early study (Gerberich & Warner, 1936) the investigators compared TC to SC for two semesters with different matched groups of students. There was no difference between the TC and SC groups on an objective final examination given at the conclusion of each semester, indicating no difference between the TC and SC modes in terms of student performance.

In a more recently reported study (Eglash, 1954) the experimenter imposed material growing out of class discussions in the SC condition on to the TC group in an attempt to more precisely control the experimental situation. The results from this study, consistent with those of Gerberich and Warner, indicated no difference between the TC and SC methods in terms of exam score.

Haigh and Schmidt (1956) reported a study wherein the Ss were permitted to choose between a TC and a SC section of the same class. The experimenters' hypothesis that differences would become evident when students' stated preferences were taken into consideration was not supported by the results, which revealed no difference between the TC and SC sections in terms of final exam performance.

Another study that compared TC, SC, and an eclectic approach (actually a combination of the TC and SC modes) resulted in no differences among the mean exam scores (Krumboltz & Farquhar, 1957). The authors concluded that differences in outcomes as measured by course exam were not

significantly different for the three methods.

The results of most other studies (e.g., Leton, 1961; Olson, 1959; and Rasmussen, 1956) examining TC vs. SC modes of instruction in terms of learning outcome have consistently supported the conclusion that there does not seem to be a difference in learning outcome as a result of either of these two general instructional approaches.

A plausible explanation for these consistent no-difference results is that each of the studies previously mentioned presumably used the same text for all treatment groups. As McKeachie (1962, 1963) has pointed out, the effect of the text (in itself a treatment common to all groups) may be so powerful that it masks any differential effect due to the experimental treatments. The plausibility of this explanation is further supported by the fact that in the aforementioned studies the final exam, which was used as the criterion measure, was primarily based on the text.

Nachman & Opochinsky (1958) provided a demonstration of this masking effect in a study examining the effect of group size on performance. The traditional criterion measure of final exam was used, as well as pop-quizzes given during the last 10 minutes of particular lectures. The quizzes specifically covered material presented in that and the previous lecture only, thus eliminating any effect due to the text (it is assumed that the quiz material was not taken from the text, but this is not explicitly

stated by the authors). The results of this study indicated no difference due to group size in terms of final exam. However, when the groups were compared on the basis of the pop-quizzes the small group scored significantly higher. The experimenters' conclusion was that differential performance was found on the quizzes but not on the final exam because of the masking effect of the text. Presumably the students had recourse to the text for the final exam but not for the pop-quizzes, and the effect of the text masked the differential effect of group size.

The results of the Nachman & Opochinsky study substantiate McKeachie's inference that perhaps inappropriate criterion measures have been used in past research. One alternative, as was just shown, is to give pop-quizzes. However, the inclusion of this measure introduces another intervening variable, anxiety. The emotional reaction to a pop-quiz is presumably different from that due to an announced exam, where the student is emotionally as well as intellectually prepared. Thus, it may be inappropriate to compare score on a pop-quiz to score on a previously announced exam. An approach that would seem to interfere less with a student's test performance would be to include items from outside sources (i.e., not in the text) in the class sessions. Criterion measures would then consist of items testing information <u>not</u> covered in the text. This would effectively eliminate the possibility of a masking

effect due to the text without the use of pop-quizzes.

The relatively few studies yielding a significant difference between the TC and SC modes in terms of student performance have yielded inconsistent results. Spence (1928) reported a study in which the author described the TC condition as a didactic mode of instruction and the SC condition as one in which relevant cases were discussed. In Spence's words, "For example, when intelligence tests were being studied, intelligence test data for an elementary school class were presented." Discussion and individual reports were also included within this mode. The results of this study indicated that the TC mode was more effective than the SC mode in terms of test scores.

Another study yielding results consistent with those of Spence's (Guetzkow, Kelly & McKeachie, 1954) described the TC mode as did Spence, but the SC mode seems to have been different. The SC mode was described as a discussion mode "attempting to create an atmosphere in which the student feels free to expose his misconceptions, so they can be corrected". Thus these two studies seem to have used entirely different treatments to arrive at the same result. Spence's SC group used a case-centered mode where the stress seems to have been on showing the relevance of the material; Guetzkow's SC mode seems to have fostered a Rogerian-like acceptance of each student whether he responded correctly or not. This type of operating definition of the SC mode has been used in the past by Faw (1949) and has resulted in

results directly contradicting those previously mentioned. Thus an additional problem limiting past research and possibly accounting for the inconsistent findings just mentioned is that ephemeral definitions open to different interpretations have been used to define the SC treatment condition. This limitation has possibly been responsible for the lack of replication of those few studies yielding significant results. One way of dealing with this problem would be to operationally define each treatment (and therefore treatment distinctions). This would remove much of the ambiguity inherent in previous definitions as well as facilitate replication of the treatment conditions.

In summary, the majority of studies investigating TC vs. SC modes of college classroom instruction have reported no difference in performance on exams due to either of these teaching methods, and those few studies that have shown significant differences have been inconsistent in their conclusions. A plausible explanation for the no-difference results is the masking effect of a text common to all treatment groups, and the inconsistent results are seen to be at least in part due to a lack of operationally defined treatment distinctions.

Aptitude X Treatment Interactions

Another source of difficulty with previous studies is that they have typically been analyses of main effects (e.g., method A versus method B) concerned with finding

the "one best method" of college classroom instruction. As a result previous studies have typically used measures of central tendency to evaluate the results. As Davis, Marzocoo, and Denny (1970) have stated: "Studies which employ these measures (of central tendency) to test differences between groups inevitably mask the effects of individual differences. Thus, some students may do better with one mode of instruction while other students do better with a second mode, but the use of the statistical mean would disguise this fact and on the average no differences between groups would be observed." As Davis et al. imply, in an attempt to find the one best method previous studies have ignored individual differences. It may be that the only way to detect treatment differences is to incorporate into the designs and analyses relevant individual difference measures. In other words, "the search for generally superior methods must be supplemented by a search for ways of adapting instruction to the individual." (Cronbach & Snow, 1969).

This research approach reduces to a search for aptitude-treatment interactions (ATI) where an aptitude is defined as any individual difference that is correlated with the outcome of a particular treatment. An ATI exists, in effect, when the regression of outcome under treatment A, upon one or more previously obtained aptitude measure, differs in slope from the regression for the same variables under treatment B.

Although the concept of ATI as an approach to educational research is appealing, there is a lack of ATI studies that have revealed significant interactions. One reason for this is that since the concept of ATI was first introduced (Cronbach, 1957) experimenters have essentially performed main effect studies with some aptitude measures taken as an afterthought. Hence, in most of the early ATI studies the particular individual measures were usually not chosen because they were differentially related to each of the treatments used and were therefore usually unrelated to the outcome of the study. This "first stage" of ATI studies which yielded few significant results has been reviewed by Bracht (1970) and discussed in detail by Salomon (1971). The application of the ATI concept has recently shifted into a second phase where ATIs are being hypothesized on an a priori basis. As a result, there are some recently reported studies that indicate the existence of ATIs in educational situations.

Gaudry & Spielberger (1970) recently examined the interactive effects of manifest anxiety X intelligence X stages of learning in a paired-associate learning task. Results indicated a second-order interaction such that while high anxiety facilitated learning for high IQ Ss and interfered with learning for low IQ Ss early in the learning task, later in learning high anxiety tended to facilitate learning for both high and low IQ Ss equally.

A second study more closely related to an actual

instructional situation was reported by Davis et al. (1970). The experimenters investigated the possibility of ATIs between modes of presenting programmed instruction and individual characteristics. The treatments were essentially different modes of responding, and the learner characteristics used were english, reading, verbal, arithmetic, and composite ability measures. No significant interactions were found but this is not surprising in view of the fact that apparently "available" individual measures were used. The authors provided no a priori theoretical basis for the inclusion of these particular variables, and concluded that "Obviously, the present study may not have measured the "right" individual differences or provided the "right" instruction options to elicit the predicted effects."

In a study examining the possibility of ATIs between two different problem solving strategies and intelligence, Keislar & Stern (1970) found that for high IQ children a complex "hypothesis testing" strategy was optimal where for low IQ children the reverse was true.

Another recently reported study is one most relevant to the present study in that it examined the relationship between "selected characteristics of students (anxiety, compulsivity, exhibitionism, and convergent-minus-divergent thinking style) and the relative degrees of success they have in learning from programmed vs. conventionally structured learning tasks." (Ripple, Millman, & Glock, 1969).

With all cases there were no significant interactions found. Again, the possibility exists that inappropriate individual measures were used. For example, high anxiety was associated with lower criterion scores in all instances. If a particular aptitude is to differentially predict as a function of treatment, the slopes of the regression lines of the outcome regressed on aptitude under each of the treatments must be significantly different from each other. As was previously stated, such was not the case in this study.

A last series of studies to be considered examined the existence of an ATI between student activities during lecture and memory for letter-span (Berliner, 1971). The student activities or treatments were note-taking (NT), paying attention (PA), and responding to test-like events (TLE). A disordinal interaction was revealed in both studies such that while TLE was optimal for those Ss low in memory ability, NT resulted in optimal learning for those high memory Ss. An illustration of one of Berliner's results, representative of all results from both studies, is presented in Fig. 1.

In summary, there is a lack of documented ATIs dealing directly with classroom learning situations. In most cases those few that have yielded significant results were based on strong a priori theoretical grounds rather than simply including in the analyses some "available" individual differences. A strong theoretical basis for the

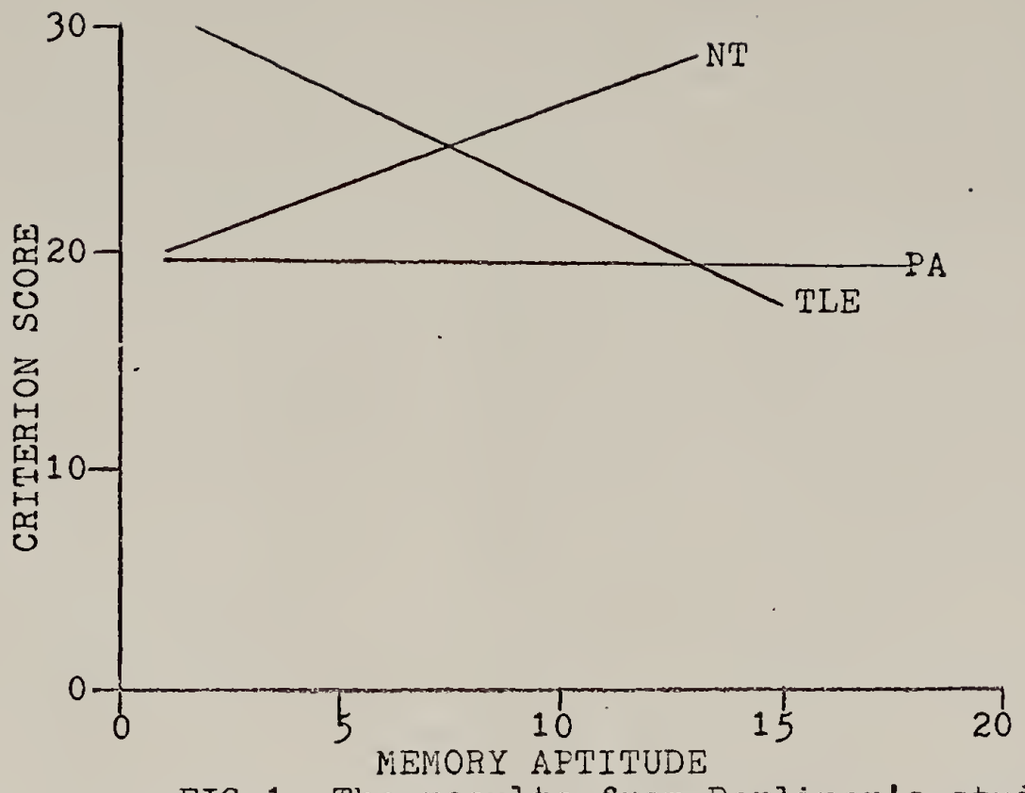


FIG.1. The results from Berliner's study indicating an ATI between letter span and student activity during lecture.

inclusion of particular variables in ATI studies is seen to be an essential component of studies resulting in significant ATIs.

Individual Difference Measures

Since the first appearance of the Taylor Manifest Anxiety Scale (TMAS) (Taylor, 1951) a host of studies have substantiated the general finding that for simple tasks, such as eyelid conditioning, a high score on the TMAS facilitates learning and a low score on the TMAS results in poorer learning (e.g., Spence, 1958; Spence & Beecroft, 1954; and Spence & Farber, 1953). Taylor and Chapman (1955) have extended this finding to simple verbal learning situations, such as paired associate learning. The converse has been found to be the case in more complex learning.

Farber and Spence (1953) examined the relationship between levels of the TMAS and the learning of a ten-choice stylus maze. The results were that low TMAS Ss were superior to high TMAS Ss in terms of number of errors and number of trials to criterion. Williams and Corder (1957) extended this finding to a concept-formation situation, where the concept to be learned was the rule determining the correct light on a display panel of several lights. Results indicated that low TMAS Ss worked more efficiently than high TMAS Ss.

The theoretical explanation for these findings, as presented by Spence (1958), is that the effect of different levels of TMAS is a function of the relative strengths of

the correct and the erroneous response tendencies elicited by the situation. On simple tasks high TMAS facilitates performance because the strength of the correct response tendency far outweighs the strengths of any erroneous response tendencies. In fact, in the case of simple eyelid conditioning there are presumably few if any erroneous tendencies elicited. The opposite is the case in complex tasks, where many competing response tendencies may be elicited by the learning situation. A number of studies incorporating experimental situations that range from paired-associate learning to college achievement have supported this theory (Gaudry & Spielberger, 1970; Spence, Taylor, & Ketchel, 1956; Spielberger, 1962; and Spielberger & Katzenmeyer, 1959). Another facet of Spence's theory that is relevant is his emotional reactivity hypothesis, which states that high TMAS Ss have lower thresholds of emotional responsiveness and therefore react with a stronger emotional response than low TMAS Ss to an equal situation. Spence, Farber, & Taylor (1954) provided support for this hypothesis. In an experiment examining the effects of threat of shock on eyelid conditioning, the experimenters found a shock X TMAS interaction, such that there was a significant difference between the shock and no-shock group only within the high TMAS group. The direction of the difference, consistent with previous findings, was that the shock group (whose anxiety had presumably been increased) performed better than

the no-shock group. Seigman (1956) found that high TMAS Ss had significantly lower scores on the time-limited subtests of the WAIS than on the untimed tests.

The logical implication of Spence's emotional reactivity hypothesis is that a particular classroom situation will elicit an emotional reaction, anxiety, as a function of score on the TMAS. McKeachie (1951) has shown that anxiety, a factor influencing classroom performance, is differentially aroused as a function of the amount of structure the learning situation provides. On the final exam the highly structured "recitation" group scored significantly higher than the less structured "study-tutorial" group. A third "discussion" group scored between the recitation and study-tutorial groups. The results revealed a trend such that mean score on the final exam was a function of classroom structure. McKeachie concludes that a less structured situation is more anxiety provoking because of the ephemeral nature of student requirements. In his words, "The student's anxiety is heightened or reduced by the instructor's teaching behavior. In general, control of anxiety is easiest if the student is in a highly structured situation, where he knows exactly what he must do."

Statement of the Problem

In conjunction with Cronbach's urging experimenters to include individual variables in their investigations, Spence's emotional reactivity hypothesis, and McKeachie's finding

that anxiety is differentially aroused as a function of classroom structure, one hypothesis of this experiment was that college students scoring high on the TMAS would perform better in a teacher-centered mode of classroom instruction than in a student-centered mode of classroom instruction.

Another individual-difference of this study was memory aptitude. In view of Berliner's results (previously mentioned) it seemed plausible that a student-centered mode of instruction, because of its redundant and student-involving nature, would result in superior performance for students scoring low in memory aptitude than would a teacher-centered mode for low memory aptitude students. Therefore, a second hypothesis of this study was that, for those Ss scoring low in memory aptitude, the student-centered mode would be more facilitive than would be a teacher-centered mode.

# METHOD

## Subjects and Setting

The Ss were 68 junior college students enrolled in two separate sections of an introductory psychology course at Greenfield Community College, Greenfield, Massachusetts. Males and females were evenly distributed in both sections, each of which was arbitrarily designated as either the TC group or the SC group. At the outset of the semester there were 35 Ss in the TC group and 33 Ss in the SC group. The Ss were unaware that they were participating in an experiment throughout the entire experimental period. The course was presented as a traditional, semester-long introductory psychology course using a standard text on the subject matter (Kagan & Havemann, 1968). The E was a part-time faculty member at Greenfield Community College and taught both sections involved in this study.

## Individual Difference Measures

The following three individual difference measures were used.

1) Taylor Manifest Anxiety Scale (TMAS): originally constructed by Taylor (1951), this scale, in its present form, consists of 28 items to be answered true or false. The published test-retest reliability, based on scores obtained from 179 students in an introductory psychology course, is .88. This reliability estimate was taken with an intertest interval of four weeks, and rank order remained essentially

the same (Taylor, 1953).

2) Wechsler Adult Intelligence Scale Digit-Span test (DS) (Wechsler, 1955): though this test is intended to be used on an individual basis, the number of Ss used necessitated it being administered to the entire group simultaneously. Because each of the two groups were tested separately, the digit span test was given via a tape recording. This insured both groups receiving the same directions and format for taking the test. The digit spans were presented in ascending order of length from three to nine at a rate of one digit per second.

3) Mental ability test (MAT): this measure consisted of 50 verbal and math items randomly drawn from the Otis-Lennon Mental Ability Test-Advanced level, form J (Otis & Lennon, 1967).

Because intact groups were used, it was not possible to randomly assign Ss to treatments. However, a comparison of the two groups with respect to each of these three variables indicated no significant differences. Additionally, Cochran's test indicated homogeneity of variance for all variables tested. The assumption that the groups were equal was made on the basis of these results, reported in Table 1. The means and standard deviations of these variables for the TC and the SC groups are reported in Table 2.

Design and Procedure

The individual-difference measures were taken during the first two weeks of class for each section. Each of the

Table 1

Summary of main effect analyses of independent variables

| | F test | | | Cochran's test | |
|---|---|---|---|---|---|
| variable | df | F | p | df | C |
| Digit Span | 1,65 | .38 | .55 | 2,35 | .58 (NS) |
| TMAS | 1,65 | .75 | .61 | 2,35 | .64 (NS) |
| MAT | 1,55 | .35 | .56 | 2,30 | .56 (NS) |

Table 2

Means and standard deviations for the three individual-difference measures

| | TC | | | SC | | |
|---|---|---|---|---|---|---|
| variable | N | mean | SD | N | mean | SD |
| Digit Span | 36 | 6.08 | 1.36 | 31 | 6.35 | 1.87 |
| TMAS | 36 | 11.80 | 5.27 | 31 | 10.81 | 3.92 |
| MAT | 31 | 37.77 | 7.21 | 26 | 36.58 | 8.13 |

two sections constituted a treatment group, operationally defined as follows.

| TC | SC |
|---|---|
| 1. Instructor told the Ss on the first day of class that, due to the large number of students in the course and the nature of the subject matter, the course would follow a lecture format. | 1. Instructor told the Ss on the first day of class that, due to the small number of students in the course and the nature of the subject matter, the course would follow a discussion format. |
| 2. Instructor told the Ss on the first day of class that there would be a five-minute question-answer period at the close of each class, during which he (the instructor) would answer any questions pertaining to the material. Following this question-answer period the instructor summarized the material covered in that class period. | 2. Instructor told the Ss on the first day of class that questions were welcome at any time during the class period, addressed to either another S or to the instructor, and that one of them (Ss) would be asked during the period to summarize the material at the end of it. The summary was given from the S's seat and may have been nothing more than a verbal report of S's notes. |
| 3. S-initiated responses were discouraged by the instructor. The instructor provided answers to questions raised during the lecture as quickly as possible, and did not answer S-initiated responses with questions or leading statements. No verbal reinforcement, such as "good", or "that's interesting" was used following S responses. | 3. S-initiated responses were strongly encouraged by the instructor. The instructor paused and obviously reflected upon questions raised during the class period and, as much as was feasible, answered S-initiated responses with questions or leading statements. Verbal reinforcement, such as "good" or "that's interesting" was used following correct S responses. |
| 4. The instructor did not build upon S-initiated responses, as much as he could avoid doing so, even though a response might have been central to the material being covered. | 4. The instructor did build upon S-initiated responses, as much as was possible, whenever the S-initiated response could, in some way, be linked to the material being covered. |

| | |
|---|---|
| 5. The instructor demonstrated to the class certain phenomena that were being covered in the lectures (e.g., the knee-jerk as a demonstration of a physiolcgical reflex). | 5. As much as was possible, the Ss performed certain phenomena that were being covered in the class sessions on each other (e.g., Ss performed the knee-jerk exercise on each other). |

6. In addition to the above stated treatment criteria, the number of S-initiated responses per session was recorded for post-hoc analysis of the intertreatment distinction with regard to frequency of S-initiated responses per session.

The first several weeks of the semester served as a warm-up period in order that the aforementioned treatment distinctions would result in an inter-treatment difference with regard to active S involvement, as measured by the number of S-initiated responses per unit of time, prior to the introduction of criterion-relevant material. As may be seen in Table 3, a difference was observed such that the S-centered group was more active throughout the entire experimental period.

Both groups met for three hours of class per week for the entire semester, though this study terminated with the midterm examination. The TC group met on Monday, Wednesday, and Friday for a 50 minute class from 8:00 A.M. to 8:50 A.M. The SC group met on Tuesday and Thursday for a 75 minute class from 3:30 P.M. to 4:45 P.M. The same criterion measures were used for both groups. The same subject matter was covered in both groups (inasmuch as this was possible), though within the T-centered group "buffer" items, which were not in the criterion measures, were used. The buffer

Table 3

Frequency of S-initiated responses

| Week | Frequency | | Approximate ratio |
|---|---|---|---|
| | T-centered | S-centered | |
| 1 | 8 | 28 | 1·3 |
| 2 | 15 | 55 | 1·4 |
| 3 | 3 | 32 | 1·11 |
| total | 26 | 115 | 1·4 |
| 4 | 8 | 24 | 1·3 |
| 5 | 6 | 16 | 1·3 |
| 6 | 20 | 30 | 1·1.5 |
| total | 34 | 70 | 1·2 |

items served to compensate for the increased time needed in the SC condition to cover the same material as in the TC mode. The experimenter taught both groups and, as much as was possible, varied only the six treatment criteria as previously defined. In order to guide the differential application of these criteria the class sessions were audio-taped. This was done under the guise of providing the teacher with a content guide with which to construct exams, and to provide the teacher with feedback on his teaching ability and performance for self-improvement purposes. The audio tape also provided a reliable recording of the frequency of S-initiated responses. The treatment criteria were maintained for nine weeks following the warm-up period. The subject matter for the first three weeks of the experimental period was "principles of learning". The second three week segment was concerned with "statistics". Relevant criterion measures were taken following each of these three week periods. Content covered during the third segment did not contribute toward criterion measures, but was intended to allow a three-week delay between criterion measures, explained below.

Criterion measures

There were five criterion measures, all based on student achievement. All criterion measures consisted of 4-alternative, multiple-choice questions dealing both directly

and indirectly with the content covered in class only (i.e., not in the text). That is, items requiring integration as well as assimilation of the content covered in class only were included.

Criterion measures 1 and 3 (CT-1, CT-3): items were chosen from a pool of 71 items provided by various publishers of psychology texts and randomly assigned to either CT-1 or CT-3, such that both criterion measures would presumably test the same subject matter. Thus, CT-1 consisted of 36 items, and CT-3 consisted of 35 items, each providing a comprehensive test of the "principles of learning" covered in class.

Criterion measures 2 and 4 (CT-2, CT-4): items were chosen from a pool of 73 items provided by various publishers of statistics and measurement texts, and randomly assigned to either CT-2 or CT-4, such that each criterion measure would presumably test the same subject matter. Thus, CT-2 consisted of 36 items, and CT-4 consisted of 37 items, each providing a comprehensive test of the "statistics" covered in class.

Criterion measure 5 (CT-5): this criterion measure consisted of CT-3 and CT-4 analyzed as one dependent measure.

CT-1 was administered immediately following the first three-week segment of the experimental period. It was announced two days (one class period) in advance and was a power test. The following directions, in addition to accompanying each test booklet, were read aloud to each

class just prior to their exam:

> This examination consists of 36 multiple-choice items, each having 4 possible alternatives. In each case, choose that alternative which best completes the item or answers the question. That is, it may be that on certain questions all 4 possible alternatives are correct, but in varying degrees of specificity. In all instances choose that answer that is, in your opinion, the most specific of the 4 choices. Please indicate your answers by circling the relevant letter. There is no time limit on this exam. Answer each question as there is no penalty for incorrect responses. Turn this page and begin.

CT-2 was administered immediately following the second three-week segment of the experimental period. It was announced two days (one class period) in advance and was a power test. The same directions that were read for CT-1, with the necessary changes (e.g., number of items) both accompanied each S's exam booklet and were read aloud by the E.

CT-3 and CT-4 were administered as a midterm examination immediately following the end of the experimental period. The same instructions used on CT-1 and CT-2 accompanied CT-3 and CT-4. CT-3, presumed to be a randomly parallel form of CT-1, served as a delayed criterion measure of the material covered in the first segment of the experimental period. CT-4, presumed to be a randomly parallel form of CT-2, served as a delayed criterion measure of the material covered in the second segment of the experimental period. Split-half reliability estimates were calculated for each of the five

criterion measures and are reported in Table 4. As is shown, the reliabilities range from .75 to .95.

Plan of Analysis

Since this study was concerned with interactions (as opposed to main effects), an appropriate statistical technique was used to analyze the data. The analysis of variance was not used for the following reasons:

1) Since intact groups were to be used, it was highly unlikely that the number of Ss in each treatment condition would remain equal or proportional throughout the experiment. As Myers (1966) has stated, disproportionality complicates the analysis and requires additional assumptions to be made about the cause of the disproportionality.

2) Also due to the use of intact groups, it was highly unlikely that a sufficient number of Ss with extreme scores on the TMAS or digit span test would be observed.

Regression analysis provided a parsimonious solution for the above problems and was therefore the statistical technique employed, as preliminary analyses indicated linearity and homoscedasticity, two requirements for regression analysis. The hypothetical situation where final grade is regressed on an aptitude variable for each of two teaching methods, labelled A and B, and results in an ATI, is shown in Fig.2. As is illustrated, method B results in final grades superior to those of method A only up to that point at which the lines cross over (labelled $X_o$). From that point on the converse is indicated, that method A is the optimal method. Replication and subsequent substantiation for this particular ATI might generate the following rules to be used in method

Table 4

Reliability estimates of the five criterion tests

| Criterion test | N | r* |
|---|---|---|
| Immediate Learning | 59 | .75 |
| Immediate Statistics | 49 | .95 |
| Delayed Learning | 47 | .78 |
| Delayed Statistics | 47 | .88 |
| Midterm | 47 | .91 |

*split-half reliability coefficients corrected by the Spearman-Brown Prophecy formula (Magnusson, 1966)

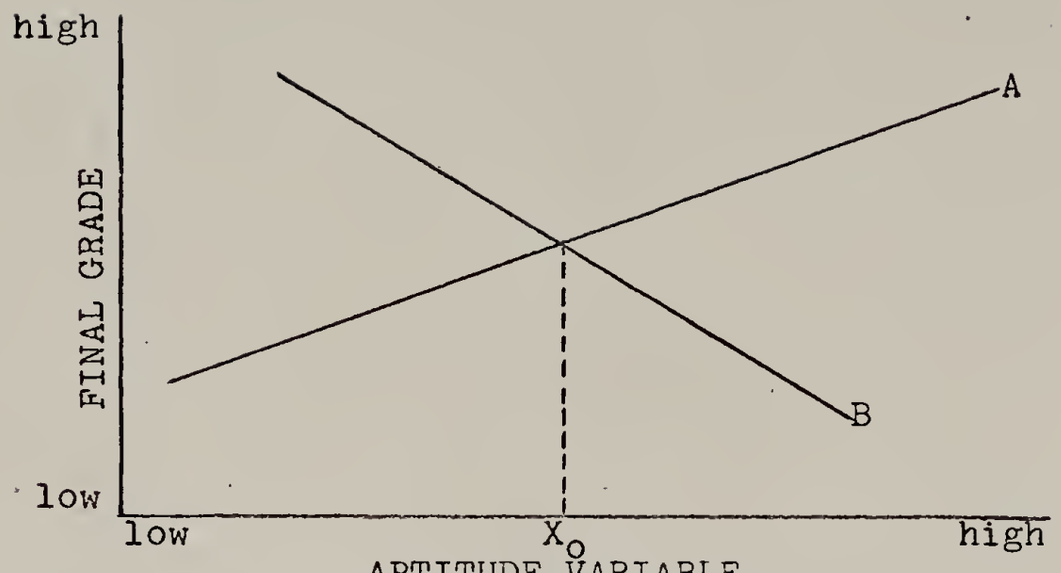


FIG. 2. A hypothetical example of an ATI.

(class) assignment:

1) Students scoring below the point $X_o$ on this aptitude measure should be assigned to class B.

2) Students scoring above the point $X_o$ on this aptitude measure should be assigned to class A.

The problem with using regression analysis (per se), even though a test for parallelism of regression slope may show the slopes to be significantly non-parallel, is that at the point denoted $X_o$ in Fig. 2 both treatments are exactly equal with regard to the predicted final grade. Additionally, homogeneity may be safely inferred to extend for some unknown distance both above and below the point $X_o$. Thus, though parallelism test may indicate whether or not an ATI exists, the limits of the "region of homogeneity" (and hence, regions of heterogeneity) are not indicated with regression analysis alone. This problem has been mentioned by others (e.g., Bracht, 1970; Cronbach & Snow, 1969) and is outlined in detail by Berliner (1971). The Johnson-Neyman technique defines a region of non-significance about the point of homogeneity of two interacting regression slopes, such as $X_o$, within which statistical homogeneity may be said to exist (Johnson & Neyman, 1936). As presented in the original article, the technique was designed to define a region of non-significance based on two independent variables and one dependent variable. A hypothetical example is presented in Fig. 3. As may be seen, the Johnson-Neyman technique

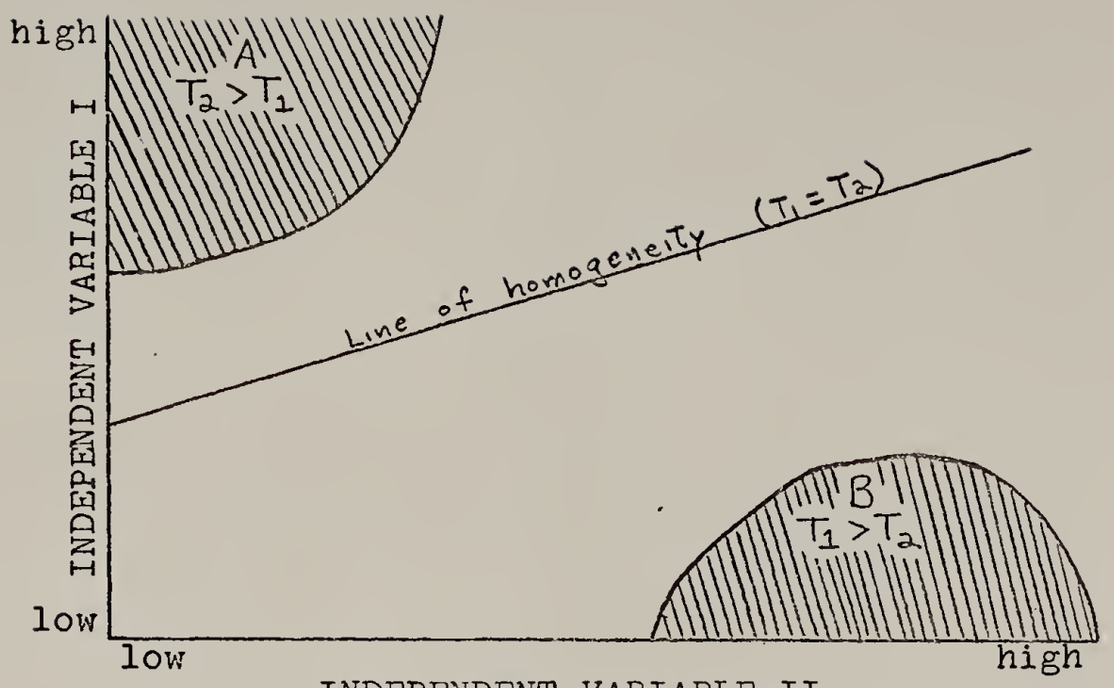


FIG. 3. A hypothetical example illustrating regions of significance computed by the Johnson-Neyman technique.

results in a line of homogeneity, about which a region of statistical homogeneity is computed. The shaded portions represent the two regions of significant heterogeneity, labelled A and B. As treatment 2 ($T_2$) results in criterion scores superior to those obtained by treatment 1 ($T_1$) in region A, $T_2$ can be said to be the optimal treatment for those Ss whose independent variable scores place them there. Conversely, as $T_1$ results in criterion scores superior to those obtained by $T_2$ in region B, $T_1$ can be said to be the optimal treatment for those Ss whose independent variable scores place them there. Differential decisions as a function of independent variable scores are not possible for those Ss whose scores place them within the region of homogeneity. These regions are calculated with a known level of probability and will approach the line of homogeneity with numerically increasing (e.g., from .01 to .05) levels of probability.

A special case of the Johnson-Neyman technique especially useful in education research, the 2-space application where one independent variable and one dependent variable are used, is shown in Fig. 4. As is shown, in the case of a disordinal interaction (i.e., where the regression lines cross over within the range of the data) the region of non-significance extends both above and below the point $X_o$. In predicting the optimum treatment, those Ss scoring below $X_b$ on the independent variable should be assigned to $T_1$, and

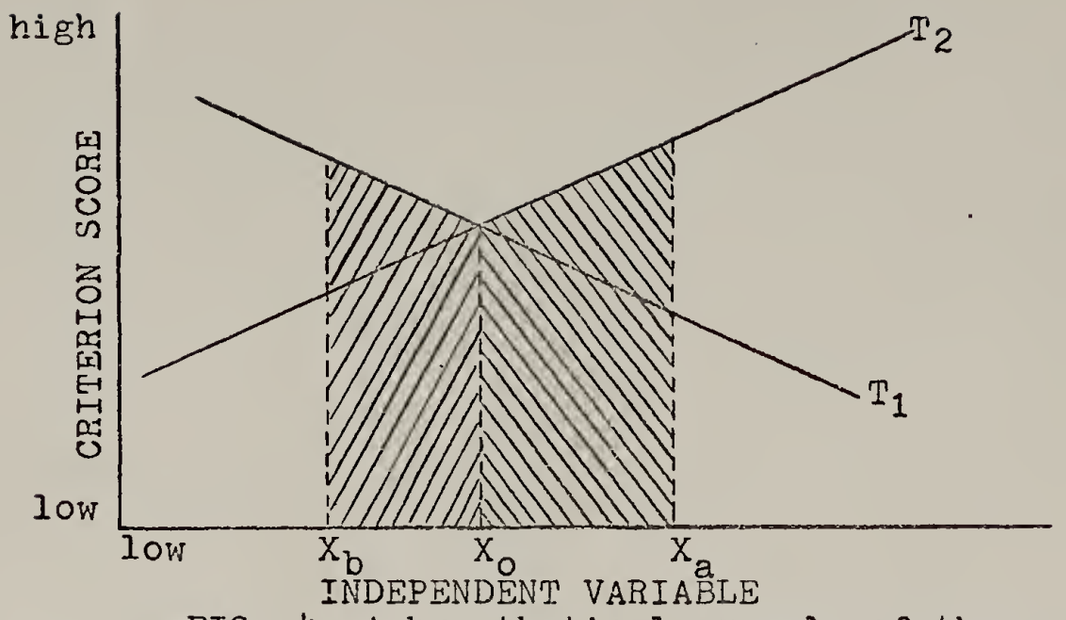


FIG. 4. A hypothetical example of the 2-space version of the Johnson-Neyman technique.

those Ss scoring above $X_a$ on the independent variable should be assigned to $T_2$. For those Ss whose independent measure falls between $X_b$ and $X_a$ both treatments are shown to yield the same results. Therefore, for these Ss, no differential prediction (i.e., as a function of a particular score on the independent variable) of optimal treatment is possible. These predictions are made with known levels of probability and as alpha is numerically increased, both $X_b$ and $X_a$ approach $X_o$.

Because of the aforementioned limitations of the use of parallelism tests alone in analyzing ATI data, the Johnson-Neyman technique as well as parallelism tests was used to analyze the data from this study.

## RESULTS and DISCUSSION

The means and standard deviations for the TC and SC groups on each of the five criterion tests are reported in Table 5. Because of the possibility that the instructor might have inadvertently performed more effectively with one group than with the other, an analysis of main effects and Cochran's test of homogeneity of variance (Winer, 1962) were performed on all criterion measures. The results, illustrated in Table 6, indicate no significant difference and homogeneity of variance for all dependent variables. Thus, when the individual-difference measures are not included in the analysis no difference between the TC and SC modes of instruction are observed. This finding is consistent with the non-significant results of previous studies comparing these and other modes of college classroom instruction, and with Dubin and Taveggia's conclusion that there is no difference between modes of instruction in terms of final exams.

### TMAS-Related Results

A test of parallelism of regression slope (Walker & Lev, 1953) was performed regressing each of the dependent variables on manifest anxiety. The results, reported in Table 7, revealed significantly different slopes when the immediate tests were regressed on manifest anxiety. These results indicate an ATI between manifest anxiety and classroom structure when immediate criterion measures are used as the

Table 5

Means and standard deviations for the five criterion tests

| Criterion test | TC | | | SC | | |
|---|---|---|---|---|---|---|
| | N | mean | SD | N | mean | SD |
| Immediate Learning | 25 | 19.68 | 5.26 | 23 | 18.48 | 5.69 |
| Immediate Statistics | 25 | 19.84 | 5.44 | 22 | 20.77 | 6.08 |
| Delayed Learning | 26 | 18.73 | 5.62 | 24 | 17.58 | 5.19 |
| Delayed Statistics | 26 | 18.69 | 5.36 | 24 | 20.00 | 6.50 |
| Midterm | 26 | 37.42 | 10.03 | 24 | 37.58 | 10.70 |

Table 6

Summary of main effect analyses of dependent variables

| variable | F test | | | Cochran's test | |
|---|---|---|---|---|---|
| | df | F | p | df | C |
| Immediate Learning | 1,46 | .58 | .54 | 2,24 | .54 (NS) |
| Immediate Statistics | 1,45 | .31 | .59 | 2,24 | .55 (NS) |
| Delayed Learning | 1,48 | .56 | .53 | 2,25 | .54 (NS) |
| Delayed Statistics | 1,48 | .61 | .55 | 2,25 | .59 (NS) |
| Midterm | 1,48 | .003 | .95 | 2,25 | .53 (NS) |

Table 7

Test of parallelism of regression slopes with manifest anxiety as the independent variable

| dependent variable | df | F | p |
|---|---|---|---|
| Immediate Learning | 1,44 | 8.572 | .005 |
| Immediate Statistics | 1,42 | 4.752 | .03 |
| Delayed Learning | 1,45 | 1.932 | .17 |
| Delayed Statistics | 1,45 | 2.176 | .14 |
| Midterm | 1,45 | 2.498 | .12 |

dependent variables. The regression slopes for each of these two measures are presented in Fig. 5 and Fig. 6. As is shown, the same disordinal relationship holds for the two different subject matters used. The SC mode resulted in higher scores on the criterion measures than did the TC mode for those Ss sufficiently low in manifest anxiety. Conversely, for those Ss sufficiently high in manifest anxiety, the TC mode resulted in higher scores.

Although the parallelism tests using the delayed measures as dependent variables did not indicate significantly different slopes, trends consistent with the results of the analyses using the immediate measures are noted. These results, shown in Fig. 7, Fig. 8, and Fig. 9, imply the same relationship between manifest anxiety and classroom structure as the results on the immediate criterion measures indicate.

These results support one hypothesis of this experiment, that college students high in manifest anxiety perform better in a teacher-centered mode of classroom instruction than in a student-centered mode. Because of the aforementioned difficulties involved with making predictions based soley upon parallelism test the Johnson-Neyman technique was used to further analyze the immediate criterion test data. Results of the "immediate learning" data revealed a region of homogeneity with a lower limit of 3.65 and an upper limit of 12.75 ($p=.05$). Thus, any predictions of optimal mode of instruction for this subject matter based on manifest

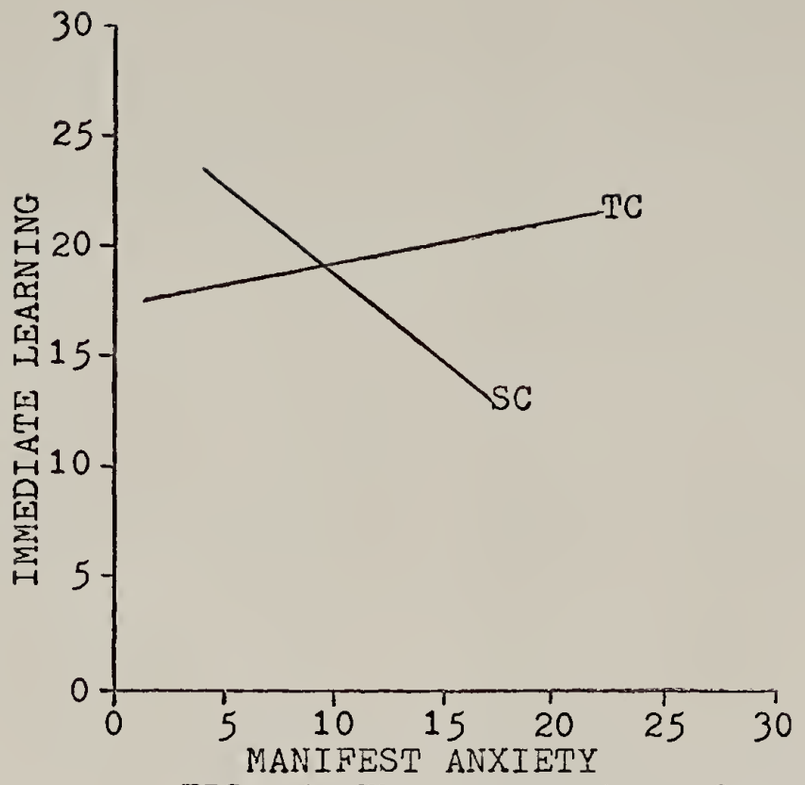


FIG. 5. The regression of score on the immediate learning exam on manifest anxiety.

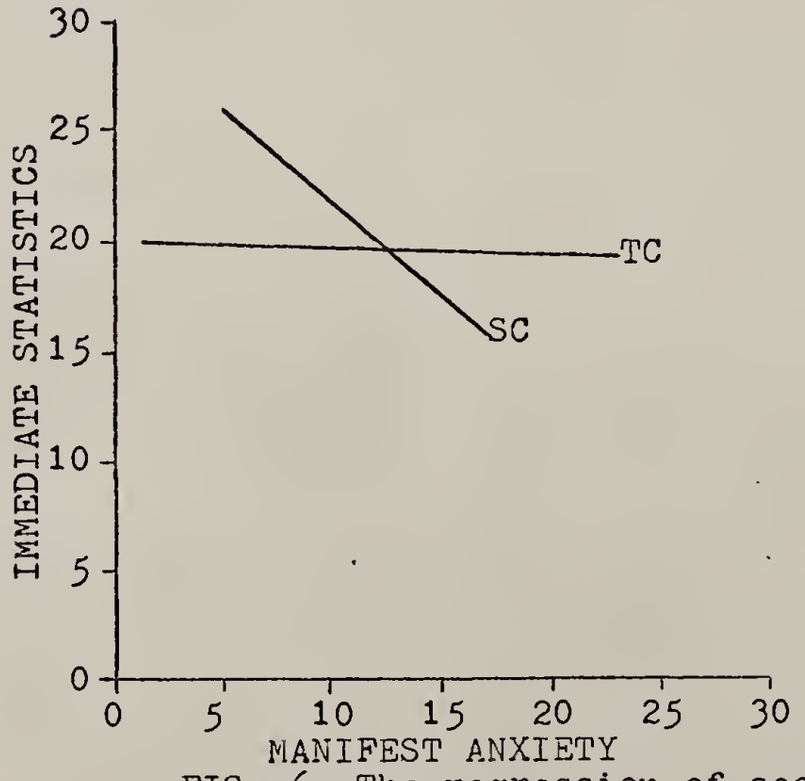


FIG. 6. The regression of score on the immediate statistics exam on manifest anxiety.

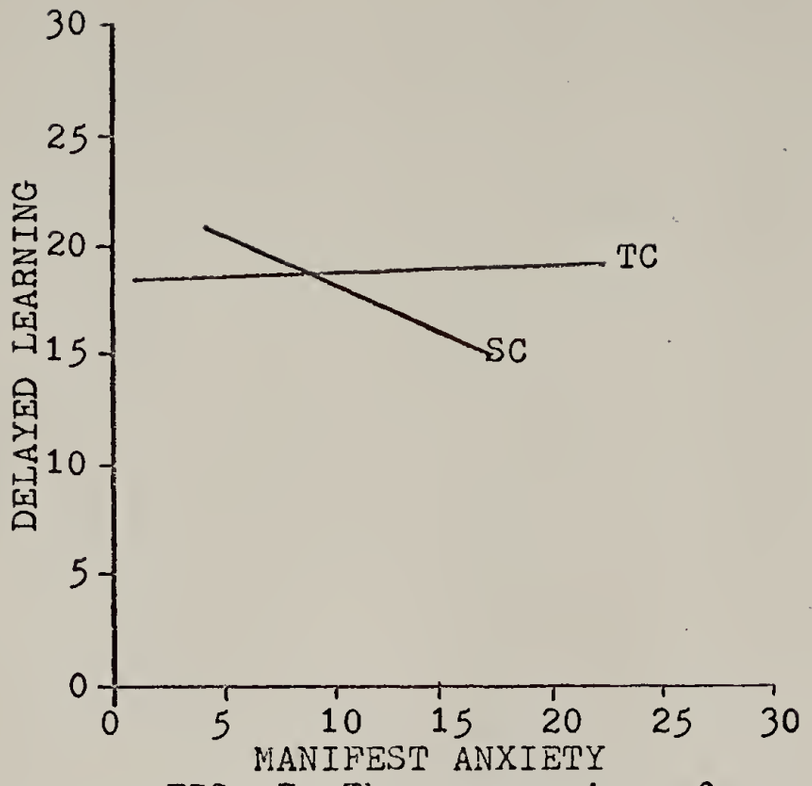


FIG. 7. The regression of score on the delayed learning exam on manifest anxiety.

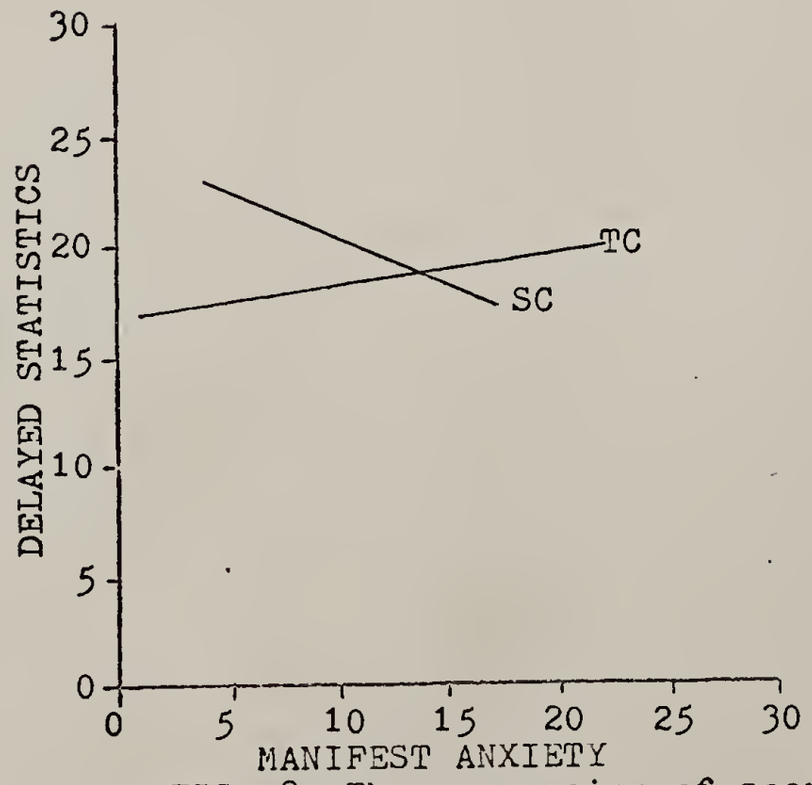


FIG. 8. The regression of score on the delayed statistics exam on manifest anxiety.

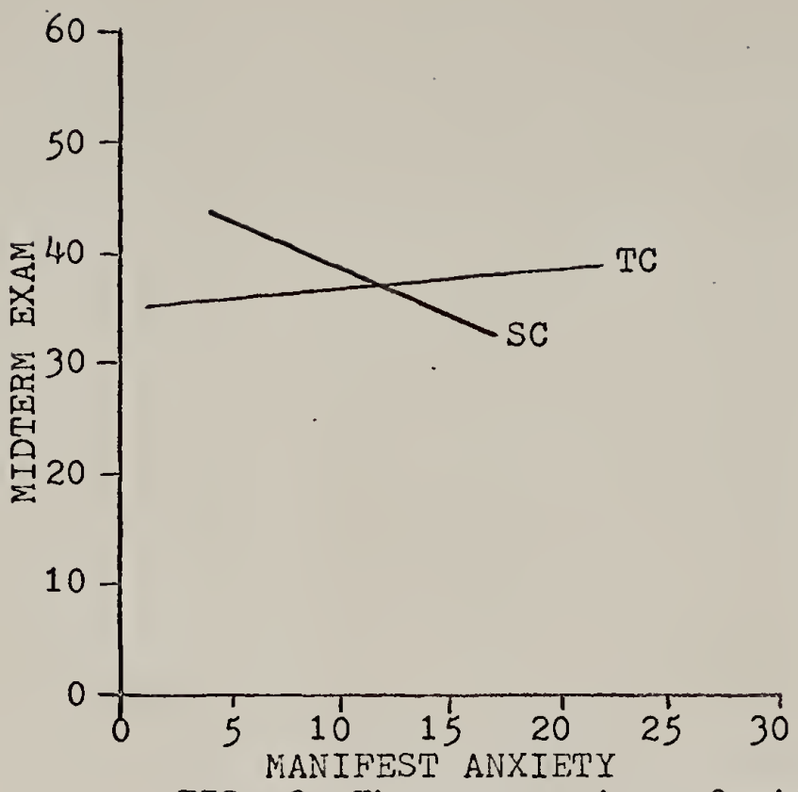


FIG. 9. The regression of midterm score on manifest anxiety.

anxiety can only be made with less than 95% confidence for anyone scoring between 3.65 and 12.75 on the TMAS. However, the optimal mode of classroom instruction for this subject matter can be predicted with 95% confidence for those scoring below 3.65 or above 12.75 on the TMAS. It is noted that these boundries reflect the actual data and no extrapolations have been made. There were 18 Ss (27.5% of the entire sample) observed above the region of non-significance and 2 Ss (4.2% of the entire sample) observed below the region of non-significance. For those scoring below 3.65 on the TMAS the SC mode is predicted as optimal, while for those scoring above 12.75 the TC mode is the predicted optimal mode.

Results of the application of the Johnson-Neyman technique to the "immediate statistics" data revealed a region of non-significance between 4.71 and 28.57 on the TMAS ($p=.05$). While there were two Ss (4.3% of the entire sample) observed below the region of homogeneity, none were observed above. Thus, any prediction of optimal mode for this subject matter based on score on the TMAS can only be made with less than 95% confidence for those Ss scoring above 4.71 on this scale. However, the prediction that the SC mode is optimal for those scoring below 4.71 on the TMAS can be made with 95% confidence.

The correlations (Pearson-Product-Moment) between manifest anxiety and each of the criterion measures for each treatment group, reported in Table 8, further strengthen these results and hence the relevance of Spence's emotional reactivity

Table 8

Correlations between the TMAS and criterion measures for each treatment group

| variable correlated with TMAS | T-centered | N | S-centered | N |
|---|---|---|---|---|
| Immediate Learning | .21 | 25 | -.54* | 23 |
| Immediate Statistics | -.02 | 25 | -.52* | 21 |
| Delayed Learning | .04 | 26 | -.35 | 23 |
| Delayed Statistics | .14 | 26 | -.26 | 23 |
| Midterm | .10 | 26 | -.33 | 23 |

*p< .02

hypothesis to the classroom situation. The significant negative correlations between manifest anxiety and each of the immediate criterion tests implies that Ss high in manifest anxiety have lower thresholds of emotional reactivity, therefore react with a stronger emotional response than do low TMAS Ss to a given classroom situation, and that this emotional reaction can interfere with learning in the classroom. These results are entirely consistent with Spence's emotional reactivity hypothesis and extend those findings to classroom situations.

The non-significant correlations between manifest anxiety and each of the delayed measures is not surprising when the intertest interval between the administration of the TMAS and each of the criterion measures is considered. The interval for the two immediate tests was 4 and 7 weeks, while the delayed measures were taken after an interval of 11 weeks. It is noted, however, that all immediate criterion test correlations with manifest anxiety are significant, and that, while none of the delayed criterion test correlations are significant, they are all negative, consistent with the immediate measures. A possible explanation for the non-significant correlations between the TMAS and delayed criterion measures is that there may have been a confounding effect due to time, studying, or both. That is, it may be that over extended periods of time where Ss may study more than usual for an important exam (i.e., the midterm) any

differential learning due to an ATI is in part compensated for by the intervening study activities. This explanation is consistent with previous results (Nachman & Opochinsky, 1958) whereby only when the investigators controlled for individual study time was the experimental effect observed.

Digit Span Related Results

A test of parallelism of regression slope was performed regressing each of the criterion measures on digit span. The results, shown in Table 9 indicate no difference in slopes and hence that an ATI does not exist between digit span and the modes of instruction used. This conclusion is supported by the essentially zero correlations between digit span and each of the dependent measures, reported in Table 10. Thus, the second hypothesis of this study, that the S-centered mode would facilitate learning for low digit span Ss, is not supported. A possible explanation for these non-significant results is that score on the digit span test may simply be reflecting rehearsal effects rather than actual memory aptitude. The digit span test used here as a measure of memory aptitude employed a rate of one digit per second. Whimbey & Leiblum (1967) have shown that when a rate of less than two digits per second is used, such as was done in this study, rehearsal is possible. Therefore, it is possible that, because of the slow rate of digit presentation used in this study, some of the Ss rehearsed the partial spans in successive interdigit intervals. Furthermore, the possibility exists

Table 9

Test of parallelism of regression slopes with digit span as the independent variable

| dependent variable | df | F | p |
|---|---|---|---|
| Immediate Learning | 1,44 | .03 | .85 |
| Immediate Statistics | 1,42 | .10 | .75 |
| Delayed Learning | 1,45 | .65 | .57 |
| Delayed Statistics | 1,45 | .001 | .98 |
| Midterm | 1,45 | .16 | .69 |

Table 10

Correlations between digit span and criterion measures for each treatment group

| variable correlated with DS | T-centered | N | S-centered | N |
|---|---|---|---|---|
| Immediate Learning | .02 | 25 | -.04 | 23 |
| Immediate Statistics | .14 | 25 | .24 | 21 |
| Delayed Learning | .08 | 26 | -.17 | 23 |
| Delayed Statistics | -.01 | 26 | 0.0 | 23 |
| Midterm | .04 | 26 | -.09 | 23 |

that this propensity to rehearse is not differentially related to the two treatments used. If this is so, then one would expect no difference in slopes when criterion measures relevant to the treatments were regressed on digit span score for each of the two treatment groups.

The same digit span analyses were performed for male and female Ss separately. The results of the parallelism tests, shown in Table 11, revealed significantly different slopes for all of the delayed measures for female Ss only. The correlations between digit span and each of the dependent measures, reported in Table 12, revealed significant negative relationships between digit span and each of the delayed criterion measures within the TC group. Essentially zero correlations are observed within the SC group.

The regression slopes for the female digit span data are shown in Fig. 10, Fig. 11, and Fig. 12. As may be seen, for those female Ss sufficiently low in digit span the TC mode resulted in superior scores. Conversely, for those female Ss sufficiently high in digit span the SC mode was more effective.

The Johnson-Neyman technique was used to define regions of homogeneity for each of the 3 delayed measures. The two analyses incorporating the delayed learning and the delayed statistics measures each with digit span resulted in infinite regions of homogeneity ($p=.05$). However, the analysis incorporating midterm and digit span resulted in a region of homogeneity between 5.0 and 8.9. There were three

Table 11

Tests of parallelism of regression slopes with digit span as the independent variable for males and females

| dependent variable | Males | | | Females | | |
|---|---|---|---|---|---|---|
| | df | F | p | df | F | p |
| Immediate Learning | 1,21 | 2.02 | .17 | 1,19 | 2.42 | .13 |
| Immediate Statistics | 1,19 | .39 | .55 | 1,19 | 1.82 | .19 |
| Delayed Learning | 1,21 | 2.57 | .12 | 1,20 | 4.32 | .05 |
| Delayed Statistics | 1,21 | .15 | .70 | 1,20 | 4.50 | .04 |
| Midterm | 1,21 | 1.01 | .33 | 1,20 | 5,85 | .02 |

Table 12

Correlations between digit span and criterion measures for each treatment group and for males and females

| variable correlated with digit span | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|
| | TC | N | SC | N | TC | N | SC | N |
| Immediate Learning | .19 | 16 | -.41 | 9 | -.55 | 9 | .18 | 14 |
| Immediate Statistics | .30 | 16 | -.05 | 7 | -.28 | 9 | .42 | 14 |
| Delayed Learning | .38 | 16 | -.35 | 9 | -.70* | 10 | -.10 | 14 |
| Delayed Statistics | .24 | 16 | -.01 | 9 | -.62* | 10 | .02 | 14 |
| Midterm | .34 | 16 | -.15 | 9 | -.73* | 10 | -.05 | 14 |

*$p < .05$

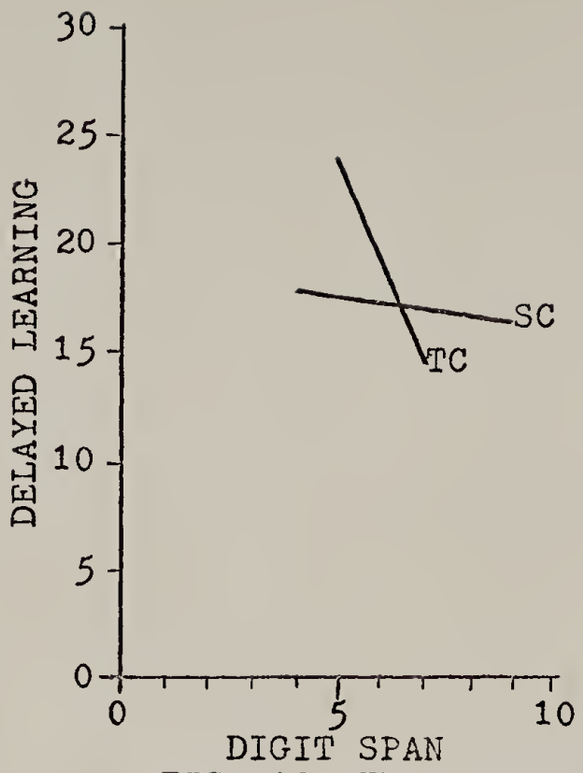


FIG. 10. The regression of score on the delayed learning exam on digit span for female Ss only.

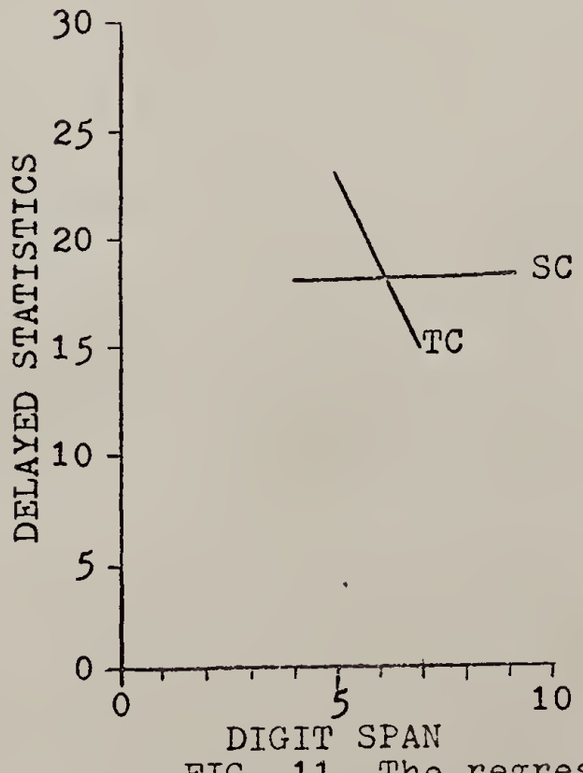


FIG. 11. The regression of score on the delayed statistics exam on digit span for female Ss only.

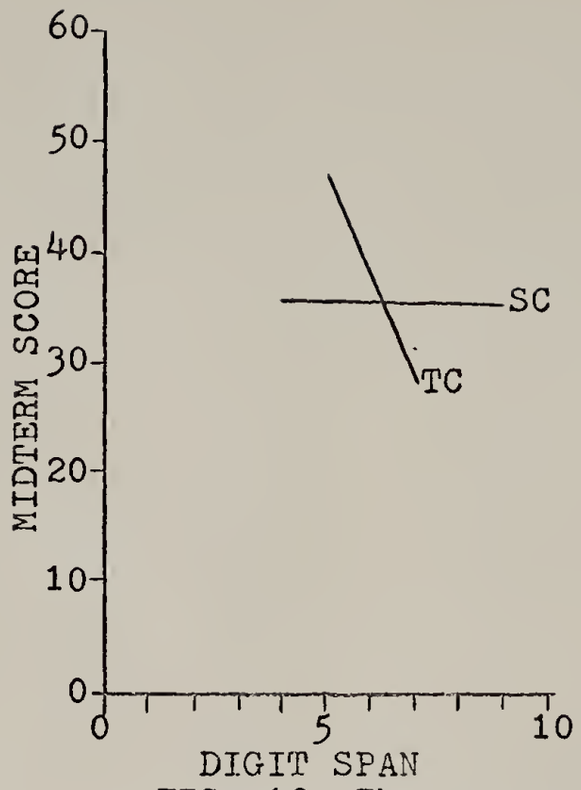


FIG. 12. The regression of total midterm score on digit span for female Ss only.

cases (12.5 % of the entire sample) above and one case (4.2 %) observed below the region of homogeneity. While these results would allow for predictions of optimal mode had they been planned, the post-hoc nature of these analyses prevents any predictions as a result of them. However, a trend implying the possibility of a second-order ATI between sex, digit span, and the modes of instruction used is noted. Whether the crucial aspect of the sex dimension in this possible ATI is female or opposite sex (i.e., the E was male) is not known.

# SUMMARY and CONCLUSIONS

The purpose of this study was to investigate the possibility of differential relationships between each of two individual difference variables and two different modes of college classroom instruction. This form of inquiry is generally termed a search for aptitude-treatment interactions (ATI) where "aptitude" is defined as any individual difference that is correlated with the outcome of a given treatment. The two modes of instruction examined were a teacher-centered approach and a student-centered approach. Results of previous studies comparing these two methods are inconsistent. Possible shortcomings in previous experiments include the finding that previous studies have been "main-effect" studies. Those few ATI studies that have been performed have resulted in no differences, possibly because of the inclusion of inappropriate aptitude measures that were either unrelated to the treatments or that were not differentially related to the different treatments used. The Taylor Manifest Anxiety Scale (TMAS) was chosen as one of the aptitudes because of Spence's emotional reactivity hypothesis, which states that high TMAS Ss react with more emotion (anxiety) than do low TMAS Ss to an equal situation. An additional reason for choosing this aptitude measure was that McKeachie has shown that anxiety is differentially aroused as a function of the amount of structure the learning situation provides. One hypothesis of this experiment was, therefore, that high TMAS Ss would perform better in a

teacher-centered (i.e., structured) mode of college classroom instruction than would their high TMAS counterparts in a student-centered (i.e., unstructured) mode.

A second aptitude measure used in this study was a digit span test. In a recent ATI study Berliner (1971) found that memory for letter span interacted with student activity during a lecture. Berliner's results indicated that, for Ss low in memory aptitude, answering a relevant question every 2.5 minutes resulted in performance superior to that of their low memory counterparts who took notes. In view of these results it seemed plausible that a student-centered mode of instruction, because of its redundant and student-involving nature, would result in superior performance for Ss scoring low in memory aptitude than would a teacher-centered mode for low memory aptitude Ss . Therefore, a second hypothesis of this experiment was that, for those Ss scoring low in memory aptitude, the student-centered mode would be more facilitative than would be a teacher-centered mode.

Students in two intact sections of an introductory psychology class, who were unaware of the experiment, were used as Ss. The aptitude measures were taken at the outset of the semester, before any differential treatment was applied. The criteria for each treatment was operationally defined in terms of activities conducive to either increasing or decreasing S-initiated responses. Additionally, the

number of S-initiated responses was recorded for each group and it was shown that the groups were different with regard to the number of S-initiated responses in the predicted direction.

Because intact groups were used and because this study was concerned with interactions the Johnson-Neyman technique, in conjunction with regression analysis, was used to analyze the data. Results indicated significant disordinal interactions between manifest anxiety and the modes of instruction used with the immediate criterion measures, and trends in the same direction were noted with the delayed criterion measures. The disordinal ATIs and trends indicated that, in support of the first hypothesis, high anxious Ss perform better in a teacher-centered classroom situation than in a student-centered classroom situation. Additionally, it was shown that the optimal mode of instruction for low anxious Ss was a student-centered mode.

The results did not support the second hypothesis of this study, that Ss low in digit span would perform better in a student-centered mode of classroom instruction than in a teacher-centered mode. Post-hoc analyses were performed on the same data for each sex separately. Results inplied an ATI such that female Ss low in digit span performed better on all delayed criterion tests as a result of the teacher-centered mode of instruction. For those female Ss high in digit span the student-centered mode resulted in superior

scores on all delayed criterion measures.

Implications

One implication of the present study is that research efforts examining the relative merits of different instructional techniques would be enhanced by the inclusion of relevant individual difference variables in their designs. Perhaps there are no "main effects" to be found and significant treatment effects will be evidenced only when successful attempts are made to account for individual differences rather than simply dismissing them as error. Additionally, perhaps many of the main effect studies that have resulted in no significant differences in the past should be replicated as interaction studies accounting for appropriate individual differences.

A second implication is that, with further substantiation of particular ATIs in the classroom situation, it may be possible to boost grades by assigning students to a particular section of a course as a function of some aptitude measure. For example, if the manifest anxiety by instructional mode interaction evidenced in this study is replicated a sufficient number of times, students' grades could be substantially boosted simply by assigning them to (or counseling them into) their optimal sections. Students sufficiently low in manifest anxiety would be assigned to or advised to join a student-centered section. Those sufficiently high in manifest anxiety would be assigned to or urged to join a teacher-centered section. Since many

introductory college courses use multiple sections already, and since instructors usually tend towards either of the two modes incorporated, the problem would essentially be one of logistics, i.e., getting certain students into certain sections.

Although one conclusion of this study is that a manifest anxiety X teacher or student-centeredness disordinal interaction exists, there were several shortcomings to the study. Even though main-effect analyses of all individual-difference variables indicated homogeneity between the treatment groups at the outset of the experiment, it can only be stated that the groups were equal with regard to these few measured traits. Another serious limitation related to the use of intact groups is that not enough extreme scorers on the aptitude measures were observed. A third limitation to this study is that of a possible anxiety X mortality interaction. That is, only Ss present on the day of the exam were included in each of the analyses, and if low anxious Ss tend to take make-up exams then they would have been excluded from the analyses.

# References

Abelson, R.P. A note on the Neyman-Johnson technique. Psychometrika, 1953, 18, 213-218.

Berliner, D.C. Aptitude-treatment interactions in two studies of learning from lecture instruction. Paper presented at the meeting of the American Educational Research Association, New York City, February, 1971.

Bracht, G.H. Experimental factors related to aptitude-treatment interactions. American Educational Research Journal, 1970, 40, 627-645.

Cronbach, L.J. The two disciplines of scientific psychology. The American Psychologist, 1957, 12, 671-684.

Cronbach, L.J., and Snow, R.E. Individual differences in learning ability as a function of instructional variables. Final Report, March, 1969, School of Education, Stanford University, Contract No. OEC4-6-061269-1217, U.S. Office of Education.

Davis, R.H., Marzocco, F.N., and Denny, M.R. Interaction of individual differences with modes of presenting programmed instruction. Journal of Educational Psychology, 1970, 61, 198-204.

Dubin, R., and Taveggia, T.C. The teaching-learning paradox. Eugene, Oregon: University of Oregon Press, 1968.

Eglash, A. A group discussion method of teaching psychology. Journal of Educational Psychology, 1954, 45, 257-267.

Farber, I.E., & Spence, K.W. Complex learning and conditioning as a function of anxiety. Journal of Experimental Psychology, 1953, 45, 120-125.

Faw, V. A psychotherapeutic method of teaching psychology. American Psychologist, 1949, 4, 104-109.

Gaudry, E., & Spielberger, C.D. Anxiety and intelligence in paired-associate learning. Journal of Educational Psychology, 1970, 61, 386-391.

Gerberich, J.R., and Warner, K.O. Relative instructional efficiencies of the lecture and discussion methods in a university course in american national government. Journal of Educational Research, 1936, 29, 574-579.

Guetkow, H., Kelly, E., and McKeachie, W.J. An experimental comparison of recitation, discussion, and tutorial methods in college teaching. Journal of Educational Psychology, 1954, 45, 193-207.

Haigh, G.V., and Schmidt, W. The learning of subject matter in teacher-centered and group-centered classes. Journal of Educational Psychology, 1956, 47, 295-301.

Johnson, P.O., and Neyman, J. Tests of certain linear hypothesis and their application to some educational problems. In J. Neyman and E.S. Pearson (Eds.), Statistical Research Memoirs, 1936, 1, 57-93.

Kagan, J., and Haveman, E. Psychology: An Introduction. New York: Harcourt, Brace & World, Inc., 1968.

Keislar, E.R., and Stern, C. Differentiated instruction in problem solving for children of different mental ability levels. Journal of Educational Psychology, 1970, 61, 445-450.

Krumboltz, J.D., and Farquhar, W.W. The effect of three teaching methods on achievement and motivational outcomes in a how-to-study course. Psychological Monographs: General and Applied, 1957, 71, 1-26.

Leton, D.A. An evaluation of course methods in teaching child development. Journal of Educational Research, 1961, 55, 118-122.

Magnusson, D. Test Theory. Reading, Massachusetts: Addison-Wesley Publishing Company, 1966.

McKeachie, W.J. Anxiety in the college classroom. Journal of Educational Research, 1951, 45, 153-160.

McKeachie, W.J. Procedures and techniques of teaching: a survey of experimental studies. In N. Sanford (Ed.), The American College. New York: John Wiley & Sons, Inc., 1962. Pp. 312-364.

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

Myers, J.L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, Inc., 1966.

Nachman, M., and Opochinsky, S. The effects of different teaching methods: a methodological study. Journal of Educational Psychology, 1958, 49, 245-249.

Olson, K.V. An experimental evaluation of a student-centered method and a teacher-centered method of biological science instruction for general education of college students. Journal of Experimental Education, 1959, 27, 225-230.

Otis, A.S., and Lennon, R.T. Otis-Lennon Mental Ability Test-Advanced level, form J. New York: Harcourt, Brace & World, Inc., 1967.

Rasmussen, G.R. An evaluation of a student-centered and instructor-centered method of conducting a graduate course in education. Journal of Educational Psychology, 1956, 47, 449-461.

Ripple, R.E., Millman, J., and Glock, M. Learner characteristics and instructional mode: a search for disordinal interactions. Journal of Educational Psychology, 1969, 60, 113-120.

Salomon, G. Heuristic models for the generation of aptitude-treatment interaction hypothesis. Paper presented at the meeting of the American Educational Research Association, New York City, February, 1971.

Siegman, A.W. Cognitive affective, and psychopathological correlation of the Taylor MAS. Journal of Consulting Psychology, 1956, 20, 137-141.

Spence, R.B. Lecture and class discussion in teaching educational psychology. Journal of Educational Psychology, 1928, 19, 454-463.

Spence, K.W. A theory of emotionally-based drive (D) and its relation to performance in simple learning situations. American Psychologist, 1958, 13, 131-141.

Spence, K.W., & Beecroft, R.S. Differential conditioning and level of anxiety. Journal of Experimental Psychology, 1954, 48, 399-403.

Spence, K.W., & Farber, I.E. Conditioning and extinction as a function of anxiety. Journal of Experimental Psychology, 1953, 45, 116-119.

Spence, K.W., Farber, I.E., & Taylor, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. Journal of Experimental Psychology, 1954, 48, 404-408.

Spence, K.W., Taylor, J., & Ketchel, R. Anxiety (drive) level and degree of competition in paired-associates learning. Journal of Experimental Psychology, 1956, 52, 306-310.

Spielberger, C.D. The effects of manifest anxiety on the academic achievement of college students. Mental Hygiene, N.Y., 1962, 46, 420-426.

Spielberger, C.D., & Katzenmeyer, W.G. Manifest anxiety, intelligence, and college grades. Journal of Consulting Psychology, 1959, 23, 278.

Taylor, J.A. The relationship of anxiety to the conditioned eyelid response. Journal of Experimental Psychology, 1951, 41, 81-92.

Taylor, J.A. A personality scale of manifest anxiety. Journal of Abnormal Social Psychology, 1953, 48, 285-290.

Taylor, J.A., & Chapman, J.P. Paired-associate learning as related to anxiety. American Journal of Psychology, 1955, 68, 671.

Walker, H.M., and Lev, J. Statistical Inference. New York: Holt, Rinehart and Winston, 1953.

Wechsler, D. Wechsler Adult Intelligence Scale, Manual. New York: Psychological Corp., 1955.

Whimbey, A.E., & Leiblum, S.L. Individual differences in memory span with and without activity intervening between presentation and recall. Journal of Educational Psychology, 1967, 58, 311-314.

Williams, J.E., & Corder, R.F. The relationship of manifest anxiety to efficiency in problem solving. American Psychologist, 1957, 12, 420. (Abstract)

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962.