SOME COMPARISONS BETWEEN CONVENTIONAL COLLEGE TEACHING
METHODS AND A COMPOSITE OF PROCEDURES INVOLVING
LARGE LECTURE GROUPS, SEMINARS, AND REDUCED
CLASS TIME

APPROVED:

Graduate Committee:

Watt L. Black
Committee Chairman

John Curry
Committee Member

Ben Coody
Committee Member

Susan Kingery
Dean of the School of Education

Robert B. Toulmin
Dean of the Graduate School
SOME COMPARISONS BETWEEN CONVENTIONAL COLLEGE TEACHING
METHODS AND A COMPOSITE OF PROCEDURES INVOLVING
LARGE LECTURE GROUPS, SEMINARS, AND REDUCED
CLASS TIME

DISSERTATION

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF EDUCATION

By

Alvin T. Bean, B.A., M.Ed.
Denton, Texas
May, 1969
# TABLE OF CONTENTS

**LIST OF TABLES** .............................................. v

**Chapter**

1. **STATEMENT OF THE PROBLEM** .......................... 1
   - Significance of the Problem
   - The Setting of the Problem
   - Common Course Procedures of the
     Experimental and Control Groups
   - The Composite of Procedures for the
     Experimental Group
   - Hypotheses
   - Definition of Terms
   - Limitations of the Study
   - Basic Assumption

2. **SURVEY OF RELATED LITERATURE** ..................... 22
   - Emphases on Junior Colleges
   - Methods of Instruction in College
     Teaching
   - Summary

3. **COLLECTION AND TREATMENT OF DATA** ............. 79
   - Subjects
   - Instruments
   - Treatment of Data

4. **PRESENTATION AND ANALYSIS OF DATA** .............. 92
   - The First Hypothesis
   - The Second Hypothesis
   - The Third Hypothesis
   - The Fourth Hypothesis
   - The Fifth Hypothesis
   - The Sixth Hypothesis
   - The Seventh Hypothesis
   - The Eighth Hypothesis
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS</td>
<td>126</td>
</tr>
<tr>
<td>Summary of Findings</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td></td>
</tr>
<tr>
<td>Implications</td>
<td></td>
</tr>
<tr>
<td>Recommendations</td>
<td></td>
</tr>
<tr>
<td>APPENDIX</td>
<td>132</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>134</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Total and Categorical Sample Sizes for Criteria Instruments</td>
<td>92</td>
</tr>
<tr>
<td>II. Degrees of Freedom for F-Ratio for Criteria Instruments and Hypotheses</td>
<td>93</td>
</tr>
<tr>
<td>III. Number Designation of Criteria Instruments</td>
<td>94</td>
</tr>
<tr>
<td>IV. Mean Gains and Standard Deviations for All Subjects on Criteria Measures</td>
<td>95</td>
</tr>
<tr>
<td>V. Mean Gains and F-Values for Treatment Groups on Criteria Measures</td>
<td>97</td>
</tr>
<tr>
<td>VI. Mean Gains and F-Values for Categorical Treatment Groups</td>
<td>100</td>
</tr>
<tr>
<td>VII. Categorical Vector Coefficients, Categorical Regression Slopes of ACT English Test and F-Value</td>
<td>103</td>
</tr>
<tr>
<td>VIII. Regression Slope of ACT English Test, Categorical Vector Coefficients, and F-Value</td>
<td>104</td>
</tr>
<tr>
<td>IX. Categorical Vector Coefficients, Categorical Regression Slopes of ACT Mathematics and F-Value</td>
<td>108</td>
</tr>
<tr>
<td>X. Regression Slope of ACT Mathematics, Categorical Vector Coefficients, and F-Value</td>
<td>109</td>
</tr>
<tr>
<td>XI. Categorical Vector Coefficients, Categorical Regression Slopes of ACT Social Studies and F-Value</td>
<td>111</td>
</tr>
<tr>
<td>XII. Regression Slope of ACT Social Studies, Categorical Vector Coefficients, and F-Value</td>
<td>112</td>
</tr>
</tbody>
</table>
Table | Page
---|---
XIII. Categorical Vector Coefficients, Categorical Regression Slopes of ACT Natural Science and F-Value | 115
XIV. Regression Slope of ACT Natural Science, Categorical Vector Coefficients, and F-Value | 116
XV. Categorical Vector Coefficients, Categorical Regression Slopes of ACT Composite, and F-Value | 119
XVI. Regression Slope of ACT Composite, Categorical Vector Coefficients, and F-Value | 120
XVII. Categorical Vector Coefficients, Categorical Regression Slopes of Criteria Pretests and F-Value | 122
XVIII. Regression Slope of Criteria Pretests, Categorical Vector Coefficients, and F-Value | 123
CHAPTER I

STATEMENT OF THE PROBLEM

Diverse instructional plans for teaching have been explored for over a century in the United States. Shane (27) has described thirty-six such plans and has suggested that his list, while by no means comprehensive, serves by its impressive length to emphasize the many ideas that have been voiced for purposes of personalizing teaching and recognizing individual differences. These concerns are being reflected by reports of research on teaching at the college and university levels (3, 7, 15).

The problem of this study was to determine the differences in achievement, critical thinking, and attitude toward subjects of junior college freshmen which could be attributed to two approaches to the teaching of English composition and American history.

The purpose of the study was to yield information for use as the basis for administrative and instructional judgments concerning pupil deployment, and plant and staff utilization.

Significance of the Problem

A conventional college instructional plan at many institutions utilizes units of twenty-five to forty students
meeting as a body three hours a week. An experimental plan, referred to in this study as the Special Program in Curricular Experimentation, utilizes large-group instruction coordinated with small seminar classes and guided independent study.

Much of the theoretical background for this latter approach has originated on the pre-college level. The best known and most comprehensively investigated plan was described by J. Lloyd Trump in 1961 (30). Essentially, this study tested the Trump Proposal with a junior college freshman population.

Prior to 1959, a compendium of reports on thirty-two studies relevant to the Trump Proposal was published and contained the following conclusions:

Size of class in itself has little relationship to achievement; students can learn materials as well in large groups of 70, 100, or even 1,100, as in traditional groups of 25 to 30 (21, p. 285).

In a report of further studies in 1960, a description of the Newton Plan was included (22). This plan was also described in detail during the same year by Bissex (4). The Newton Plan uses some class groups of from 50 to 200 students and is somewhat similar to the Trump Proposal. The Rutgers Plan was another similar proposal. Formulated by teachers of English meeting on the Rutgers campus, it was designed to insure that no English teacher need ever meet more than twenty-five students at a time except by choice (9).
In 1961, the National Association of Secondary-School Principals reported the conclusion of twenty-three of the studies related to the Trump Proposal and suggested that, although results were not definitive, the studies had resulted in much discussion, experimentation, and careful re-examination of secondary education (23, p. 10).

Reports of research on the college and university levels tend to isolate components of the Trump Proposal rather than to examine the pattern of the plan in a matrix. In three major reviews (15, pp. 1118-1172; 18, pp. 211-239; 26, pp. 312-364), the emphases are on lecture versus discussion, small versus large classes, and independent study, each treated separately.

The history of research on these topics began in the 1920's with the Edmondson and Mulder study of class size (10) and the Bane report on lecture versus discussion (2). McKeachie has reviewed subsequent relevant research through 1967 (18, pp. 211-239). Concerning the lecture-versus-discussion component, he concluded that a good deal of evidence suggests that discussion is more effective than lecturing in achieving the more complex cognitive and attitudinal objectives (18, p. 217).

Another component of the Trump Proposal, a limited amount of independent study, has been investigated at the college level with the resulting conclusion that independent study can effectively achieve ends other than knowledge of specific facts (18, pp. 223-225).
After reviewing studies of methods of teaching in colleges and universities, McKeachie offered this summary:

In most courses there are several levels of goals—in knowledge, critical thinking, attitudes toward learning and so on. The teacher's task is to find a combination of methods that will achieve an optimal balance of all these. Unfortunately, most teaching research has studied the effect of one method versus another painfully repeated day after day for a semester; thus little evidence is available on the relative effectiveness of differing combinations or degrees of flexibility in teaching methods (18, pp. 214, 215).

The continuing interest in the elements of this study is illustrated by the contemporaneity of a report of the National Education Association (24) suggesting that research on class size has not been comprehensive or multivariate. A further element of significance of the study is related to a reported paucity of studies on the junior college level (5).

The Setting of the Problem

The subjects were college students enrolled in first semester English composition and American history courses in El Centro Junior College of the Dallas County Junior College District during the fall semester of the 1968-69 school year. The institution is a public open-door junior college committed to assignment of students to sections in the order of receipt of requests. This was the population base of the control group. The population of the experimental group was also assembled through comparable enrollment procedures of the institution. The availability of the experimental schedule
was made known, and students requesting sections in it were assigned in order of request. From the point of view of students, the experimental program was a scheduling option available to them rather than an experimental plan.

The control classes met in the enrollment pattern of twenty-five to forty students three hours per week. Students in the experimental group were enrolled in both English composition and American history in the Special Program in Curricular Experimentation. Their schedules were completed with psychology, business or mathematics offered in the Special Program plan or with these or other subjects offered in conventional classes.

The formal schedule of the experimental group had two parts. One was the general assembly. The English general assembly met on Monday at 10:00 A.M. The history general assembly met on Wednesday at the same hour. Each general assembly was one hour in length, and all students enrolled in the Special Program sections were required to attend. This time was used for lectures and giving major examinations.

The other part of the schedule was a one-hour seminar each week in English and another in history. These were limited to fifteen students and were scheduled by computer for each student.

Replacing the traditional third hour was a planned tutorial program. The teacher assigned instructional meetings or activities as considered necessary and helpful for
each student's progress. Often, the student studied independently and at his own rate.

The English control classes and the history control classes were scheduled for three hours of instruction each week using the conventional pattern of scheduling.

Six teachers were involved; all had similar professional credentials. There were two teachers of the experimental group—one English and one history. Both were females with Master of Arts degrees in their teaching fields. The mean number of years of teaching experience was thirteen.

The two control group history teachers, one male and one female, had Master of Arts degrees in their teaching fields with a mean of ten years of teaching experience. There were also two control group English teachers, both female, with Master of Arts degrees in their teaching fields. The mean number of years of teaching experience was eleven. Schedule limitations prevented experimental group teachers from teaching control sections. Five classes served as the English control group, and five classes served as the history control group.

During the spring semester of 1968, the division chairman of the college had been informed of the decision to organize the Special Program in Curricular Experimentation. Teachers had opportunity to express interest in the program either to division chairmen or the Dean of Instruction. Interested teachers were designated as the experimental group
teachers by the Dean of Instruction in the order that their expression of interest became known. Assignment of the control group teachers was not within the administrative jurisdiction of the investigator.

Common Course Procedures of the Experimental and Control Groups

Both experimental and control classes followed the college's course of study for English 101 and History 101. The same departmental textbooks were used by all sections involved (1, 12, 19, 20). The importance of their roles in the experiment was discussed with all experimental and control classes during the testing periods and visits of the investigator.

The investigator worked with all experimental and control group teachers to more clearly identify the expected outcomes of the courses. Teachers of both groups also worked with the investigator in pre-semester meetings and periodically during the experimental period in order to coordinate comparable methodological approaches and procedures.

Content and Objectives

The English course began with a short unit on the use of the dictionary. This was followed by a unit on planning a composition. Topics included pre-writing, sources of materials, patterns of organization, and the outline. The next unit was related to the expression of ideas. Topics included paragraphs, sentences, and choosing the right words.
Correlated within each unit were exemplary essays on language and style, sources of ideas, and patterns of development. A handbook of grammar and usage was available to all students.

In a series of pre-semester conferences, the following statement of the purposes of the course was developed:

1. The student will write acceptable expository and persuasive papers.

2. The student will demonstrate his knowledge of selected principles of linguistics and semantics and their applications to the American English language today and his ability to recognize and apply basic principles of logic.

3. The student will demonstrate his thinking processes and his knowledge of linguistics and semantics by writing short essay answers and by discussing orally these topics.

The History 101 course included units about the following areas: the beginnings of American history; discord; revolution and independence; the birth of a nation; the roles of Jefferson, Madison, and Monroe; the growth of democracy; the Westward Movement; developments of 1790-1860; slavery; the Civil War and Reconstruction.

In a series of pre-semester conferences, the following statement of the purposes of the course was developed: to afford the opportunity for development of critical thinking skills through class exercises, open-ended historical questions, library research, and analysis of primary and secondary sources; to expose the student to the skills and
techniques of the historian with a view toward making him a more reflective thinker and more responsible citizen; to examine in depth certain selective historical problems in order to note their complexity and better understand events and situations on the current scene.

Audio-Visual Aids

Utilization was made of the departmental core of films in all control and experimental classes (6). In history, this included The Seeds of Revolution, The Revolution, The Era of the Common Man, The Real West, and The True Story of the Civil War. In English, this included the series Talking Sense, The English Language, and Language and Linguistics.

Supplementary Assignments

In history, control and experimental classes were assigned readings for Fast (13), Guthrie (16), Franklin (14), Stowe (29), Elkins (11), DeToqueville (8), and standard reference works (17). In English classes readings were limited to the essays in the departmental text (19). Students were assigned writing projects and essays relevant to the purposes as developed in the pre-semester planning conferences. In addition, all students were assigned to keep a daily journal of personal writings on events, persons, places, questions, daydreams, books, movies, television programs, classes, or any other significant experience.
The Composite of Procedures for the Experimental Group

The experimental treatment was a composite of procedures used only with the Special Program in Curricular Experimentation English and history students. This composite had certain recognizable characteristics described below.

General Objectives

As developed in pre-semester conferences with administrators and involved teachers, the overall goal of the Special Program in Curricular Experimentation was to achieve a more effective approach to student learning by making better use of teacher-student time, space, and facilities of the college. The basic concept was that of a college-withina-college. It was proposed that a small group of students taken from the total enrollment would form a learning unit which would share teachers, courses, problems, and ideas. It was theorized that under these circumstances the student's adjustment to college should be faster and better and his self-development and scholastic growth greater.

Also included in the goals was the concept that small-group sessions would lead to the student's being more at home in his new college setting and more at ease in expressing his genuine feelings. It was projected that the close interaction between student-student and student-teacher would give the opportunity to share ideas, to clear up confusions, to learn to respect others, to develop self-confidence,
and to participate in a group effort. The seminar climate would be one in which the teacher and the students would be mutual learners and sharers of knowledge.

Large-group meetings had the goals of freeing the teacher from repeating the same lecture or demonstration several times and from monitoring the same test for several hours. The purpose of the planned tutorial activities was to accommodate different levels of skill and aptitude in order that students would have the opportunity to progress at their own rates within limits specified by each instructor.

Reduced Class-Contact Time Activities

The formal schedule of the experimental classes provided for two hours of class time instead of the conventional three hours. A planned program of library research, films, slides, and tape utilization for use outside of class was developed and implemented for the experimental group only.

Tapes and slides.—As a part of their English study, students in the experimental group were assigned to listen to tapes placed in the dial-access listening center of the college. A study guide was provided for use with each tape (see Appendix). The titles of the tapes were Unity, Development, Descriptive and Factual Details, and Illustration (6). The study guides directed the students to perform certain tasks both during and following the playing of the tapes. For example, in the guide to the tape on factual details,
students were directed to mark the best models of generalizations and specific statements. They were then directed to write a paragraph to support an opinion with factual details drawn from a recent newspaper article. The guide on development directed the students to number in sequence statements taken from the tape from the most general to the most specific. They were then directed to demonstrate the differences between the general and specific by writing a specific statement based on indicated generalizations.

Other study guide activities for the tapes included the marking of neutral and slanted words, deciding whether statements were fact or judgment, and the ordering of ideas into major and minor categories.

As a part of the history study, the experimental group was also assigned to view slide-tape productions as well as to listen to supplementary tapes made by the instructor which supplied further content details. The students were expected to utilize study guides of the same type used in listening to the English tapes.

Films.—Films were assigned to English and history students in the experimental group for out-of-class viewing. Students made appointments for viewing in the media center. Film-lesson guides were also provided for use with these lessons. They included both comprehension questions as well as questions which would indicate the synthesis of ideas.
For example, the guide to the Language and Linguistics series asked how the speaker in the film defined fact and inference. It then asked whether or not the student accepted these definitions. Another guide asked for two ways of answering a question. It then asked whether or not the student saw any relationship between what the speaker called the exaggerating tendency and what another speaker in the series called polar opposites.

Library and writing laboratory assignments.—The library assignment for the experimental group involved the use of study guides relating to research tools. One major assignment sheet required that, given a topic previously unknown to him, the student must find at least ten relevant sources and label them primary or secondary and construct a logical outline of six to ten headings and subheadings.

Study guides on the process of outlining were also provided for this part of the program. One guide listed twenty historical topics, and the student was directed to select the seven major topics, circle them, and number them in the most logical order. The directions then provided that before each of the remaining topics the student was to place the number of the major topic under which it would logically belong. The student was then directed to write a title for such an outline.
In addition, students in the experimental group were assigned to make appointments for tutoring sessions in a writing laboratory supervised by a teacher aide. Study guides were provided for this activity also. For example, one guide required that sentences making up a scrambled outline for an essay be rearranged to make a logical outline.

**Systematic Use of Large and Small Classes**

The composite of procedures systematically utilized large general sessions and small seminar classes as the bases of the regular schedule for the experimental group.

**Specialized Use of Lecture and Discussion**

The instructors' use of lecture or discussion was a specified function of class size in the experimental treatment. Lecture was used exclusively in general assemblies, while in seminar sessions discussion was used exclusively.

Study guides were provided for these areas of the Special Program. The study guides for the lectures were incomplete topic outlines within which the student would write his own notes. For example, the history lecture guide on mercantilism used phrases followed by blank spaces for student use. Some of the phrases were: basic assumptions; purposes of mercantilism; *Trade and Navigation Acts*; benefits to the colonies; and *Triangular Trade Route*. This was followed by comprehension questions for out-of-class use. One English lecture guide listed twenty-three fallacies about which
the student was directed to write additional notes as the lecture proceeded.

Guides for the seminar sessions were constructed for a dialogue format. For example, the guide on fallacies gave some examples of fallacious statements around which discussion was to proceed. In history, a guide suggested discussion topics about Benjamin Franklin. These included his personality, education, business success, and views on morals and religion. Further questions for discussion explored ways in which one may estimate one's own proper worth and the relative merits of formal education as compared to self-education.

Hypotheses

The following hypotheses were formulated for testing:

I. The mean gain made by the experimental group will be significantly greater than the mean gain made by the control group on each of the following measures:

   A. The Crary American History Test
   B. The Cooperative English Expression Test
   C. The Watson-Glaser Critical Thinking Appraisal
   D. The Purdue Attitude Scale, Part A, Attitude Toward Any Subject.

II. With sex held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group.
III. With sex and ACT English score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group.

IV. With sex and ACT mathematics score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group.

V. With sex and ACT social studies score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group.

VI. With sex and ACT natural science score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group.

VII. With sex and ACT composite score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group.

VIII. With sex and pretest score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group.
Definition of Terms

The following definitions were developed to apply to certain terms which were considered basic concepts of the study:

Score held constant.—The statistical comparison of scores made on control and criteria measures which may fall within a range of ± .25 of a standard deviation.

Conventional grouping.—A pattern of enrollment of twenty to forty students meeting as a single discipline class with an instructor three hours a week.

Special Program in Curricular Experimentation.—A pattern of enrollment for instruction in large groups of approximately 150 to 180 students coordinated with seminar classes of approximately 15 students and guided independent study.

Achievement.—Scores attained on the standardized measures utilized.

Limitations of the Study

The following limitations were recognized as being applicable:

1. The data would tend to suggest gross or molar effects of the grouping matrices rather than molecular cause and effect relationships.
2. The student body of the experimental institution may be unique in that its 1967-68 freshman class had a mean of 15.91 on the ACT composite score as compared to the mean of 18.05 for the 1966-67 Texas college-bound students.

**Basic Assumptions**

1. A basic assumption was that resources for keeping content and methods comparable in the experimental and control groups were adequate through supervisory status of the investigator and commitment to the experiment on the part of the experimental institution.

2. Generalizations which may be drawn from this study would be limited by human variables and various other factors not subject to complete control.
CHAPTER BIBLIOGRAPHY


2. Bane, C. L., "The Lecture vs. the Class Discussion Method of Teaching," School and Society, XXI (March, 1925), 300-302.


CHAPTER II

SURVEY OF RELATED LITERATURE

Literature relating to the problems of higher education has increased since 1945 and the conclusion of World War II. Eckert (21) indicated that pressures of steadily mounting enrollments as well as higher aspirations of students and the society have brought these concerns into sharper focus. She also suggested that additional factors have been rising costs, the scarcity of qualified teachers, and the increasing heterogeneity in the background of students. In this context, Knoell's study (46) was undertaken to determine needs for college services in urban centers of population for persons who have not been reached by traditional institutions and practices in higher education. She concluded that while the need for society to offer opportunity for post-secondary education to all cannot easily be proven with research, there may be now what amounts to a national goal of offering such a universal opportunity (46, p. 193). Further indication of the growing extensiveness of the investigations in the college area is supplied by Bells (24) whose continuing annotated bibliography in 1962 raised the total references since 1945 to 5,656.
An emphasis on implementing innovative practices characterizes much of the reported activity in both established and new colleges.

In describing some of the new four-year institutions, Mayhew (57, pp. 19-20) suggested that there are several new trends: a desire to combine the values of small colleges with the virtues of large size; a commitment to the liberal arts tradition in curriculum; the use of media and automated instruction; acceptance of a broad range of student ability; provisions for individualized instruction; and an attempt to create the feeling of intellectual community. Baskin (4, p. 181) summarized the new developments in post-secondary education as including new organizational and structural patterns, focus on curriculum revision, better utilization of staff and plant resources, and new teaching-learning methods.

Some general studies have organized the presentation of theoretical positions by means of essays by authorities in the field in which data and opinions have been subjected to assessment and commentary by colleagues holding divergent views. Lee (49) examined innovations in college teaching in this manner. He concluded that innovations in classroom techniques are frequently ignored or condemned before any effort is made to explore their practicality. He indicated that society, students, and changing roles necessitate assessment with a view to instituting reforms (49, pp. 2-3).
Wilson (86) has also viewed new developments in a compendium format to explore the position that they represent a changing environment which will result in basic institutional modifications. In another similar collection of position papers (72), the conclusions related to the need for research, consultation, and initiation of changes by means of special understandings and techniques (72, p. 1014). Havighurst developed a thesis in 1960 that the future of higher education will be shaped by the forces of economy, ideology, and population (40). As early as 1957, the Committee on Education Beyond the High School pointed to an oncoming increase in college-age population and predicted a conflict between the demand to expand enrollments and the needs for quality (67). Lineberry has organized a further examination of the quantity-quality dilemma in which the conclusion was proposed that, in some respects, students have much of substance to complain about (53, pp. 49-50), with rising enrollments and mounting costs acting as further perplexing influences.

Emphases on Junior Colleges

As the emphasis on problems of post-secondary education has continued to gain attention, the role of the two-year college has received intensified exploration. Two presidential bodies have pointed to the potential of these institutions (41, 67). One recommended that the number of community colleges be increased and the activities multiplied (41, p. 67),
while a later one suggested that these institutions meet the needs of greater variety and accessibility at the same time having a role and integrity of their own (67).

Medsker (59) conducted a survey study of seventy-six two-year institutions in fifteen states. He concluded that, while they had made significant contributions as a democratizing agency in adding diversity to higher education and producing students successful in transfer to higher institutions, he did not consider their records as impressive in emphasis on terminal education, student personnel services, and the achievement of institutional identity. Reynolds (68) interpreted the growth in number of these institutions and the accreditation of their programs as evidence of the inevitability of continued growth, grounded in part on vast social changes of the twentieth century.

An emphasis on innovation in curriculum and instruction has characterized many of the reports about these two-year institutions. Blocker (8) has provided an explanation of this as being related to needed modifications based on the variety of motivations and abilities of the students (8, p. 229). Tyler (83) has suggested that it is especially appropriate for the junior college to examine teaching methods and strategies carefully and reject those which are inappropriate. He indicated that the reasons older methods may not be appropriate to junior college students may be explained by a lack of motivation toward abstract thinking,
minimal skill competency, poor academic work habits, and competing out-of-class environments (83, pp. 526-527).

The teaching obligation in junior colleges has found institutional expression in instructional innovations in two studies conducted by Johnson. One was a collation of reports by representatives of several two-year institutions concerning their activities in planning and implementation of new instructional programs and techniques (44). The other was an exploratory survey of the utilization of junior college faculty services (45). Among major findings on class size he noted facilities in some junior colleges for large classes and the use of innovations in the class schedule (45, pp. 10-11). His study described several programs but gave no indication as to provisions for evaluation of results (45, pp. 30-32; pp. 47-58). In this context, Eckert (22) commented that "unfortunately" little research could be reported dealing specifically with the instructional program of the junior college (22, p. 295).

Experimental studies in teaching junior college American history and English are not plentiful in the literature at this time. The general approach is one of survey and theoretical exploration of curriculum rather than of comparing and testing components of teaching methods.

In exploring junior college social science offerings, Gross (38) found that the number of courses offered in the social sciences increased with the size of the school. Less
than 3 per cent of the colleges surveyed offered an integrated social studies course. He concluded that the aspects of the programs analyzed revealed the domination of traditional western civilization-type history courses and near minimal offerings on important regions, countries, and continents. He interpreted the findings as suggesting that the social science offerings in many junior colleges are dominated by the freshman and sophomore social science requirements of the state university to which many of the students aspire to advance.

Two reports by the National Council of Teachers of English surveyed the problems of research and development of English programs in the junior college (62, 63). Both of these reports were status studies with recommendations for administrative and logistical procedures. They both suggested needed research strongly oriented to the gathering of information. Both gave space to the problems of keeping junior college English instructors comparably competent with their senior college colleagues.

Methods of Instruction in College Teaching

The general literature pertaining to methods of instruction is broad and voluminous. For this reason, the emphasis of this review will be on that which is pertinent to the specific area of college instruction.
Although Canfield (13) and Eckert (22) indicated that there is a noticeable lack of research dealing specifically with the program of instruction in the junior college, there is a useful body of empirical investigation on undergraduate instruction which is relevant to the components of this study. For example, Umstattd's survey (84) of twenty-eight colleges, while pointing up the problems of evaluation when attempts to use different methods are not accompanied with descriptions of context, did indicate increased activity with teachers using a greater variety of methods than had been reported in an earlier survey.

**Studies on Lecture and Discussion**

In an early study, Bane (2) reported controlled experiments involving 510 juniors and seniors in university departments of education and psychology. Five experienced instructors participated and cooperated in keeping constant in both a lecture and discussion section the factors of teaching time, subject matter, identical assignments, objective tests, standardized intelligence tests, and efforts to make each method yield the best results possible. He defined lecture as a class situation in which the teacher did all of the talking and discussion as a class situation in which both teacher and students participated. From the statistical treatment of the standardized and objective instruments, he concluded that the lecture and discussion methods appeared
equally effective in immediate recall of material, but that the discussion method was more effective in delayed recall.

Spence (77) also was among the early investigators reporting studies of the lecture and discussion methods. He concluded that for two large sections of graduate students in educational psychology, meeting once a week for two hours with little previous experience or feeling in favor of the discussion method, the lecture method was superior to the discussion method in producing improvement in things measured by the tests. He concluded his study with the generalization that there are times when the lecture method is effective and there are other times when the lecture method is harmful.

The lecture was compared to reading in effectiveness in another early study by Greene (36). He reported finding no significant differences between the average scores of those who heard a lecture and those who read the same material. His population was two groups of college men of nearly equal ability totaling about 125 men each. He did find some indications that the highest quarter of students tended to benefit more from the reading while the opposite was true for the poorest quarter of students.

Pursuing the same field of investigation, Corey (16) applied correlation techniques to lecture and reading results and measures of vocabulary, reading ability, and a psychological test. He also investigated the dimensions of immediate and delayed recall. He interpreted his results as indicating
that immediate recall was better for read materials, that there was no significant differences in effects upon delayed recall, and that the correlations were higher between read materials and standardized test results than for lecture and the same type of tests. He also found that students in the highest quartile did better on reading than on lecture tests.

A survey through 1937 by Cole (15) summarized the results of ten studies. She interpreted their results as indicating that the lecture method was as good as any other with either identical or varying results favoring either one or the other method slightly from study to study.

The effects of participation as compared to the non-participation characteristics of the lecture situation were studied by Gibb and Gibb (32). Eleven sections of a general psychology class were used. Ten sections were taught by traditional lecture methods and the eleventh by participative action methods without formal lectures on course content. Reading and subgrouping methods were used instead, and the instructor played a minimum role. In pretest and post-test measures given to the experimental group and one randomly selected control group, the experimental group was reported as having made significant gains in role flexibility and leadership related ratings with no apparent loss of normal content acquisition.
In 1953, Ruja (70) provided a survey of contemporary research pertaining to experimenting with discussion in college teaching. After a summary and critique of seventeen related studies, he raised several methodological questions which he recommended be translated into specific research issues: the relationship between subject matter and method; the importance of differences in instructors; the importance of differences in students; the importance of the differences in social climate; and the long-run effects of the two methods and ensuing outcomes when methods are combined. He also concluded that while research in this area had been energetic, it had been diffuse and undisciplined because of lack of uniformity in procedures (70, pp. 335-341).

Bloom (9) has reported attempts to answer questions related to the thoughts students have during lecture classes as compared with discussion classes and the relationship of these thoughts to differences in the two methods of teaching. He used a method of stimulated recall with auditory cues as the bases of recall. In comparison of percentages of thoughts about other persons and the self reported in lectures and discussions, the thoughts in which the student applied the ideas being considered were approximately the same for the discussions and lectures. However, the thoughts in which the student attempted to answer a problem or question or to synthesize and integrate ideas were only 1 per cent in the lecture as compared with 8 per cent in the
discussion. Four-fifths of the discussions evoked more thoughts of evaluation and synthesis than the average lecture. Bloom concluded that the lecture was more efficient in communicating knowledge about a topic or field, but that in the development of abilities and skills which were problem-solving in nature, discussion was superior.

A doctoral dissertation by Warren (85) was concerned with the effect of variations in proportion of lecture to recitations in teaching physics to engineering students. He compared a one-lecture and four-recitation ratio pattern with a two-lecture and three-recitation ratio pattern, keeping total number of hours and sizes of classes the same. His criterion measure was the score on the Cooperative College Physics Test. He found that for students of medium ability the two plans were equally effective in all of the topics tested, but there were variations in effectiveness among students of high and low ability. Student opinion indicated that the two-lecture plan was popular among high ability students and unpopular among the low ability students. In all ability groups, however, he reported a significant tendency toward the opinion that the two-lecture plan was best.

Guetzkow (39) conducted a study in which there was a greater variety in teaching methods than in most other experiments which had been reported up to that time. Three methods were used. In method one (recitation-drill), the instructor
directed, paced, dominated, and administered a quiz at each class meeting. In method two (group discussion), the meetings were regarded as supplementary tools used in conjunction with the text, lectures, and readings, operating on the assumption that telling the student or having him read was insufficient. Method three (tutorial-study), emphasized an individualized approach with the student pacing his own work within the framework of the course. These methods were used among twenty-six small class sections as part of a pattern in which each student spent one hour a week in a large lecture section and two hours in the small sections. Total population was 865 students enrolled in elementary psychology at the University of Michigan.

On the criterion measure of final examination, method one ranked first, method two ranked second, and method three ranked third. The recitation method was found to be more effective for all levels of intelligence in the group. Results of an attitude-toward-subject scale showed students in method two sections slightly above the method three sections while students in method one sections did not differ significantly from either of the others. A significantly larger number of students in recitation sections chose to concentrate in psychology than did students in the other two types of sections. The study concluded with the general statement that, by and large, no differences between the three teaching methods from the point of view of educational outcome could be reported.
Eglash (25) designed an experiment to compare achievement and students' reactions in a group-discussion class and a lecture class. He reported that while keeping the emotional climate similar, an instructor taught one elementary psychology class by group discussion and the other by lecture. Such matters as examinations, grades, and course content were covered by discussion in the experimental discussion class. The instructor imposed the same decisions on the control lecture class and covered the same material by means of lectures and demonstrations. Achievement on course content was not significantly different, but the morale of the lecture class was reported as being significantly higher than that of the discussion class.

Additional data on outcomes of lecture and discussion were sought by Ruja (71) in three college courses. The experiment sought to assess the validity of the hypothesis that students in discussion classes in comparison with students in lecture classes would show greater subject-matter mastery, greater gains in emotional adjustment, more favorable attitudes toward the instructor, and become better acquainted with other students. His criteria measures were an objective-type examination, the Bell Adjustment Inventory and a Likert-type instructor rating form. Two philosophy classes and one psychology class were used, with one lecture and two discussion sections the first semester and the opposite design the second semester. Results suggested that lecture
proved superior in subject-matter mastery for the psychology students. Discussion proved superior in names known for all classes. In attitude toward instructor, discussion was superior in the philosophy classes. In all other regards the two methods showed no significant differences.

The dimensions of group effort, sociometric factors, and leadership were included in the design of a lecture-discussion experiment by Smith (76). He divided a class in general psychology with forty-eight students into five subgroups. The groups were equated for initial achievement but were varied systematically in their attitudes toward class participation. Each member was matched with a member of a lecture class taught by the same instructor. The group-discussion class made greater achievement gains than did the lecture class, and the trend in four different criteria measures was significantly upward. Compared to controls in the lecture class, however, students with low anxiety and high initial achievement scores in the team or discussion classes were less satisfied and made smaller achievement gains than those with high anxiety and low initial achievement.

Maloney (55) applied sociological research techniques in an investigation of discussion as a learning medium. The study evaluated the effects of a leader following a precise method of implementing discussion and a leader without a precise method. Results of two achievement tests indicated that learning was neither facilitated nor impeded in the
precise method group. Analysis of discussion content tended to suggest that learning was facilitated in the precise method group. Other criteria procedures of analysis favored learning in the precise control group as indicated by verbal participation, behavioral participation, and more unity as evidenced in the geometric configuration of spontaneous seating arrangements. Group cohesion, as measured by sociometric choice, also increased in the precise method group.

A doctoral dissertation by Burke (12) compared a class-whole pattern of organization with class sections of cooperative-group patterns. In the conventional pattern, the instructor assumed a central position and set the goals. In the experimental pattern, efforts were made to develop small-group cohesiveness. A total of sixty-nine test items from twenty-two tests were used for raw data and treated to obtain F-ratios. Significant F-ratios were obtained on four items only. It was concluded that the evidence suggested that either pattern of instruction may be used with equal confidence.

A course in physiology in a medical school provided the setting for a further comparison between lecture and conference methods of instruction (52). The control group's weekly schedule consisted of six hours of lecture, six of laboratory, two of demonstration, and one of a conference or seminar meeting. The work of the experimental group differed in that discussion sessions were substituted for about one-third of
the lectures. The hours thus freed from lectures were devoted entirely to student discussion of papers. The total scores of the experimental and control groups of both mid-quarter and final examinations were almost identical. Students seemed to achieve better on items constructed by their own instructors. The authors found evidence, however, that the experimental group outperformed the control group on application-type items. They also found that the attitudes of the experimental group were significantly more favorable than those of the control group as measured by weights assigned to student responses to questions related to this area. The conclusions were described as being limited, however, because of defects in experimental design.

A doctoral dissertation by Deignan (19) reported an experiment using as subjects 130 fine arts college freshmen divided into ten sections in general psychology. They were systematically assigned to experimental or control sections on the basis of the alphabetical order of their surnames. The experimental variable was the method used by the instructors. In experimental sections, students were given the responsibility for the initiation or the change of discussion topics. The control group leader conducted classes with lecture methods more usually used in college teaching. He chose the topics, asked questions of the class, corrected inaccurate or incomplete statements, and maintained complete control. Post-test data on which analyses of variance were
completed required the rejection of the major hypotheses of the study relating to subject-matter gains of the experimental group, except that frequency of participation was found to be positively related to subject-matter gains for the control group and negatively related in the experimental group.

Marr (56) reported a study which compared two methods of conducting an introductory course in psychology, four lectures a week versus one question-and-answer period a week. Each of four instructors taught two classes, one class by each of the two methods. All classes had the same reading assignments. An analysis of variance design was employed to determine the effects of different instructors, different methods, and different levels of students as measured by grade point averages on the four examinations given. All interactions resulting from the analysis were examined, and none was found to be consistently significant.

Lancaster (48) has reported an experiment in which the achievement of students in college physics increased significantly as the number of discussion-recitations increased and the number of lectures decreased. To test the achievement made in instruction by lecturing with that of instruction by discussions, the second physics course for engineers on the subject of electricity and magnetism and light was given by three distinct procedures: three recitations, two lectures and one practicum per week; two recitations, two lectures and one practicum per week; one recitation, three
lectures and one practicum per week. All groups followed the same general outline of subject matter. Students for the three treatments were selected at random. The same one-hour tests and final examination were given to all students. The composite numerical grade the students made in the first college physics course was used as a correlating variable. An analysis of covariance was made to determine whether the results were attributable to the different methods of instruction. A highly significant F-value of 5.20 was found for the low lecture group. It was reported that the analysis indicated that the differences could not be attributed to differences in teachers or to any assignable cause other than the treatments.

Leton (50) evaluated course methods in teaching child development. There was a total of 145 students in the study which made up three sections of the course. The sections were taught utilizing three methods: lecture, case-centered, and group-centered. Although positive attitudinal change as reflected by the Minnesota Teacher Attitude Inventory was a definite course outcome for all sections, no one method of teaching proved superior to the other methods in producing this change.

In a related doctoral study with high school American history students, Barratt (3) compared growth in critical thinking, use of sources of information, and mastery of factual material in a teacher-dominated lecture class with
the growth attained in the same areas by students taught in a classroom-laboratory environment. There seemed to be no significant difference in growth in critical thinking abilities and in growth in factual material mastery between classes taught by the lecture method and the laboratory method, although the lecture-method group did significantly better in growth in ability to use sources of information.

Elkins (26) investigated the generation of new ideas as a result of subgroup discussion participation with grades seven and eight in American history. A post-meeting reaction questionnaire was used to measure student satisfaction with classroom activities, and an idea questionnaire was used to trace the origin of students' ideas. Student and teacher verbal behavior was recorded on tape. Under teacher control, the class described as quieter and deliberate produced more ideas.

Pape and Miller (66) described an experiment in teaching college sociology. Rather than assigning students to lecture for three hours each day for five days a week, the class of 120 freshmen was broken into 11 cells of 10 or 11 students each, with interacting assignments. They reported that, on the basis of test performance and personal evaluation of the instructors based on conferences with students and general observation, the students enrolled in the Principles of Sociology course learned sociology principles
as well as, if not better, than they would have had they taken a conventional lecture course.

Gayles (31) published a recent collation of research findings on the lecture versus discussion problem. After summarizing briefly fifty-nine relevant references, she concluded:

The skillful teacher has many methods and techniques at his command. Not one of them can be regarded as the best, for there is no best technique. In fact, techniques which are good for one subject or for one group of students may be quite unsatisfactory for another. First, teaching is a highly variable performance. Second, teaching procedures not only vary greatly, but the factor of appropriateness further complicates analysis. A third factor which makes comparative studies practically impossible is that different general patterns of instruction serve, in fact, different purposes. Depending on which authoritative analyses and experimental investigations are selected, lecture and discussion methods can be supported as the best or poorest of all possible teaching procedures (31, pp. 98-99).

Studies on Class Size

The early study by Edmondson and Mulder (23) delineated the persistent issues of this type of research problem. During the first semester of the school year 1922-23, the authors conducted an experiment at the University of Michigan to determine what differences existed in the measurable results of instruction in a large and a small class in an education course offered at the University of Michigan. Their conclusions were: results in terms of semester averages indicated no appreciable difference in achievement; student
opinion favored small classes; the chief advantages in a small class appeared to be more personal contact with the instructor; the chief advantage of the large class was the contact with varied opinions; and physical properties of the room made the large class less effective.

Soon thereafter, Mueller (60) did a follow-up replication of this experiment and reported that the small group was 17.5 per cent superior in achievement in psychology when measured by an objective test. He questioned the methodology of the Michigan experiment for not having a valid criterion of achievement and for using classes of forty-five students as the small classes. He indicated that his experiment did not determine the exact size of class for the greatest efficiency of instruction, but that on the basis of his findings, it might be reasonable to assume that the critical point lies somewhere below forty-five since an appreciable difference existed in the efficiency of instruction in classes of twenty and forty.

Husband (43) applied a later statistical sophistication to the large versus smaller class problem in a study with a population of approximately 1,700 Iowa State College undergraduates covering six quarters. Lecture sections averaged about 200, with a range from 140 to 320. The smaller classes were from 30 to 60, usually 50 to 55. Final course grades were computed by totalling all hourly quizzes and the final examination, and combining all sections of all instructors
into one master distribution. Out of 275 possible points, the median was usually about 185 to 190. He reported the main results were that the lecture sections averaged for the six quarters three points higher than the smaller classes, with superiority appearing five quarters out of six and the sixth coming out a tie. In no case did the smaller group earn a higher course total. He concluded that the often cited disadvantages of large classes at least do not lower the final grade.

A further concern with the advantages versus the disadvantages inherent in group size prompted Gibb (33) to attempt to determine the effects of size of group upon creativity and the effects of induced reduction of threat upon creativity in groups of various sizes. He used 1,152 college students selected at random in elementary psychology classes. The groups of varying sizes were given a set to produce as many solutions as possible to a series of problems permitting multiple solutions. The groups varied in size, containing one, two, three, six, twelve, twenty-four, forty-eight, and ninety-six members. Under Condition A, the groups contained an equal number of each sex seated alternately. Under Condition B, the conditions were the same with the exception that the experimenter read a different set of instructions meant to reduce threat. Data were taken by absolute number of solutions produced and scores weighted for independently judged "goodness" of solutions. Results
reported were interpreted as indicating that the number of ideas produced was found to increase in a negatively accelerated function of size of group in each of the conditions.

Stephan and Mishler (78) devised a simple exponential model to study distribution of participation in small groups. The size of the group was found to be an important parameter affecting the size of the ratio in the basic equation of the model used in estimating participations by linear functions. That is, as groups increased in size, members of adjacent ranks became more like each other in their relative rates of participation. It was postulated that three conditions had to be fulfilled for this model to be applicable to a study of group size: a range of verbal participation potential among the members, no systematic interference with the "free competitive expression" of these potentials, and a lack of well-differentiated roles among the members.

Rohrer (69) attempted to evaluate objectively some measured changes in college students who completed a beginning course in American Government. The two experimental variables operating were size of class and lecture versus discussion method of presentation. Two criteria were used in the evaluation of the variable manipulated: the students' measured attitudes toward the difficulty of the course material and their interest in it, and the measured achievement of the students. Three instructors took part. Instructor A had a
large and small class, Instructor B had a large and small
class, and Instructor C had two small classes. Seven addi-
tional variables were either equated, controlled, or con-
founded in the experiment. These were the age of the student;
sex; veteran and non-veteran status; classification; academic
aptitude of the student as measured by the Ohio State Psy-
chological Examination; the degree of mastery of the student
in the field of American Government at the beginning of the
course as measured by the Cooperative American Government
Test; and the subject matter taught in all sections. Students
were not assigned to sections through a random-sampling
procedure. Enrollment depended upon fitting the course into
the schedule of the other courses that they were taking. A
chi-square test was computed on scholastic aptitude and
subject-matter scores obtained at the beginning of the
experiment to determine if the enrollment procedure resulted
in a distribution that favored one or more of the sections.
The value obtained was significant at the 30 per cent level
of confidence, indicating that the distribution of aptitude
and knowledge at the beginning of the course could be
reasonably considered the same for each section.

A measure of learning, by sections, was available by
comparing the preliminary and final test scores earned by
the students in the various sections on the Cooperative
American Government Test. Although there was a mean score
gain for all sections, none of the observed differences
between small sections was significant statistically. The average gain per pupil for all sections was statistically significant beyond the 1 per cent level of confidence. On the Cooperative American Government Test, a slight average difference of 0.80 was observed in favor of the large sections. The achievement scores of the various groups did not differ significantly by sections. Felt difficulty expressed by the students was significantly related to the teacher and to the size of class, and there was interaction between these two variables. According to the investigator, the most significant finding of the study was that the amount of achievement, as measured by standardized tests, and the attitudes of students toward American Government varied as a function of the course instructor and did not vary as a function of size of class. No statistically significant differences were observed between the small classes taught by the lecture or discussion methods, but differences were revealed in the achievement of students when taught by different instructors.

Nachman (61) compared twenty-one students in a small class on examination performance with a matched group of students who were in a large class. He hypothesized that the small class would do better on quizzes which specifically covered the classroom material and for which the students had not prepared, but that the two groups would do equally well on final examinations for which they had studied. The hypothesis was confirmed. The investigator proposed that
the effects of different teaching methods have been masked in the measurement process and that in order to test the effect of a particular classroom technique evaluation should be done immediately after the technique is employed since waiting until the final or another announced examination may confound the problem by permitting many other variables to operate.

Siegel (74) conducted a series of studies at Miami University related to the use of large-group teaching procedures at the college level. The primary purposes were to investigate the effectiveness of large-group instruction and to find ways of improving upon the quality of the educational experience thereby provided for the students. Three general approaches to large-group instruction with eleven courses were investigated: large sections with relatively little discussion in class; large sections taught by a combination of lecture and problem-solving or case-study techniques; and multiple sections of approximately thirty-five students, each receiving simultaneous instruction by television.

He reported that large-group instruction did not have deleterious effects upon achievement as measured within the framework of the study. He indicated, however, that the issue of comparative achievement under the experimental and conventional conditions of instruction was far from settled. This was the case because of possible defects in the criterion instruments that purported to measure achievement. Additional
research based upon the use of unique measures emphasizing other kinds of achievement than the retention of facts was called for. He also suggested that achievement for the most part has been restricted in scope to immediate rather than to long-term retention. He investigated this problem in a later study (75). The purpose of that investigation was to determine the effect of instructional procedure upon subject-matter retention one year or more after completion of a course. Retention scores of students who had received conventional instruction in nine courses were compared with the scores of students who had been instructed in the same courses by large group procedures. The retention test scores were compared with control on total ACE scores for eight of the courses and on the Cooperative English Test score in the ninth course being exercised by analysis of covariance. None of the F-ratios approached statistical significance. The conclusion was offered that retention of subject matter a year or more after completion of a course is not adversely affected by increased class size.

A related study of the effectiveness and feasibility of using certain types of large group instruction at the college level was also reported by Macomber (54). Courses were taught by television, by lecture, by a problem or case study approach, and by graduate student assistants. Evaluations involved comparisons between these experimental methods and
conventional instruction. No great difference was found between academic achievement in experimental and control classes.

Schellenberg (73) designed a study to discover whether there was any relationship between the size of a group and measures of student satisfaction. Thirty-two academic discussion groups were varied in sizes from four to ten students to measure the effect of group size upon student satisfaction, instructor satisfaction, and student achievement. The most significant finding to the investigator was a consistent inverse relationship between group size and student satisfaction. Students claimed greater satisfaction in the smaller groups. A second reported result was the differences between the perspective of instructors and those of students. Instructors were more inclined than students to show satisfaction with larger groups. There was limited evidence that smaller groups showed slightly higher academic achievement than did larger groups.

Griffin and Bowers (37) designed a study to determine whether the change in student performance brought about by two methods of teaching Fundamentals of Speech was significantly different. The first method involved speech classes of approximately twenty students meeting three times a week with its instructor and assignments made in a standard text. The second method of teaching employed one mass lecture per week by an experienced member of the department. The following variables were controlled: the same graduate assistants
served as instructors for both experimental and control groups; all classes in both groups met in the morning on the same days of the week; control and experimental classes met in the same or comparable classrooms. Students were evaluated by their performance in knowledge of speech principles and speaking skill. The design was replicated for a second semester. According to the results reported, there were no statistically significant differences in the post-test performance of groups subjected to the two methods of teaching. Both groups, however, made significant positive changes in knowledge of speech theory and in speaking skill between the beginning and the end of each of two semesters. Measurable results of the two methods compared were almost identical. The investigators concluded that it appeared that the mass-lecture method of teaching could be employed without significantly diminishing the amount of learning.

A critical review of studies conducted on the sociological aspects of group size has been made by Thomas and Fink (79). On the basis of their review, they suggested that it is apparent that group size has significant effects on aspects of individual and group performance, on the nature of interaction and distribution of participation of group members, on group organization, on conformity and consensus, and on member satisfaction. They suggested that the variable of group size should be included in theories of group behavior distinguishing where possible between the effects that result
from the interaction of group size with other independent variables and the effects arising from intervening variables that are dependably and nondependably associated with size. They also concluded that future research on group size should proceed systematically making efforts to vary size in complete sequence over a suitably large range, to identify relevant intervening variables, to determine in advance whether these variables should be expected axiomatically to be correlated with size or would be only contingent variables, and to use multivariate designs in which variables are manipulated experimentally.

Feldhusen (29) studied the learning outcomes in small and large class sections of educational psychology. A group of twenty-two students enrolled in a large class of seventy-two students in educational psychology learned the subject matter as well as a matched group of twenty-two students enrolled in a smaller class of twenty-eight students. Terminal measures of interest in teaching and evaluation of the course and teacher revealed no significant differences between the groups. However, terminal Minnesota Teacher Attitude Inventory measures were significantly greater for the students enrolled in the small class. The investigator noted that 40 per cent of all the students in the large class indicated a preference for a small class.

The dimensions of achievement and attitude have been investigated by Eash and Bennett (20) in relation to class
size. Three classes of 189, 208, and 167 students who took general psychology during the fall, winter, or spring terms of 1961-62, and for whom entrance-test data were available composed the experimental group. Three classes of 31, 36, and 46 students who took general psychology in the winter or spring terms of 1960-61, and for whom entrance-test data were available, composed the control group. During the fall and spring terms of 1961-62, the same professor who had taught the control group classes the previous year delivered all the lectures to the large classes. During the winter terms, another professor gave some of the lectures to the large class, keeping content as similar as possible. To obtain data on attitudes and perceptions, random samples of sixteen students were drawn and subjected to focused interview techniques. Results indicated that the experimental group (large lecture classes plus small discussion groups) did significantly better than the control group (lecture-discussion classes ranging from 31 to 46 students) on objective tests covering the content of a general psychology course. The two groups did not differ significantly in gains on an objective test of psychological misconceptions. The two hypotheses of the study (first, that students taught in the large-lecture-class-plus-small-discussion-group pattern would perform as well on objective measures as students taught in conventional lecture-discussion classes ranging in size from 30 to 50; second, that there would be
an absence of negative effects on students' perceptions of selected referents in the course from having experienced large-class instruction) were supported by the results.

The research division of the National Education Association has a research summary on class size in process (64). The report suggests that research findings do not indicate that there is one best class size nor one best teacher-pupil ratio. The report summarizes results from several studies illustrating positive effects which small classes have on teachers' behavior such as inventing of more new practices, giving of more attention to individual pupils, and using a greater variety of instructional methods. Other studies, however, found teachers of small classes using mass instruction in small classes. A conclusion suggested is that while few significant differences have been found between small and large classes, the results indicated not that large classes are as good as small classes, but that small classes did not prove superior on the criteria measures.

Brasted (11, p. 63) has proposed the use of very large lecture groups as a solution of the logistical problems of teaching general college chemistry. The technique was reported to require improved stock control and attention to safety practices. It was suggested that instrumentation in major and honors classes should vary from that in the general classes.
Trump (82) has warned that working with small groups challenges very high professional competences on the part of the teachers. He suggested that a study of group dynamics helps in the development of these competences. Acquaintance with the principles of sociometry and behavioral psychology was also suggested. He further suggested that if teachers do not take advantage of the unique factors of small groups there is no sound educational reason for such an organization. Trump (81) has also suggested that large-group instruction has three types of purposes: motivational, informational, and directional. He suggested an emphasis on the necessity of varying the size of student groups with the purposes of learning. Unless there is a structural relationship, he holds the position that the potential outcomes will suffer.

After reviewing the history of the debate on class size and the resulting inconclusive generalizations, Goodlad (35) proposed that although experimental evidence does not provide a clearcut answer to the class-size issue, the general trend of the evidence places the burden of proof upon the proponents of small classes. Burich (27) suggested that the issue is no longer one of class size, but one of how the whole range of possibilities can be used to bring the best instruction to the student. He proposed that clusters of people and clusters of jobs be related to clusters of objectives as the modern approach to college instruction.
Studies on Reduced Class Contact

The most often used designation for reduced class contact options is independent study. Beggs (7) has solicited articles from professional educators and incorporated them in a single reference on the subject. Most of the articles discuss independent study as an extension of the classroom or as a ramification of team teaching. Not accepting this definition of independent study, parts of the study advise against a mere individualization of subject matter limiting the student's responsibility to doing the drills, reading, and exercises expected of him by adult authority.

Trump (80) has indicated that independent study for most students has two dimensions. First, the pupils need to master the minimum essential knowledge and skills required in the subject. The second level of independent study provides for enriching knowledge and skills, referred to as depth education. He proposed that independent study may involve several patterns of components. One component is a resource center where students read, write, listen, view, think, and at times talk with other students. Another component is the library where students may work who prefer that type of environment and have access to the storehouse for books and other materials. Another component is the small-group conference of three or four students discussing projects, exchanging ideas, and making suggestions for further investigation. He defined the teacher's goal as
becoming increasingly dispensable as students develop more responsibility.

Dearing (18) has added further support to the idea that independent study may occur in a variety of ways. He warned that it is easy to confuse by generalization, and that the nature of the independence under consideration should be made explicit. Among the forms of independence he listed was one involving reduced class-contact time in which the student is freed from attending a number of his regular class meetings but is expected to cover, either on his own or in teams or groups with other students, the material that might normally have been covered in the class sessions from which he has been excused.

Increased attention to these approaches to learning was indicated by a 1957 report by Bonthius and his colleagues (10) which listed a total of 334 such programs in 256 institutions. A report of the Fund for the Advancement of Education (30) listed some 16 institutions that have conducted experiments in the use of independent study as a part of their regular teaching procedures. In a later study, Felder (28) surveyed a total of 520 institutions which offered four-year degree programs and had enrollments exceeding 200. Responses were received from 445 of the institutions, 68 percent of which responded that independent study was used in undergraduate instruction. In a few cases, however, practices were described that the investigator classified as
reflecting a limited use of independent study rather than total commitment to its basic principles.

Baskin (5) has described experimental programs involving new arrangement in the use of instructor-student contact time. At Oberlin College, an experiment was designed to test the hypothesis that students participating in freshman courses and working independently for one-third of their regularly scheduled class time would learn as well as a comparable group of students who met with their instructor the usual number of times. At Antioch College, an experiment reduced class time by amounts ranging from 20 to 60 per cent in a total of nineteen different courses. At Vanderbilt University, an experiment reduced class-contact time by amounts ranging from 25 to 50 per cent. Dearing (18, p. 65) has reported, however, that in the few experimental studies available to date, notably at Antioch and Oberlin, the results were not as significant as proponents of the independent study approach might have hoped. He reported that few differences have been found between achievement of students working independently and that of those taught in conventional classes.

Craig (17) has investigated some psychological factors of directed versus independent discovery. He attempted to determine the effect of directing the learner's discovery of established relationships upon retention and the ability to discover new relations. Each of two groups of college
students was given a different amount of direction during discovery of the bases of determining solution of multiple-choice verbal items. The group receiving the greater direction, including short summary statements of the bases determining correct responses, learned more relations in each of three trials. Three days after learning, the two groups did not differ in the proportion of learned relations retained. After thirty-one days, the directed group retained a greater proportion of those learned than the more independent discovery group. He interpreted the data as suggesting that teachers and experimenters should be liberal with information designed to assist learners in the discovery of principles. He concluded that large amounts of external direction may help to insure that the learner will have an adequate background of knowledge to direct his future discovery.

McKeachie (58) has investigated an individualized program for beginning students in psychology. The program was comprised of two alternative plans: regular tutorial and laboratory tutorial. In the regular tutorial section, the semester was divided into three segments involving an increasing amount of student initiative. During the first five or six weeks, the students read an assigned introductory textbook and met in groups of fifteen students with an instructor for discussion. After this initial period, each student read in some specific area of his choice. In the
final phase, each student designed and executed an experimental study. In the laboratory tutorial, the textbook was assigned and additional readings were done in the experimental literature. There was one two-hour seminar session each week. Students in conventional sections attended two lectures a week and two hours of discussion section. The results indicated that the conventional method for teaching was superior to the tutorial method in communicating information as measured by a multiple-choice examination, but the differences were not significant on other measures of cognitive outcomes. The tutorial students, however, were more favorable in their ratings of the course and on other motivational measures than were conventional students.

A doctoral dissertation by Churchill (14) compared lecture-discussion with two methods of independent study (small groups and individual work) in seven representative first-level general education courses at Antioch College. In both independent study methods, class time was reduced roughly one-half. One instructor taught all sections of each course. Content and other features were the same. Students were randomly assigned to treatments.

The basic statistical procedure was analysis of variance and covariance. Analysis of covariance was used where necessary to control differences in background among sections within a course or to take account of correlation between pretest scores and gains. In all courses, all sections
gained substantially and significantly in achieving course objectives measured by multiple-choice tests. For the courses as a group, sections taught by different methods did not differ in gains on these tests. Gains in skills and attitudes favoring independent study were irregular, slight, and generally non-significant. Students were satisfied with instructors and courses and with lecture-discussion; they tended to be dissatisfied with both methods of independent study. The investigator concluded that since all teaching methods led to similar substantial gains in course achievement, serious doubt was cast on definitions of time needed in the classroom. She suggested that independent-study methods appeared to be as useful as lecture-discussion, provided that independent study can be made more acceptable to students.

Hovey (42) conducted a study with two classes in educational psychology equated with respect to aptitude. One class was taught by the traditional lecture method, three lectures per week, while the other was handled by a self-directed technique involving a considerable amount of small-group discussion. There was a small but non-significant difference in favor of the self-directed group in the mastery of course material. Ten months after the course was completed, the subjects were reassembled for post-tests covering course material and measures of curiosity. The self-directed subjects were still slightly superior in retention of course
material. On thirteen of fifteen items measuring curiosity, the self-directed subjects were superior. The investigator generalized that the technique employed in the experiment of placing a major responsibility on the student for his own education suggests interesting possibilities for developing attitudes toward learning which will result in the student's continuing a search for knowledge after the formal classroom experience is over.

Studies of Composites of Procedures

A doctoral dissertation by Nesbitt (65) analyzed the evidence during the final year of an experiment on staff utilization at the secondary level. Its purpose was to test the hypothesis that teaching the same students in large groups of 70 to 100 part of the time, and in small groups of 10 to 15 and individually the remainder of the time, is more productive educationally than methods usually associated with teaching classes of 25 or 30. The data were collected by both standardized and teacher-made tests in biology, general science, and language arts. Pretests and post-tests were used. The findings from the statistical treatment of the data indicated that each class did achieve significant growth; that with the exception of language arts, individuals within each class tended to become significantly more variable as the experiment progressed; that there was a significant difference between the two methods of instruction in biology
and general science but not in language arts. No significant differences were noted in the retention scores.

Becker (6) coordinated an experiment in college American Government using a composite of procedures involving television discussion, television observation, small-group discussion, large-group discussion and lecture. Sections using each of the methods as the primary instructional medium were organized. Evaluative criteria were: achievement as measured by two mid-term and one final examination; attitudes toward concepts of liberal democracy; opinions concerning the various methods; and attitudes of the discussion instructors toward these methods of instruction. The three discussion methods were also compared on the basis of class participation. The hypothesis tested was that the five methods of instruction would not differ significantly on any of the four criteria.

The investigator reported that the hypothesis of no difference between these methods of instruction in their effect on acquisition of knowledge and attitude change was retained. No significant differences were found on the amount of participation in the three discussion methods. The only significant differences found were in the area of student opinion. It was found that students prefer a course taught at least partially by the discussion method.

A study of a composite of procedures in a block-teaching project on the college level has been reported by
Kreisman (47). The project brought together 100 students and seven faculty members, covering six areas of subject matter for a total of 12 credit hours per quarter for three quarters. The method was used to determine whether or not the subject matter could be taught as efficiently in 12 sections as under the traditional system which took 17 sections per 100 students. Test results indicated that the two groups (block and control) were comparable on the basis of covariance analysis of the achievement test scores of both groups. The investigator concluded that the use of the block method did not result in educational loss since results showed no significant difference between these two groups on achievement tests.

Anthony and his colleagues (1) studied composite procedures at Boston University Junior College in the general education program. The first report was a description of the core curriculum involving the areas of humanities, social relations, science, guidance, and communications. A later evaluation by Glanz (34) subjected the process to a more rigorous analysis. The system used provided for informal student-teacher relationships. The faculty was divided into units, each unit comprised of five instructors drawn from the five major divisions of the college. Each student was assigned to a section of twenty to twenty-seven students, and each team of instructors had the responsibility of teaching several of these sections. In the reported study,
scholastic growth was evaluated in terms of the change in score on the Graduate Record Examinations Area Achievement Tests. Parallel forms were used as pretest and post-test. The pretest was given at the beginning of the freshman year and the post-test was given at the end of the sophomore year. Results indicated that as first semester freshmen, the members of the class tested were below the national freshman average, but that as second semester sophomores, they were above the national sophomore average. The investigators emphasized that they considered the data of the study to be evidence for, not proof of, the value of the team-and-composite system in a college setting.

Lienemann (51) organized two studies in 1966 and 1967 to determine empirically if, by assigning students to a block of classes with a cross-discipline team of teachers, significant differences could be found in interaction and academic achievement. Each study was conducted over a span of one sixteen-week trimester. In both studies, the two groups were compared on the basis of American College Test composite scores. Although the control group was slightly higher in 1966 and lower in 1967, the two groups were judged to be comparable on this variable. Objective measuring devices were used on both the experimental and control groups. Out-of-class discussion was measured in both experiments by an interaction questionnaire. The Graduate Record Examinations Area Tests in the humanities and in social science were used
to measure academic achievement in 1966. Standardized tests were used to measure achievement in three other specific course areas. Mean scores for both groups on each device were determined and the difference between means was tested for significance. The hypothesis that the experimental group would have greater student-student and student-faculty out-of-class discussion was not accepted. On the achievement measures, the experimental group scored higher on every measure in both studies and was significantly higher in the social science area on both occasions. Therefore, the hypothesis that the experimental group would have greater academic achievement than the traditionally taught students was accepted in both studies although not all results were significant.

Summary

The concern with college instruction has been reflected in a growing body of literature and studies (21, 24). The proposal that universal post-secondary education be made available to all is now taken as a serious possibility (46).

A trend toward innovative practices in instruction characterizes much of the activity now reported by colleges and universities (4, 57). Better utilization of staff and facilities is a factor receiving much attention. These developments are viewed as being symptomatic of more basic changes in the culture and society at the present time.
The basic dilemma is one of maintaining quality in a setting of increased numbers of extension of physical and financial resources (40, 53, 67). Collations of essays of authorities in the field continue to relate to the need for research and the initiation of changes by means of sophisticated techniques (49, 72, 86).

The role of the junior college in recent developments in post-secondary education has been explored by distinguished commissions and specialists in the area (8, 41, 59, 67, 68). The teaching obligation in these institutions, while not having been the subject of intensive research, has been recognized as central to their reasons for existing (13, 22, 83). Many studies in this area at the present time are status surveys (44, 45, 84).

For the purposes of the present study, the literature on instruction has been limited primarily to studies of college instruction as such. The components whose patterns of combination distinguish the experimental and control treatments have been designated as lecture-versus-discussion, class size, and class contact time.

The literature on lecture-versus-discussion has been at variance in reported findings. Studies have reported findings indicating that lecture and discussion methods appear equally effective in immediate recall of material (2). The lecture method has been considered superior in producing improvement in things measured by tests (77). Comparisons
between reading and lecture produced no significant differences (36). Other results indicated that immediate recall is better for read materials with no differences in delayed recall (16).

Consideration of possible flaws in research design on lecture-versus-discussion methods have called attention to diffusion and lack of uniformity in procedures (15, 70).

Thought processes and differentiation of results in the two methods have received more recent attention (32). Conclusions resulting from controlled experiments have indicated that the lecture is more efficient in communicating knowledge about a topic or field, but that in the development of abilities and skills which are problem-solving in nature, discussion is superior (9). The ratio of lecture to discussion time has been reported as favoring a minimum lecture time for specific purposes (85). Multivariate designs of several combinations of lecture-discussion approaches resulted in conclusions that no differences between the teaching methods from the point of view of educational outcome could be reported (39). Morale of students has been reported better in lecture classes in some situations (25). In other situations, non-academic factors were reported as being superior for discussion classes with no significant differences in other regards.

Studies on the psychological and sociological dimensions of lecture-versus-discussion indicated gains in non-academic
variables in favor of discussion as opposed to lecture (71, 76), but with definitive group leadership being advantageous (55). Whether the cohesiveness was the result of strong instructor leadership or group leadership did not make for significant differences in other studies (12).

Attitudes have been reported as being better and participation more frequent for discussion methods as opposed to lecture (26, 52, 66), but the same lack of significant differences in academic achievement persisted (3, 19, 50, 56) with one exception (48). The author of a collation of recent studies (31) on lecture versus discussion concluded that techniques which are good for one subject or for one group of students may be quite unsatisfactory for another.

Studies on class size are also at variance in reported findings. Early studies indicated no differences in achievement for large or small classes (23). A general trend of findings supported the position that large classes at least did not lower academic performances (43, 60). Non-academic factors have generally been reported as being assisted by small classes, however (33, 78, 79). Size of class and group has been found to be related to participation in specific ratios in non-scholastic settings. In academic settings, results have tended to suggest that differences between small and large classes have been more a function of instructor personality than the variable of size (69). Research design in this area has also been criticized for
not testing in the immediate situation (61). In other studies concerned with immediate and long-term retention, however, no significant differences were found in large class and small class situations (20, 54, 74, 75). Student satisfaction with small classes was generally agreed upon, however (29, 73). The conclusion has been offered that a mass-lecture method of teaching can be employed without significantly diminishing the amount of learning (37).

With the continued report of no significant differences attributable to class size (64), attention has been directed to the teaching approaches within the variations of class size as being the key to differences in outcome (82). A contemporary evaluation of the problem suggests that the issue is no longer one of class size but one of how the whole range of possibilities can be used for the best instructional combinations (27, 35).

Studies on reduced class contact time have been largely narrative and exploratory (7, 10, 28). There has been some reported confusion as to definition of the factors which would be considered as independent study or reduced class contact time. The general trend was to consider several approaches as being appropriate to inclusion in the category of independent study practices (18, 80).

Experimental studies in the area of reduced class contact time have not produced the favorable results which proponents of the method had hoped for (18). The experiment
at Antioch College resulted in the conclusion that since all teaching methods led to similar substantial gains in course achievement, serious doubt was cast on definitions of time needed in the classroom. However, independent study was not looked upon with favor by the student involved in it (14). Other studies also dealt with the cognitive versus non-cognitive advantages and disadvantages of reduced class time with conflicting findings (17, 42, 58).

Studies of composites of procedures with experimental controls are rare in the literature. One related study on the secondary level (65) reported differences between science subjects but not in language arts when a composite of lecture, discussion, and small and large classes was used. All classes, however, had significant growth regardless of method. On the college level, an extensive experiment with a composite of five methods of instruction resulted in the finding of no significant differences except in the area of student opinion which favored a discussion or dialogue method (6). Other findings supported the position that, since a composite method cannot be shown to hinder academic achievement, its use is justified on the basis of administrative efficiency (47). Results of studies of composite procedures on the junior college level so far, however, have tended to favor the procedures in terms of academic achievement rather than non-academic gains (1, 34, 51).
CHAPTER BIBLIOGRAPHY


2. Bane, C. L., "The Lecture vs. the Class Discussion Method of Teaching," School and Safety, XXI (March, 1925), 300-302.


42. Hovey, Donald E., Howard E. Gruber, and Glenn Terrell, "Effects of Self-Directed Study on Course Achievement, Retention, and Curiosity," Journal of Educational Research, LVI (March, 1963), 346-351.


45. New Directions for Instruction in the Junior College, Los Angeles, The University of California, 1965.


84. Umstattd, James G., Teaching Procedure Used in Twenty-eight Midwestern and Southwestern Colleges and Universities, Austin, University Cooperative Society, 1954.


CHAPTER III

COLLECTION AND TREATMENT OF DATA

Subjects

The subjects were students enrolled in first-semester freshman English composition and American history courses in El Centro Junior College of the Dallas County Junior College District during the fall semester of the 1968-69 school year. The institution is a public open-door junior college committed to assignment of students to sections in the order of receipt of requests. This was the population base of the experimental and control groups. The availability of the Special Program schedule was made known, and students requesting sections in it were assigned in order of request. From the point of view of students, the experimental program was a scheduling option available to them rather than an experiment.

Students who had elected to take concurrently both their English 101 and History 101 courses in the Special Program schedule pattern comprised the experimental group. In short, the students in the Special Program English and the students in the Special Program history were the same students.

Five regular-schedule sections of English 101 chosen randomly outside the jurisdiction of the investigator comprised the English control group. In like manner, five
regular-schedule sections of History 101 chosen randomly comprised the history control group.

Data were collected within the first fourteen class days and the final fourteen class days of the semester. The American College Testing Program battery was administered to entering students as a part of the enrollment procedure of the college.

Students who had failed the equivalent of English 101 or History 101, students over thirty years of age, and students concurrently enrolled in the Audio-Tutorial Biology courses of the college were not included in the analyses.

The instruments were administered during regular class meetings of the groups. Because of this, the number of subjects with complete data varied with each criterion instrument since there were the factors of attrition and absences during the pretest and post-test situations, as well as the lack of availability of ACT scores for some students. The number of usable subjects for each criterion measure is reported in the presentation of the data in Chapter IV.

Instruments

Alternate forms of four instruments were administered before and after the experimental treatment.

Forms E and F of the 1965 revision of the Crary American History Test were administered to the experimental group and the five sections of control history classes. The test was
constructed and revised in accordance with standards of the American Educational Research Association and National Council of Measurements Used in Education (10). The item-tryout and standardization were carried out in a dual research program conducted in the spring of 1964 in 16 geographically balanced regions involving 9,523 subjects. In a general sense, the test was constructed to provide a representative coverage of the important outcomes of American history instruction as attested to by a consensus of current practices (4). Both forms provide proportionate samplings of content within each basic aspect of American history and are approximately equivalent to each other in terms of content and statistical balance (4, p. 4). Odd-even split-half correlation coefficients corrected by the Spearman-Brown Formula are reported for the college preparatory normative group at .90 for Form E and .88 for Form F.

Forms 1A and 1B of the Cooperative English Expression Test were administered to the experimental group and the five sections of control English classes. Adequate content validity is reported for the test (5). The first section takes fifteen minutes to administer and has thirty items for measuring effectiveness, defined as the choice of the written expression which precisely conveys the meaning intended. The section on mechanics allows twenty-five minutes for sixty items designed to measure competences in usage, spelling, punctuation, and capitalization. The student is given a
series of three-line sentences and asked to identify in each sentence the line in which any mechanics error occurs, or to indicate that the sentence contains no such error.

Reviews are unanimously favorable (3) and suggest that, although no objective test measures the ability to write, this instrument tests certain critical powers which are related to this ability by coming near to the actual writing of themes, and stimulating one part of the writing process of proofreading and revision. Reliabilities based on correlation between the parallel forms used are reported at .84 each (13).

Forms YM and ZM of the Watson-Glaser Critical Thinking Appraisal were administered to the experimental group as well as to the five sections of control history and the five sections of control English classes. This instrument has five subtests designed to measure different though interdependent aspects of critical thinking. The Inference section of twenty items samples ability to discriminate among degrees of truth or falsity of inferences drawn from given data. The Recognition of Assumptions section of sixteen items samples ability to recognize unstated assumptions or presuppositions which are taken for granted in given statements or assertions. The Deduction section of twenty-five items samples ability to reason deductively from given statements or premises. The Interpretation section of twenty-four items samples ability to weigh evidence and to distinguish between various
types of generalizations. The Evaluation of Arguments section of fifteen items samples ability to distinguish between arguments which are strong and relevant and those which are weak or irrelevant to a particular question at issue (7, p. 2).

The 1963 revision represents the culmination of over twenty-five years of experimentation on the measurement of critical thinking abilities. Administration to representative samples of well-defined national populations of high school students and liberal arts college freshmen was a component of the standardization program. The liberal arts freshman population numbered 5,297. Odd-even split-half reliability coefficients corrected by the Spearman-Brown formula for this population for Form YM was .85. It is not reported for Form ZM. Correlations of .56 to .79 between the various subtests and the total test is cited as evidence of the substantial relationship of the various subtests to the instrument as a whole (7).

Forms A and B of the Purdue Master Attitude Scale, Part A, Attitude Toward Any Subject, were administered to the experimental group as well as to the five sections of control history and the five sections of control English classes. The publishers of the scale point out that the scaling procedure is the psycho-physical principle that equally often-observed differences are equal, often referred to as the Thurstone attitude scaling technique. This principle is
considered applicable to the scaling of attitude statements relevant to specific attitudes (9, 11). Beyond their face validity, these scales are reported to have demonstrated validity both against Thurstone's specific scales with which they show strong correlations and in differentiating among attitudes known to differ among various groups (9). Both high school and college subjects have been involved in a considerable variety of situations (9). The correlations of Form A with B have ranged from .81 to .90. Reliabilities from .84 to .86 have been reported with high school subjects and from .81 to .90 for college students (12).

The scale was revised in 1960 and reduced to seventeen items. With the exception of this scale, reviews are largely negative. Campbell, however, refers to generalized scales such as this one as being essential to the field of testing (2, p. 46; 3, p. 359) and Clark indicates that the test is adequate (2, p. 46).

**Treatment of Data**

All hypotheses were restated in null form for testing with the equivalent of a one-way (single classification) analysis of variance design for two groups (8). Levels of significance were identified. A multiple-linear regression computer program (1) was used with a Model 40 IBM 360 computer in the Dallas County Junior College District Data Processing Center.
The following ordered vectors, each with \( n \) elements, were constructed to test the first hypothesis:

- \( Y \) = criterion vector.
- \( X^{(1)} \) = a vector with 1 in elements for which the corresponding element in \( Y \) is for a Special Program student, 0 otherwise.
- \( X^{(2)} \) = a vector with 1 in elements for which the corresponding element in \( Y \) is for a student in the control group, 0 otherwise.

The following full model was generated, with least squares solutions for the vector coefficients, using the multiple linear regression computer program.

\[
Y = cU + a_1 X^{(1)} + a_2 X^{(2)}
\]

The vector \( U \) was the unit vector which contained 1 in each element. It was added mathematically by a computing procedure based on a correlation matrix. This model resulted in an \( R^2_f \) value where \( f \) designated the full model \( R^2 \) and had two linearly independent predictor vectors.

Hypothesis I, stated in the null form, placed the following restrictions on the full model.

\[
a_1 = a_2 = b.
\]

Substituting in the full model and algebraically simplifying resulted in the following restricted model:
\[ Y = c_1 U + bX^{(1)} + bX^{(2)} \]

\[ Y = c_1 U + b(X^{(1)} + X^{(2)}) \]

\[ Y = c_1 U \]

since \( X^{(1)} + X^{(2)} = U \).

This model, when solved, resulted in an \( R^2 = 0.0 \) (r for restricted model). An F ratio for testing the null hypothesis \( a_1 = a_2 = b \) was then computed

\[ F = \frac{R^2_f / dfn}{(1 - R^2_f) / dfd} \]

where the degree of freedom for the numerator (dfn) equals unity and the degrees of freedom for the denominator (dfd) equal n-2.

To test Hypothesis II, the following additional vectors were constructed.

\( X^{(3)} = \) a vector with 1 in elements for which the corresponding element in Y is for a male, 0 otherwise.

\( X^{(4)} = \) a vector with 1 in elements for which the corresponding element in Y is for a female, 0 otherwise.

\( Z^{(1)} = \) the product of vectors \( X^{(1)} \) and \( X^{(3)} \) which has 1 in the elements for which the corresponding elements in Y are Special Program males, 0 otherwise.
\( Z^{(2)} = \) the product of vectors \( X^{(1)} \) and \( X^{(4)} \) which has 1 in the elements for which the corresponding elements in \( Y \) are Special Program females, 0 otherwise.

\( Z^{(3)} = \) the product of vectors \( X^{(2)} \) and \( X^{(3)} \) which has 1 in the elements for which the corresponding elements in \( Y \) are control group males, 0 otherwise.

\( Z^{(4)} = \) the product of vectors \( X^{(2)} \) and \( X^{(4)} \) which has 1 in the elements for which the corresponding elements in \( Y \) are control group females, 0 otherwise.

The appropriate full model was

\[ Y = c_2 U + d_1 Z^{(1)} + d_2 Z^{(2)} + d_3 Z^{(3)} + d_4 Z^{(4)} \]

which had four linearly independent predictors. This classified the subjects into the following 2 by 2 set of categories.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special Program Students</td>
<td>( Z^{(1)} )</td>
<td>( Z^{(2)} )</td>
</tr>
<tr>
<td>Control Group Students</td>
<td>( Z^{(3)} )</td>
<td>( Z^{(4)} )</td>
</tr>
</tbody>
</table>

Equal cell frequencies were not expected nor necessary.

In null form, Hypothesis II placed the following restrictions on the full model:

\[ d_1 = d_3 = b_2 \text{ and } d_2 = d_4 = b_3 \]

Substituting and doing the algebra resulted in
\[ Y = c_2 U + b_2 z^{(1)} + b_3 z^{(2)} + b_4 z^{(3)} + b_5 z^{(4)} \]

\[ Y = c_2 U + b_2 (z^{(1)} + z^{(3)}) + b_3 (z^{(2)} + z^{(4)}) \]

\[ Y = c_2 U + b_2 x^{(3)} + b_3 x^{(4)} \]

since \( z^{(1)} + z^{(3)} = x^{(3)} \) and \( z^{(2)} + z^{(4)} = x^{(4)} \).

This restricted model had two linearly independent predictors.

Using the squared multiple correlation coefficients resulting from the solution for the vector coefficients in the full and restricted models, the null hypothesis was tested by the \( F \) ratio

\[
F = \frac{(R_f^2 - R_r^2) / df_n}{(1 - R_f^2) / df_d}
\]

with \( df_n = 2 \) and \( df_d = n - 4 \).

To test Hypotheses III, IV, V, VI, VII, and VIII, the following additional vectors were generated for the control variables being considered.

\( x^{(5)} \) = a vector containing the scores on the control variable for each subject.

\( z^{(5)} \) = the product of vectors \( z^{(1)} \) and \( x^{(5)} \) which contained the control variable scores in those elements for which the corresponding element in \( Y \) is for a male Special Program student, 0 otherwise.

\( z^{(6)} \) = the product of vectors \( z^{(2)} \) and \( x^{(5)} \), which contained the control variable scores in those elements for which the corresponding
element in $Y$ is for a female Special Program student, 0 otherwise.

$Z^{(7)} = \text{the product of vectors } Z^{(3)} \text{ and } X^{(5)}, \text{ which contained the control variable scores in those elements for which the corresponding element in } Y \text{ is for a male control group student, 0 otherwise.}$

$Z^{(8)} = \text{the product of vectors } Z^{(4)} \text{ and } X^{(5)}, \text{ which contained the control variable scores in those elements for which the corresponding element in } Y \text{ is for a female control group student, 0 otherwise.}$

The appropriate full model was then

$$Y = c_1 U + d_1 Z^{(1)} + d_2 Z^{(2)} + d_3 Z^{(3)} + d_4 Z^{(4)} + d_5 Z^{(5)} + d_6 Z^{(6)} + d_7 Z^{(7)} + d_8 Z^{(8)}$$

which had 8 linearly independent predictor vectors.

Before the question under consideration could be asked, it was necessary to determine whether the relationship between the control variable and the criterion variable could be considered to be the same (not statistically significantly different) in the four subgroups (Special Program males, control males, Special Program females and control females). Graphically stated, if four regression lines were computed, would they be parallel?
This restriction was stated as
\[ d_5 = d_6 = d_7 = d_8 = b_5 \]
which through substitution yielded the restricted model
\[ Y = c_{4, U} + d_1 z_1 + d_2 z_2 + d_3 z_3 + d_4 z_4 + b_5 x_5 \]
with 5 linearly independent predictors. The formula
\[ F = \frac{(R_f^2 - R_r^2)}{(1 - R_f^2) / dfd} \]
with \( dfn = 3 \) and \( dfd = 8 \) provided a test of the null hypothesis. This made provision for instances in which the null hypothesis was not rejected (the relationship between criterion and control variable could be considered the same) in order that the question of interest could be asked directly. As before, this was stated as
\[ d_1 = d_3 = b_2 \] and \( d_2 = d_4 = b_3 \)
which yielded the following restricted model
\[ Y = c_{4, U} + b_2 x_3 + b_3 x_4 + b_5 x_5 \]
with 3 linearly independent predictor vectors, which was compared with the previous model
\[ Y = c_{4, U} + d_1 z_1 + d_2 z_2 + d_3 z_3 + d_4 z_4 + b_5 x_5 \]
which was considered as the full model
\[ F = \frac{(R_f^2 - R_r^2)}{dfn} \]
\[ \frac{(1 - R_f^2)}{dfd} \]
with \( dfn = 2 \) and \( dfd = n - 5 \).
CHAPTER BIBLIOGRAPHY


CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

Statistical results have been organized in tabular form with both the 1 and 5 per cent levels of confidence identified. The tables are the bases of the interpretations.

An examination of Table I indicates the number of usable subjects with each criterion measure. The total number of subjects varied with each instrument since administration took place in regular class meetings, and an ACT score was not available for some students.

**TABLE I**

TOTAL AND CATEGORICAL SAMPLE SIZES FOR CRITERIA INSTRUMENTS

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Crary American History Test</td>
<td>178</td>
</tr>
<tr>
<td>Cooperative English Expression Test</td>
<td>159</td>
</tr>
<tr>
<td>Watson-Glaser Critical Thinking Appraisal English</td>
<td>160</td>
</tr>
<tr>
<td>Watson-Glaser Critical Thinking Appraisal History</td>
<td>190</td>
</tr>
<tr>
<td>Purdue Attitude Scale English</td>
<td>156</td>
</tr>
<tr>
<td>Purdue Attitude Scale History</td>
<td>170</td>
</tr>
</tbody>
</table>

*E, Experimental; C, Control.
It should be noted that although the Watson-Glaser Critical Thinking Appraisal is listed twice in Table I, as well as in subsequent tables, the experimental group numbers represent a single administration situation. In contrast, the Purdue Attitude Scale actually represents separate experimental group administrations, one for English attitude and one for history attitude.

Table II is provided to serve as a basic reference for subsequent analyses.

**TABLE II**

DEGREES OF FREEDOM FOR F-RATIO FOR CRITERIA INSTRUMENTS AND HYPOTHESES

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Hypotheses</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
<td>Part A</td>
</tr>
<tr>
<td>Crary American History Test</td>
<td>N*</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>D*</td>
<td>176</td>
<td>174</td>
<td>170</td>
</tr>
<tr>
<td>Cooperative English Expression Test</td>
<td>N</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>157</td>
<td>155</td>
<td>151</td>
</tr>
<tr>
<td>Watson-Glaser Critical Thinking Appraisal--English</td>
<td>N</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>158</td>
<td>156</td>
<td>152</td>
</tr>
<tr>
<td>Watson-Glaser Critical Thinking Appraisal--History</td>
<td>N</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>188</td>
<td>186</td>
<td>182</td>
</tr>
<tr>
<td>Purdue Attitude Scale--English</td>
<td>N</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>154</td>
<td>152</td>
<td>148</td>
</tr>
<tr>
<td>Purdue Attitude Scale--History</td>
<td>N</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>168</td>
<td>166</td>
<td>162</td>
</tr>
</tbody>
</table>

*N, numerator; D, denominator.
The dfn figure is determined by subtracting the number of independent predictors in the restricted model from the number in the full model. The dfd figure represents the number of full-model independent predictors subtracted from the criterion number of cases.

In Tables III through XVIII, the criteria instruments and their subtests are designated by Arabic numerals 1 through 18 as shown in Table III.

### TABLE III

#### NUMBER DESIGNATION OF CRITERIA INSTRUMENTS

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Crary American History Test</td>
</tr>
<tr>
<td>2</td>
<td>Cooperative English Expression Test: Effectiveness</td>
</tr>
<tr>
<td>3</td>
<td>Cooperative English Expression Test: Mechanics</td>
</tr>
<tr>
<td>4</td>
<td>Cooperative English Expression Test: Total</td>
</tr>
<tr>
<td>5</td>
<td>Watson-Glaser Critical Thinking Appraisal: Inference--English groups</td>
</tr>
<tr>
<td>6</td>
<td>Watson-Glaser Critical Thinking Appraisal: Recognition of Assumptions--English groups</td>
</tr>
<tr>
<td>7</td>
<td>Watson-Glaser Critical Thinking Appraisal: Deduction--English groups</td>
</tr>
<tr>
<td>8</td>
<td>Watson-Glaser Critical Thinking Appraisal: Interpretation--English groups</td>
</tr>
<tr>
<td>9</td>
<td>Watson-Glaser Critical Thinking Appraisal: Evaluation of Arguments--English groups</td>
</tr>
<tr>
<td>10</td>
<td>Watson-Glaser Critical Thinking Appraisal: Total--English groups</td>
</tr>
<tr>
<td>11</td>
<td>Watson-Glaser Critical Thinking Appraisal: Inference--History groups</td>
</tr>
<tr>
<td>12</td>
<td>Watson-Glaser Critical Thinking Appraisal: Recognition of Assumptions--History groups</td>
</tr>
<tr>
<td>13</td>
<td>Watson-Glaser Critical Thinking Appraisal: Deduction--History groups</td>
</tr>
<tr>
<td>14</td>
<td>Watson-Glaser Critical Thinking Appraisal: Interpretation--History groups</td>
</tr>
<tr>
<td>15</td>
<td>Watson-Glaser Critical Thinking Appraisal: Evaluation of Arguments--History groups</td>
</tr>
<tr>
<td>16</td>
<td>Watson-Glaser Critical Thinking Appraisal: Total--History groups</td>
</tr>
<tr>
<td>17</td>
<td>Purdue Attitude Scale--English groups</td>
</tr>
<tr>
<td>18</td>
<td>Purdue Attitude Scale--History groups</td>
</tr>
</tbody>
</table>
Table IV is a compilation of the mean gains, standard deviations, and \( t \)-scores for the criteria instruments for all subjects. The \( t \)-score data for Table IV are based on a derivation of the \( t \)-ratio formula for testing the significance of differences between correlated pairs of means

\[
\frac{t}{=\frac{M}{\sqrt{\frac{1}{\sigma^2}}}} \text{ \( \frac{N}{N-1} \)}
\]

where \( df = N - 1 \).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Mean Gain</th>
<th>Standard Deviation</th>
<th>( t )</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.83</td>
<td>5.73</td>
<td>8.90</td>
<td>.01</td>
</tr>
<tr>
<td>2</td>
<td>1.06</td>
<td>3.73</td>
<td>3.57</td>
<td>.01</td>
</tr>
<tr>
<td>3</td>
<td>-1.68</td>
<td>5.70</td>
<td>-3.71</td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>-.53</td>
<td>7.04</td>
<td>-.95</td>
<td>N.S.</td>
</tr>
<tr>
<td>5</td>
<td>-.13</td>
<td>4.67</td>
<td>-.35</td>
<td>N.S.</td>
</tr>
<tr>
<td>6</td>
<td>1.61</td>
<td>4.74</td>
<td>4.29</td>
<td>.01</td>
</tr>
<tr>
<td>7</td>
<td>.95</td>
<td>5.71</td>
<td>2.12</td>
<td>.05</td>
</tr>
<tr>
<td>8</td>
<td>.40</td>
<td>5.66</td>
<td>.89</td>
<td>N.S.</td>
</tr>
<tr>
<td>9</td>
<td>.98</td>
<td>3.53</td>
<td>3.50</td>
<td>.01</td>
</tr>
<tr>
<td>10</td>
<td>3.71</td>
<td>19.48</td>
<td>2.40</td>
<td>.05</td>
</tr>
<tr>
<td>11</td>
<td>-.34</td>
<td>4.47</td>
<td>-1.05</td>
<td>N.S.</td>
</tr>
<tr>
<td>12</td>
<td>1.61</td>
<td>4.45</td>
<td>4.97</td>
<td>.01</td>
</tr>
<tr>
<td>13</td>
<td>.64</td>
<td>5.42</td>
<td>1.62</td>
<td>N.S.</td>
</tr>
<tr>
<td>14</td>
<td>-.30</td>
<td>5.58</td>
<td>-.74</td>
<td>N.S.</td>
</tr>
<tr>
<td>15</td>
<td>1.02</td>
<td>3.30</td>
<td>4.25</td>
<td>.01</td>
</tr>
<tr>
<td>16</td>
<td>2.58</td>
<td>18.43</td>
<td>1.92</td>
<td>.05</td>
</tr>
<tr>
<td>17</td>
<td>-.01</td>
<td>1.38</td>
<td>.08</td>
<td>N.S.</td>
</tr>
<tr>
<td>18</td>
<td>.38</td>
<td>1.36</td>
<td>3.63</td>
<td>.01</td>
</tr>
</tbody>
</table>

It may be seen that in several instances the obtained \( t \)-scores surpassed the required ratios of 1.66 for the .05
level of confidence and 2.61 for the .01 level. It is interesting to note that on the Cooperative English Expression Test: Mechanics (criteria 3), there was a loss significant at the .01 level while there was a gain significant at the .01 level for the same instrument's subtest of effectiveness. Total score gains for the Crary American History Test, the Watson-Glaser Critical Thinking Appraisal, and the Purdue Attitude Scale for history were significant. This is not the case for the total score gains of the Cooperative English Expression Test, some subtests of the Watson-Glaser Critical Thinking Appraisal, and the Purdue Attitude Scale for English.

It may be noted that the values of the standard deviations as related to the reported mean gains in several instances are such as to suggest a platykurtic trend; however, it should be recalled that the actual gain differences on the criteria measures ranged from negative values to positive values rather than from zero to positive only.

The First Hypothesis

The first research hypothesis was stated as follows: The mean gain made by the experimental group will be significantly greater than the mean gain made by the control group on each of the following measures:

a. The Crary American History Test
b. The Cooperative English Expression Test
c. The Watson-Glaser Critical Thinking Appraisal
d. The Purdue Attitude Scale, Part A, Attitude Toward Any Subject.

Restated in null form, the statistical hypothesis was that the mean gain made by the experimental group on each of the four measures will not be significantly greater than the mean gain made by the control group. The findings of Hypothesis I are presented in Table V.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Mean Gain Experimental</th>
<th>Mean Gain Control</th>
<th>F-Value</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.10</td>
<td>4.80</td>
<td>3.87</td>
<td>N.S.</td>
</tr>
<tr>
<td>2</td>
<td>1.10</td>
<td>1.00</td>
<td>0.00</td>
<td>N.S.</td>
</tr>
<tr>
<td>3</td>
<td>-1.32</td>
<td>-2.18</td>
<td>0.85</td>
<td>N.S.</td>
</tr>
<tr>
<td>4</td>
<td>-0.22</td>
<td>-0.98</td>
<td>0.47</td>
<td>N.S.</td>
</tr>
<tr>
<td>5</td>
<td>0.22</td>
<td>-0.72</td>
<td>1.49</td>
<td>N.S.</td>
</tr>
<tr>
<td>6</td>
<td>2.29</td>
<td>0.41</td>
<td>5.96</td>
<td>.05</td>
</tr>
<tr>
<td>7</td>
<td>2.10</td>
<td>-1.07</td>
<td>12.09</td>
<td>.01</td>
</tr>
<tr>
<td>8</td>
<td>1.32</td>
<td>-1.22</td>
<td>7.76</td>
<td>.01</td>
</tr>
<tr>
<td>9</td>
<td>1.59</td>
<td>-0.09</td>
<td>8.65</td>
<td>.01</td>
</tr>
<tr>
<td>10</td>
<td>7.43</td>
<td>-2.83</td>
<td>10.82</td>
<td>.01</td>
</tr>
<tr>
<td>11</td>
<td>0.22</td>
<td>-0.99</td>
<td>3.46</td>
<td>N.S.</td>
</tr>
<tr>
<td>12</td>
<td>2.29</td>
<td>0.81</td>
<td>5.38</td>
<td>.05</td>
</tr>
<tr>
<td>13</td>
<td>2.10</td>
<td>-1.05</td>
<td>17.14</td>
<td>.01</td>
</tr>
<tr>
<td>14</td>
<td>1.32</td>
<td>-2.18</td>
<td>20.47</td>
<td>.01</td>
</tr>
<tr>
<td>15</td>
<td>1.59</td>
<td>0.36</td>
<td>6.69</td>
<td>.05</td>
</tr>
<tr>
<td>16</td>
<td>7.43</td>
<td>-3.05</td>
<td>16.43</td>
<td>.01</td>
</tr>
<tr>
<td>17</td>
<td>-0.16</td>
<td>0.18</td>
<td>2.23</td>
<td>N.S.</td>
</tr>
<tr>
<td>18</td>
<td>0.52</td>
<td>0.07</td>
<td>8.15</td>
<td>.01</td>
</tr>
</tbody>
</table>

Examination of Table V reveals an F-value of 10.82 and 16.43 for the Watson-Glaser Critical Thinking Appraisal in the English and history comparison situations respectively.
The .01 level of confidence required F-values of 6.80 and 6.77 respectively. Furthermore, the findings on certain subtests of this instrument (criteria 6, 7, 8, 9, 12, 13, 14 and 15) also were significant. Differences on one of the subtests (criterion 11), while not reaching the F-value of 3.91 required for the .05 level of confidence, did indicate a trend in the direction of significance. It may also be seen that the F-value of 3.87 for the Crary American History Test (criterion 1) approached the .05 level of confidence F-value of 3.90 in favor of the control group. The 3.91 F-value for the .05 level of confidence for the Cooperative English Expression Test was not approximated.

It may also be seen that a significant difference favoring the experimental group at the .01 level was found for the Purdue Attitude Scale for history. By contrast, the trend related to the Purdue Attitude Scale for English is negative for the experimental group.

In some instances (criteria 7, 8, 9, 10, 13, 14, 16), the significance of the gains for the experimental group were related to the losses experienced by the control group.

In instances where the findings indicated no significant differences, the trend was in favor of the experimental group five times and the control group two times. On two criteria there were negative trends for both the experimental and control groups. These were the Cooperative English Expression Test: Mechanics and Total.
The null hypothesis of no significant differences could not be rejected, and the research hypothesis was rejected as related to the Crary American History Test, the Cooperative English Expression Test, and the Purdue Attitude Scale for English. As related to the Watson-Glaser Critical Thinking Appraisal and the Purdue Attitude Scale for history, however, the null hypothesis was rejected. Therefore, the research hypothesis as it related to these criteria was accepted.

The Second Hypothesis

The second hypothesis was stated as follows: With sex held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group. Data for this hypothesis are reported in Table VI.

An examination of Table VI reveals that the F-values of 5.49 and 8.52 for the Watson-Glaser Critical Thinking Appraisal (criteria 10 and 16) were significant in both the English and history comparison situations since the .01 level of confidence required F-values of 4.75 and 4.72 respectively. Furthermore, the differences were significant on seven of the subtest situations of this instrument. The .05 confidence level F-value of 3.05 was approached on the instrument's subtest of deduction (criterion 12).
TABLE VI
MEAN GAINS AND F-VALUES FOR CATEGORICAL TREATMENT GROUPS

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Male</th>
<th>Female</th>
<th>Constant</th>
<th>F-value</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exper.</td>
<td>Control</td>
<td>Exper.</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.00</td>
<td>1.58</td>
<td>-.50</td>
<td>1.46</td>
<td>3.26</td>
</tr>
<tr>
<td>2</td>
<td>-.82</td>
<td>.00</td>
<td>-.18</td>
<td>-2.05</td>
<td>1.68</td>
</tr>
<tr>
<td>3</td>
<td>1.83</td>
<td>.00</td>
<td>.48</td>
<td>1.43</td>
<td>-2.66</td>
</tr>
<tr>
<td>4</td>
<td>.97</td>
<td>.00</td>
<td>.29</td>
<td>-.16</td>
<td>-.93</td>
</tr>
<tr>
<td>5</td>
<td>.98</td>
<td>.00</td>
<td>-.64</td>
<td>-1.48</td>
<td>-.19</td>
</tr>
<tr>
<td>6</td>
<td>.00</td>
<td>-2.54</td>
<td>-.45</td>
<td>-1.17</td>
<td>2.45</td>
</tr>
<tr>
<td>7</td>
<td>2.92</td>
<td>.00</td>
<td>1.66</td>
<td>-1.91</td>
<td>-.38</td>
</tr>
<tr>
<td>8</td>
<td>2.98</td>
<td>.00</td>
<td>1.96</td>
<td>.20</td>
<td>-1.29</td>
</tr>
<tr>
<td>9</td>
<td>.00</td>
<td>-1.81</td>
<td>-.01</td>
<td>-1.45</td>
<td>1.59</td>
</tr>
<tr>
<td>10</td>
<td>11.46</td>
<td>.04</td>
<td>6.84</td>
<td>-1.27</td>
<td>-2.40</td>
</tr>
<tr>
<td>11</td>
<td>1.45</td>
<td>.00</td>
<td>-.17</td>
<td>-.99</td>
<td>-.66</td>
</tr>
<tr>
<td>12</td>
<td>.00</td>
<td>-1.81</td>
<td>-.45</td>
<td>-1.32</td>
<td>2.45</td>
</tr>
<tr>
<td>13</td>
<td>3.33</td>
<td>.00</td>
<td>2.06</td>
<td>-.81</td>
<td>-.78</td>
</tr>
<tr>
<td>14</td>
<td>3.82</td>
<td>.00</td>
<td>2.80</td>
<td>-.14</td>
<td>-2.14</td>
</tr>
<tr>
<td>15</td>
<td>.00</td>
<td>-1.34</td>
<td>-.01</td>
<td>-1.00</td>
<td>1.59</td>
</tr>
<tr>
<td>16</td>
<td>11.76</td>
<td>.00</td>
<td>7.14</td>
<td>-1.06</td>
<td>-2.70</td>
</tr>
<tr>
<td>17</td>
<td>.00</td>
<td>.64</td>
<td>.11</td>
<td>-.05</td>
<td>-.20</td>
</tr>
<tr>
<td>18</td>
<td>.00</td>
<td>-.50</td>
<td>.13</td>
<td>-.61</td>
<td>.61</td>
</tr>
</tbody>
</table>

It may also be seen that the F-value in the case of the Purdue Attitude Scale for history was significant at the .01 level for the experimental group.

In the case of the Crary American History Test and the Cooperative English Expression Test, the .05 level of confidence would have required F-values of 3.05 and 3.06 respectively. The Purdue Attitude Scale for English would have required an F-value of 3.06 for the .05 level of confidence. The trend was in the direction of the control group.

In Table VI, and in subsequent tables, the comparison was between males in control and experimental groups and
between females in control and experimental groups. It was interesting to note the high incidence of negative trends in the female population in Table VI for both the experimental and control groups.

At this point it should be noted that in the case of the Purdue Attitude Scale for history, the statistical procedures used for the first and second hypotheses did not control for ability levels by using the ACT tests and the pretests of the criteria as control variables. Under these conditions, significant differences were found. In Hypotheses III through VIII it may be seen that there were non-significant differences relating to the Purdue Attitude Scale for history. This may be attributed to accounting for ability differences in the experimental and control groups with the use of the ACT and pretest scores as control variables.

Unlike the Purdue Attitude Scale for history, significant differences were not found in the case of the Purdue Attitude Scale for English in the analyses of the first and second hypotheses. Like the history situation, however, the application of statistical procedures using the ACT scores and pretest scores to account for ability differences in the English experimental and control groups resulted in findings of non-significant differences in the subsequent analyses involving the Purdue Attitude Scale as a criterion instrument. Trends, however, were in the direction of the English control group.
The null hypothesis of no significant differences could not be rejected as it related to the Crary American History Test, the Cooperative English Expression Test, and the Purdue Attitude Scale for English. Therefore, the research hypothesis in these instances was rejected. In the case of the Watson-Glaser Critical Thinking Appraisal and the Purdue Attitude Scale for history, the null hypothesis was rejected and the research hypothesis was accepted.

The Third Hypothesis

The third hypothesis stated that with sex and ACT English score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group. Table VII and Table VIII (see pages 103 and 104) present the data relevant to this hypothesis. These two tables represent two sequential phases of statistical procedures used in testing the hypothesis and should be examined consecutively.

Table VII is a compilation of results of the preliminary analyses which were necessary before proceeding to the main issue of the hypothesis. It includes data resulting from testing the null hypothesis that there were no significant differences in the expected change in the criterion per unit change in the control variable for the four subgroups (Special Program males, control males, Special Program
**TABLE VII**

CATEGORICAL VECTOR COEFFICIENTS, CATEGORICAL REGRESSION SLOPES OF ACT ENGLISH TEST, AND P-VALUE

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Categorical Vector Coefficients</th>
<th>Slope of Regression Act English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male Exper. Control Female Exper. Control</td>
<td>Male Exper. Control Female Exper. Control</td>
</tr>
<tr>
<td>1</td>
<td>.66 3.32 - .07 -1.31</td>
<td>.09 .04 .09 .28</td>
</tr>
<tr>
<td>2</td>
<td>-2.84 - .56 -4.70 - .34</td>
<td>.16 .09 .26 .05</td>
</tr>
<tr>
<td>3</td>
<td>4.79 - .01 8.67 4.28</td>
<td>-.18 -.12 -.42 -.18</td>
</tr>
<tr>
<td>4</td>
<td>1.57 - .71 3.56 4.03</td>
<td>-.03 .05 -.16 -.24</td>
</tr>
<tr>
<td>5</td>
<td>3.84 - .09 -4.54 -3.35</td>
<td>-.15 .01 .21 .13</td>
</tr>
<tr>
<td>6</td>
<td>.00 -7.40 -9.54 -9.41</td>
<td>-.33 -.05 .16 .15</td>
</tr>
<tr>
<td>7</td>
<td>6.23 1.63 6.47 - .69</td>
<td>-.19 -.11 -.24 -.08</td>
</tr>
<tr>
<td>8</td>
<td>9.42 - .02 -4.65 4.16</td>
<td>-.38 -.01 .32 -.24</td>
</tr>
<tr>
<td>9</td>
<td>-1.14 -2.20 2.78 -3.71</td>
<td>-.04 -.09 -.22 .03</td>
</tr>
<tr>
<td>10</td>
<td>34.73 7.64 8.53 4.06</td>
<td>-1.09 -.19 .14 -.05</td>
</tr>
<tr>
<td>11</td>
<td>5.50 .20 -2.86 1.48</td>
<td>-.15 .08 .21 -.06</td>
</tr>
<tr>
<td>12</td>
<td>.33 -8.78 -9.22 -8.74</td>
<td>-.33 .10 .17 .12</td>
</tr>
<tr>
<td>13</td>
<td>5.48 1.61 5.66 .23</td>
<td>-.19 -.18 -.24 -.13</td>
</tr>
<tr>
<td>14</td>
<td>10.07 - .47 -4.03 4.67</td>
<td>-.38 .01 .32 .30</td>
</tr>
<tr>
<td>15</td>
<td>- .99 -.57 2.92 -3.22</td>
<td>-.04 -.15 -.22 .04</td>
</tr>
<tr>
<td>16</td>
<td>31.53 3.20 5.45 5.94</td>
<td>-1.09 -.14 .13 -.35</td>
</tr>
<tr>
<td>17</td>
<td>.24 - .60 -2.72 -.13</td>
<td>-.07 .01 .09 -.07</td>
</tr>
<tr>
<td>18</td>
<td>1.32 1.17 -1.93 -1.57</td>
<td>.04 -.08 .07 .02</td>
</tr>
</tbody>
</table>

**Constant** | **F-value** | **L.S.**
--- | --- | ---
.94 | .21 | N.S.
.91 | .76 | N.S.
-2.38 | .55 | N.S.
-1.03 | 2.25 | N.S.
- .51 | .68 | N.S.
8.19 | 1.58 | N.S.
- .32 | 1.11 | N.S.
-1.13 | 1.77 | N.S.
3.34 | .53 | N.S.
-6.92 | .53 | N.S.
-2.17 | .80 | N.S.
7.84 | 1.95 | N.S.
.43 | .05 | N.S.
-1.79 | 2.35 | N.S.
3.19 | .94 | N.S.
-3.72 | .59 | N.S.
.84 | 1.83 | N.S.
1.22 | 2.03 | N.S.
### TABLE VIII

**REGRESSION SLOPE OF ACT ENGLISH TEST, CATEGORICAL VECTOR COEFFICIENTS, AND F-VALUE**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Regression Line Slope</th>
<th>Categorical Vector Coefficients</th>
<th></th>
<th></th>
<th></th>
<th>Constant</th>
<th>F-value</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACT English</td>
<td>Male Exper.</td>
<td>Control</td>
<td>Female Exper.</td>
<td>Control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.11</td>
<td>.84</td>
<td>2.58</td>
<td>-.01</td>
<td>2.23</td>
<td>.54</td>
<td>2.32</td>
<td>N.S.</td>
</tr>
<tr>
<td>2</td>
<td>.10</td>
<td>-.29</td>
<td>.72</td>
<td>.08</td>
<td>-1.45</td>
<td>-.63</td>
<td>2.09</td>
<td>N.S.</td>
</tr>
<tr>
<td>3</td>
<td>-.17</td>
<td>.77</td>
<td>-1.40</td>
<td>-.14</td>
<td>.25</td>
<td>1.45</td>
<td>1.82</td>
<td>N.S.</td>
</tr>
<tr>
<td>4</td>
<td>-.08</td>
<td>1.18</td>
<td>.06</td>
<td>.70</td>
<td>.00</td>
<td>.24</td>
<td>.34</td>
<td>N.S.</td>
</tr>
<tr>
<td>5</td>
<td>.04</td>
<td>1.27</td>
<td>.35</td>
<td>-.47</td>
<td>-1.17</td>
<td>-1.15</td>
<td>.58</td>
<td>N.S.</td>
</tr>
<tr>
<td>6</td>
<td>-.03</td>
<td>.32</td>
<td>-2.26</td>
<td>-.03</td>
<td>-.86</td>
<td>2.72</td>
<td>3.65</td>
<td>.05</td>
</tr>
<tr>
<td>7</td>
<td>-.15</td>
<td>3.02</td>
<td>-.11</td>
<td>2.22</td>
<td>-1.87</td>
<td>2.09</td>
<td>6.97</td>
<td>.01</td>
</tr>
<tr>
<td>8</td>
<td>-.10</td>
<td>2.86</td>
<td>-.26</td>
<td>2.17</td>
<td>-.04</td>
<td>.63</td>
<td>4.44</td>
<td>.05</td>
</tr>
<tr>
<td>9</td>
<td>-.07</td>
<td>-.21</td>
<td>-2.12</td>
<td>.00</td>
<td>-1.69</td>
<td>3.02</td>
<td>4.81</td>
<td>.01</td>
</tr>
<tr>
<td>10</td>
<td>-.33</td>
<td>11.87</td>
<td>.00</td>
<td>8.29</td>
<td>-1.00</td>
<td>2.90</td>
<td>5.86</td>
<td>.01</td>
</tr>
<tr>
<td>11</td>
<td>.02</td>
<td>1.52</td>
<td>.09</td>
<td>-.16</td>
<td>-.92</td>
<td>-1.04</td>
<td>1.78</td>
<td>N.S.</td>
</tr>
<tr>
<td>12</td>
<td>.03</td>
<td>.12</td>
<td>-1.65</td>
<td>-.42</td>
<td>-1.19</td>
<td>1.86</td>
<td>2.64</td>
<td>N.S.</td>
</tr>
<tr>
<td>13</td>
<td>-.18</td>
<td>3.38</td>
<td>-.18</td>
<td>2.67</td>
<td>-.77</td>
<td>2.22</td>
<td>10.36</td>
<td>.01</td>
</tr>
<tr>
<td>14</td>
<td>-.11</td>
<td>3.97</td>
<td>.01</td>
<td>3.29</td>
<td>.80</td>
<td>.44</td>
<td>11.09</td>
<td>.01</td>
</tr>
<tr>
<td>15</td>
<td>-.09</td>
<td>-.27</td>
<td>-1.72</td>
<td>.00</td>
<td>-1.29</td>
<td>3.38</td>
<td>4.14</td>
<td>.05</td>
</tr>
<tr>
<td>16</td>
<td>-.34</td>
<td>12.21</td>
<td>.00</td>
<td>8.67</td>
<td>-.66</td>
<td>2.77</td>
<td>9.14</td>
<td>.01</td>
</tr>
<tr>
<td>17</td>
<td>-.01</td>
<td>.01</td>
<td>.62</td>
<td>.15</td>
<td>-.04</td>
<td>.05</td>
<td>2.51</td>
<td>N.S.</td>
</tr>
<tr>
<td>18</td>
<td>.00</td>
<td>.00</td>
<td>-.50</td>
<td>.13</td>
<td>-.61</td>
<td>.64</td>
<td>5.0</td>
<td>N.S.</td>
</tr>
</tbody>
</table>
females, control females). Graphically, if the four regression lines were plotted, would they be parallel?

In Table VII, the categorical vector coefficient added to the regression constant indicates the point of interception on the Y or criterion axis where it actually plotted on a graph. The slope of regression column indicates the amount of gain on the criterion for each point of increase of ACT English Test score. As reported in Table VII, the F-ratio tests with the models described in Chapter III, revealed no significant differences among the four sub-groups. The .05 level of confidence in this step required F-values ranging from 2.66 to 2.67. Since the null hypothesis could not be rejected, it was possible to proceed with the main question of interest.

The data for the subsequent phase of the analyses are presented in Table VIII. In this table, the slope of the ACT English Test regression line identifies the change on the criteria measures for each point of change of ACT English Test score. The categorical vector coefficients indicate the differences within groups on the criterion measures for students with the same ACT English Test score. For example, by referring to Table VIII, the Crary American History Test gain score (criterion 1) of a male student in the experimental group with an ACT English Test score of 20 could be identified by multiplying 20 by .11 and adding to this the categorical vector coefficient of .84 and the constant of .54. The same process would apply to any other given ACT
English Test score. The comparison was between males in the experimental and control groups and females in the experimental and control groups.

It is also interesting to note in Table VIII the incidence of negative categorical vector coefficients in the female population for both the experimental and control groups. While not statistically significant, it may also be noted that the trend for the Crary American History Test and the Purdue Attitude Scale for English were in favor of the control group while the trend for the total of the Cooperative English Expression Test was in favor of the experimental group. In addition, the high incidence of negative values in the regression lines for the ACT English Test indicate that for a majority of the criteria, as the ACT scores increased the gains on the criteria decreased.

The .05 level of confidence required F-values ranging from 3.05 to 3.06 while the .01 level required F-values ranging from 4.72 to 4.74 for the respective criteria. It may be seen that for the Watson-Glaser Critical Thinking Appraisal totals (criteria 10 and 16), the values were significant. There were also significant values for certain subtests of this instrument. In the case of the Watson-Glaser Critical Thinking Appraisal, the null hypothesis was rejected and the research hypothesis was accepted. As related to the other criteria, the null hypothesis could not be rejected, and the research hypothesis was rejected.
The Fourth Hypothesis

The fourth hypothesis was stated as follows: With sex and ACT mathematics score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group. Table IX and Table X (see pages 108 and 109) should be examined consecutively in sequence to interpret the two phases of the statistical treatment.

Again, the preliminary test of the null hypothesis that there were no statistically significant differences in the change in the criteria per unit change in the control variable in the four subgroups could not be rejected. Data for this phase of the analysis are presented in Table IX. An examination of this table reveals the finding of no significant differences on any criterion. Again, F-values of 2.66 and 2.67 were required for the .05 level of confidence.

The data for the second phase of the procedure are reported in Table X. It may be seen that the F-value for the Cray American History Test did not reach the necessary 3.05 for the .05 level of confidence. The same may be said for the Cooperative English Expression Test (3.06) and the Purdue Attitude Scales (3.06 for English and 3.05 for history). This was not the case for the Watson-Glaser Critical Thinking Appraisal in both the English and history comparison situations. F-values of 5.35 and 7.97 were obtained as compared to the 4.75 and 4.72 values necessary for the .01 level of
### Table IX

CATEGORICAL VECTOR COEFFICIENTS, CATEGORICAL REGRESSION SLOPES OF ACT MATHEMATICS, AND F-VALUE

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Categorical Vector Coefficients</th>
<th>Slope of Regression ACT Mathematics</th>
<th>Constant</th>
<th>F-value</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>Regression Line Slope</td>
<td>Categorical Vector Coefficients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>ACT Math</td>
<td>Male Exper.</td>
<td>Male Control</td>
<td>Female Exper.</td>
<td>Female Control</td>
</tr>
<tr>
<td>1</td>
<td>.08</td>
<td>.00</td>
<td>1.75</td>
<td>-.41</td>
<td>1.69</td>
</tr>
<tr>
<td>2</td>
<td>.06</td>
<td>-.71</td>
<td>.15</td>
<td>.00</td>
<td>-1.73</td>
</tr>
<tr>
<td>3</td>
<td>-.08</td>
<td>1.46</td>
<td>-.43</td>
<td>.00</td>
<td>.77</td>
</tr>
<tr>
<td>4</td>
<td>-.01</td>
<td>1.02</td>
<td>.04</td>
<td>.32</td>
<td>-.14</td>
</tr>
<tr>
<td>5</td>
<td>.01</td>
<td>.97</td>
<td>.00</td>
<td>-.64</td>
<td>-1.45</td>
</tr>
<tr>
<td>6</td>
<td>.00</td>
<td>.02</td>
<td>-2.52</td>
<td>-.44</td>
<td>-1.16</td>
</tr>
<tr>
<td>7</td>
<td>.03</td>
<td>2.92</td>
<td>.00</td>
<td>1.66</td>
<td>-1.85</td>
</tr>
<tr>
<td>8</td>
<td>.05</td>
<td>2.97</td>
<td>.00</td>
<td>1.96</td>
<td>.30</td>
</tr>
<tr>
<td>9</td>
<td>.00</td>
<td>.01</td>
<td>-1.80</td>
<td>.00</td>
<td>-1.45</td>
</tr>
<tr>
<td>10</td>
<td>.11</td>
<td>11.39</td>
<td>.00</td>
<td>6.80</td>
<td>-1.09</td>
</tr>
<tr>
<td>11</td>
<td>.08</td>
<td>1.51</td>
<td>.18</td>
<td>-.10</td>
<td>-.71</td>
</tr>
<tr>
<td>12</td>
<td>.05</td>
<td>.00</td>
<td>-1.74</td>
<td>-.45</td>
<td>-1.18</td>
</tr>
<tr>
<td>13</td>
<td>-.02</td>
<td>3.35</td>
<td>.00</td>
<td>2.08</td>
<td>-.82</td>
</tr>
<tr>
<td>14</td>
<td>-.01</td>
<td>3.83</td>
<td>.00</td>
<td>2.81</td>
<td>-.15</td>
</tr>
<tr>
<td>15</td>
<td>.01</td>
<td>.00</td>
<td>-1.35</td>
<td>-.01</td>
<td>-1.02</td>
</tr>
<tr>
<td>16</td>
<td>.12</td>
<td>11.59</td>
<td>.00</td>
<td>6.99</td>
<td>-.91</td>
</tr>
<tr>
<td>17</td>
<td>.02</td>
<td>.00</td>
<td>.66</td>
<td>.13</td>
<td>.00</td>
</tr>
<tr>
<td>18</td>
<td>.01</td>
<td>.00</td>
<td>-.49</td>
<td>.14</td>
<td>-.58</td>
</tr>
</tbody>
</table>
confidence. Furthermore, certain subtests of this instrument (criteria 6, 7, 8, 9, 13, 14 and 15), also reached or exceeded the F-values of 3.06 (English comparison) and 3.05 (history comparison) for the .05 level of confidence.

A negative trend seems to be evident in the female population as indicated by the high incidence of negative values in those columns. It may also be seen that the control group's performance on the Crary American History Test and the Purdue Attitude Scale for English were statistically better than that of the experimental group, but the trend was reversed in the case of the Cooperative English Expression Test.

The null hypothesis of no significant differences could not be rejected, and the research hypothesis was rejected as it related to the Crary American History Test, the Cooperative English Expression Test and the Purdue Attitude Scale. In the case of the Watson-Glaser Critical Thinking Appraisal, however, the null hypothesis was rejected, and the research hypothesis was accepted.

The Fifth Hypothesis

The fifth hypothesis stated: With sex and ACT social studies score held constant, the mean gain of the experimental group on each of the above measures will be significantly greater than the mean gain of the control group. Table XI and Table XII (see pages 111 and 112) present the data
TABLE XI

CATEGORICAL VECTOR COEFFICIENTS, CATEGORICAL REGRESSION SLOPES OF
ACT SOCIAL STUDIES, AND F-VALUE

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Categorical Vector Coefficients</th>
<th>Slope of Regression ACT Social Studies</th>
<th></th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male Exper.</td>
<td>Control</td>
<td>Female Exper.</td>
<td>Control</td>
</tr>
<tr>
<td>1</td>
<td>-1.99</td>
<td>3.99</td>
<td>2.91</td>
<td>4.33</td>
</tr>
<tr>
<td>2</td>
<td>-3.53</td>
<td>-.16</td>
<td>1.18</td>
<td>-3.34</td>
</tr>
<tr>
<td>3</td>
<td>.40</td>
<td>-4.06</td>
<td>5.56</td>
<td>1.68</td>
</tr>
<tr>
<td>4</td>
<td>3.07</td>
<td>-3.63</td>
<td>6.38</td>
<td>-1.64</td>
</tr>
<tr>
<td>5</td>
<td>3.22</td>
<td>3.49</td>
<td>-2.54</td>
<td>-.06</td>
</tr>
<tr>
<td>6</td>
<td>1.79</td>
<td>.41</td>
<td>-2.15</td>
<td>-.48</td>
</tr>
<tr>
<td>7</td>
<td>3.57</td>
<td>2.20</td>
<td>-.53</td>
<td>-2.19</td>
</tr>
<tr>
<td>8</td>
<td>8.11</td>
<td>2.70</td>
<td>-.02</td>
<td>6.74</td>
</tr>
<tr>
<td>9</td>
<td>.00</td>
<td>-2.13</td>
<td>-.88</td>
<td>-2.38</td>
</tr>
<tr>
<td>10</td>
<td>19.57</td>
<td>8.91</td>
<td>-2.98</td>
<td>4.70</td>
</tr>
<tr>
<td>11</td>
<td>1.56</td>
<td>1.08</td>
<td>-4.19</td>
<td>1.37</td>
</tr>
<tr>
<td>12</td>
<td>-.43</td>
<td>-3.37</td>
<td>-.43</td>
<td>-2.43</td>
</tr>
<tr>
<td>13</td>
<td>3.89</td>
<td>3.58</td>
<td>-.24</td>
<td>-.55</td>
</tr>
<tr>
<td>14</td>
<td>7.80</td>
<td>3.85</td>
<td>-.33</td>
<td>6.61</td>
</tr>
<tr>
<td>15</td>
<td>.28</td>
<td>-.21</td>
<td>-.60</td>
<td>-1.47</td>
</tr>
<tr>
<td>16</td>
<td>15.82</td>
<td>8.11</td>
<td>-6.77</td>
<td>6.37</td>
</tr>
<tr>
<td>17</td>
<td>-.01</td>
<td>.98</td>
<td>-.06</td>
<td>.57</td>
</tr>
<tr>
<td>18</td>
<td>.49</td>
<td>-.09</td>
<td>-.19</td>
<td>-.64</td>
</tr>
</tbody>
</table>
### TABLE XII

**REGRESSION SLOPE OF ACT SOCIAL STUDIES, CATEGORICAL VECTOR COEFFICIENTS, AND F-VALUE**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Regression Line Slope ACT Social Studies</th>
<th>Categorical Vector Coefficients</th>
<th>F-value</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Male Exper.</td>
<td>Male Control</td>
<td>Female Exper.</td>
</tr>
<tr>
<td>1</td>
<td>.10</td>
<td>.00</td>
<td>1.65</td>
<td>-.64</td>
</tr>
<tr>
<td>2</td>
<td>.05</td>
<td>-.61</td>
<td>.32</td>
<td>.00</td>
</tr>
<tr>
<td>3</td>
<td>.01</td>
<td>1.82</td>
<td>.00</td>
<td>.47</td>
</tr>
<tr>
<td>4</td>
<td>.08</td>
<td>.81</td>
<td>.00</td>
<td>.09</td>
</tr>
<tr>
<td>5</td>
<td>.06</td>
<td>.94</td>
<td>.02</td>
<td>-.73</td>
</tr>
<tr>
<td>6</td>
<td>.01</td>
<td>.00</td>
<td>-2.53</td>
<td>-.47</td>
</tr>
<tr>
<td>7</td>
<td>.01</td>
<td>2.93</td>
<td>.00</td>
<td>1.67</td>
</tr>
<tr>
<td>8</td>
<td>.04</td>
<td>2.95</td>
<td>.01</td>
<td>1.88</td>
</tr>
<tr>
<td>9</td>
<td>.01</td>
<td>.00</td>
<td>-1.80</td>
<td>-.03</td>
</tr>
<tr>
<td>10</td>
<td>.11</td>
<td>12.36</td>
<td>1.06</td>
<td>7.56</td>
</tr>
<tr>
<td>11</td>
<td>.01</td>
<td>1.44</td>
<td>.00</td>
<td>-.20</td>
</tr>
<tr>
<td>12</td>
<td>.03</td>
<td>.00</td>
<td>-1.79</td>
<td>-.51</td>
</tr>
<tr>
<td>13</td>
<td>-.05</td>
<td>3.36</td>
<td>.00</td>
<td>2.19</td>
</tr>
<tr>
<td>14</td>
<td>-.03</td>
<td>3.84</td>
<td>.00</td>
<td>2.87</td>
</tr>
<tr>
<td>15</td>
<td>-.02</td>
<td>.00</td>
<td>-1.35</td>
<td>.02</td>
</tr>
<tr>
<td>16</td>
<td>-.06</td>
<td>11.81</td>
<td>.01</td>
<td>7.30</td>
</tr>
<tr>
<td>17</td>
<td>.02</td>
<td>.02</td>
<td>.66</td>
<td>.07</td>
</tr>
<tr>
<td>18</td>
<td>-.02</td>
<td>-.13</td>
<td>-.65</td>
<td>.01</td>
</tr>
</tbody>
</table>
relevant to this hypothesis. Again, the tables should be examined in sequence.

An examination of Table XI reveals no significant differences in the gain relationships in the criteria and the ACT Social Studies Test scores among the four subgroups in the preliminary step of the analyses. The .05 level of confidence required an F-value of at least 2.66. Therefore, the null hypothesis could not be rejected. It was then appropriate to proceed with the test of the null hypothesis of the main question. Table XII indicates the findings. In the case of criterion 1, the required F-value for the .05 level of confidence was 3.05 and for criteria 2, 3, and 4 it was 3.06. For criteria 17 and 18, the necessary F-value for the .05 level of confidence was 3.06 and 3.05 respectively. It may be seen that these were not reached. In the case of the Watson-Glaser Critical Thinking Appraisal, however, the .01 level of confidence F-values of 4.75 and 4.72 were surpassed in the English and history comparisons respectively (criteria 10 and 16). Furthermore, F-values for some sub-tests of this instrument also reached this level while others were at or better than the required 3.06 and 3.05 F-value for the .05 level of confidence.

Again, negative trends seem to be evident in the female experimental and control groups. Also, the F-value for the history comparison (criterion 16) was noticeably larger than that for the English comparison (criterion 10) in the case
of the Watson-Glaser Critical Thinking Appraisal. It may also be noted that the trend continued in favor of the control group on the Crary American History Test and the Purdue Attitude Scale for English, but in favor of the experimental group on the Cooperative English Expression Test.

Since the null hypotheses could not be rejected, research Hypotheses VA, B, and D were rejected. In the case of Hypothesis VC involving the Watson-Glaser Critical Thinking Appraisal, however, the null hypothesis was rejected and the research hypothesis was accepted.

The Sixth Hypothesis

The sixth hypothesis was stated as follows: With sex and ACT natural science score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group. Sequentially, Table XIII and Table XIV (see pages 115 and 116) present the data for an analysis of results relevant to this hypothesis.

The lack of significant differences noted in Table XIII among the four subgroups indicated that the null hypothesis for this preliminary phase of the analyses could not be rejected. The F-values of 2.66 and 2.67 necessary for the .05 level of confidence were not approximated. It was then possible to proceed with tests of the null form of the research hypothesis. These results are reported in Table XIV.
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Categorical Vector Coefficients</th>
<th>Slope of Regression ACT Natural Science</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>1</td>
<td>- 4.23</td>
<td>.11</td>
<td>2.59</td>
<td>1.01</td>
</tr>
<tr>
<td>2</td>
<td>- .77</td>
<td>.26</td>
<td>2.46</td>
<td>-4.52</td>
</tr>
<tr>
<td>3</td>
<td>3.98</td>
<td>- 1.22</td>
<td>3.95</td>
<td>.99</td>
</tr>
<tr>
<td>4</td>
<td>4.40</td>
<td>.25</td>
<td>2.72</td>
<td>-2.34</td>
</tr>
<tr>
<td>5</td>
<td>.19</td>
<td>1.10</td>
<td>2.81</td>
<td>-5.68</td>
</tr>
<tr>
<td>6</td>
<td>3.80</td>
<td>.13</td>
<td>.93</td>
<td>-1.49</td>
</tr>
<tr>
<td>7</td>
<td>1.97</td>
<td>1.11</td>
<td>2.71</td>
<td>-3.22</td>
</tr>
<tr>
<td>8</td>
<td>3.80</td>
<td>-.19</td>
<td>.05</td>
<td>2.02</td>
</tr>
<tr>
<td>9</td>
<td>-.01</td>
<td>.24</td>
<td>.55</td>
<td>-4.99</td>
</tr>
<tr>
<td>10</td>
<td>20.76</td>
<td>12.19</td>
<td>12.14</td>
<td>-2.34</td>
</tr>
<tr>
<td>11</td>
<td>1.12</td>
<td>1.17</td>
<td>1.89</td>
<td>- .17</td>
</tr>
<tr>
<td>12</td>
<td>1.28</td>
<td>3.21</td>
<td>1.55</td>
<td>-.89</td>
</tr>
<tr>
<td>13</td>
<td>2.09</td>
<td>3.20</td>
<td>2.85</td>
<td>1.33</td>
</tr>
<tr>
<td>14</td>
<td>6.42</td>
<td>1.00</td>
<td>2.69</td>
<td>5.49</td>
</tr>
<tr>
<td>15</td>
<td>-.19</td>
<td>3.40</td>
<td>.37</td>
<td>-5.08</td>
</tr>
<tr>
<td>16</td>
<td>13.23</td>
<td>1.20</td>
<td>4.63</td>
<td>3.28</td>
</tr>
<tr>
<td>17</td>
<td>.07</td>
<td>.95</td>
<td>.76</td>
<td>.09</td>
</tr>
<tr>
<td>18</td>
<td>1.53</td>
<td>.76</td>
<td>.58</td>
<td>-.38</td>
</tr>
</tbody>
</table>
TABLE XIV

REGRESSION SLOPE OF NATURAL SCIENCE, CATEGORICAL VECTOR COEFFICIENTS, AND F-VALUE

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Regression Line Slope</th>
<th>Categorical Vector Coefficients</th>
<th>Male</th>
<th>Female</th>
<th>Constant</th>
<th>F-value</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACT Natural Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.18</td>
<td>Exper. 1.79</td>
<td>.41</td>
<td>1.93</td>
<td>.24</td>
<td>2.63</td>
<td>N.S.</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>- .73</td>
<td>.22</td>
<td>-.01</td>
<td>-1.70</td>
<td>.23</td>
<td>2.22</td>
</tr>
<tr>
<td>3</td>
<td>.04</td>
<td>1.29</td>
<td>-.46</td>
<td>.00</td>
<td>1.03</td>
<td>-2.85</td>
<td>1.39</td>
</tr>
<tr>
<td>4</td>
<td>.10</td>
<td>.73</td>
<td>-.02</td>
<td>.19</td>
<td>.01</td>
<td>-2.81</td>
<td>.15</td>
</tr>
<tr>
<td>5</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td>-.64</td>
<td>-1.42</td>
<td>-.52</td>
<td>.69</td>
</tr>
<tr>
<td>6</td>
<td>-.04</td>
<td>.44</td>
<td>-2.12</td>
<td>-.04</td>
<td>-.89</td>
<td>2.90</td>
<td>3.70</td>
</tr>
<tr>
<td>7</td>
<td>.01</td>
<td>2.94</td>
<td>.02</td>
<td>1.68</td>
<td>-1.86</td>
<td>-.56</td>
<td>5.88</td>
</tr>
<tr>
<td>8</td>
<td>-.04</td>
<td>2.92</td>
<td>-.08</td>
<td>1.89</td>
<td>.00</td>
<td>-.45</td>
<td>4.12</td>
</tr>
<tr>
<td>9</td>
<td>-.10</td>
<td>.00</td>
<td>-1.85</td>
<td>-.05</td>
<td>-1.80</td>
<td>3.53</td>
<td>5.04</td>
</tr>
<tr>
<td>10</td>
<td>-.14</td>
<td>11.49</td>
<td>.01</td>
<td>6.81</td>
<td>-1.75</td>
<td>.38</td>
<td>5.58</td>
</tr>
<tr>
<td>11</td>
<td>-.01</td>
<td>1.47</td>
<td>.01</td>
<td>-.16</td>
<td>-1.01</td>
<td>-.49</td>
<td>1.95</td>
</tr>
<tr>
<td>12</td>
<td>-.04</td>
<td>.44</td>
<td>-1.41</td>
<td>-.03</td>
<td>-.99</td>
<td>2.72</td>
<td>3.04</td>
</tr>
<tr>
<td>13</td>
<td>-.07</td>
<td>3.40</td>
<td>.00</td>
<td>2.10</td>
<td>-.95</td>
<td>-.46</td>
<td>9.13</td>
</tr>
<tr>
<td>14</td>
<td>-.06</td>
<td>4.15</td>
<td>.26</td>
<td>3.10</td>
<td>-.02</td>
<td>-1.25</td>
<td>10.73</td>
</tr>
<tr>
<td>15</td>
<td>-.02</td>
<td>.02</td>
<td>-1.34</td>
<td>.00</td>
<td>-1.05</td>
<td>1.97</td>
<td>3.43</td>
</tr>
<tr>
<td>16</td>
<td>-.19</td>
<td>11.96</td>
<td>.01</td>
<td>7.26</td>
<td>-1.48</td>
<td>.83</td>
<td>8.79</td>
</tr>
<tr>
<td>17</td>
<td>.03</td>
<td>-.01</td>
<td>.66</td>
<td>.12</td>
<td>.03</td>
<td>-.69</td>
<td>2.93</td>
</tr>
<tr>
<td>18</td>
<td>.01</td>
<td>.00</td>
<td>-.49</td>
<td>.14</td>
<td>-.57</td>
<td>.40</td>
<td>.57</td>
</tr>
</tbody>
</table>
In relation to the Crary American History Test, the F-value of 2.63 did not reach the 3.05 value which was necessary for the .05 level of confidence. The necessary 3.06 level was not reached in connection with the Cooperative English Expression Test, nor were the necessary 3.06 and 3.05 levels reached in relation to the Purdue Attitude Scales, although the statistical trend was in the direction of the English control group. However, the 3.05 (.05) and 4.72 (.01) values were reached in nine instances involving the Watson-Glaser Critical Thinking Appraisal. In one instance (criterion 12) the 3.04 level was reached, closely approximating the required level of 3.05 for the .05 level of confidence. It may be seen also that the F-value relative to the Watson-Glaser Critical Thinking Appraisal in the history comparison (criterion 16) was larger than the F-value of the same instrument in the English comparison (criterion 10). Again, it is interesting to note the high incidence of negative trends in the female population. By contrast, the males in the experimental group in all but one instance had positive trends.

Since the null hypothesis relating to three of the four instruments could not be rejected, the research hypothesis in these instances was rejected. In relation to the Watson-Glaser Critical Thinking Appraisal, however, the null hypothesis was rejected, and the research hypothesis was accepted.
The Seventh Hypothesis

The seventh hypothesis was stated as follows: With sex and ACT composite score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group. Table XV and Table XVI (see pages 119 and 120) present the data for the two statistical phases of analysis of this hypothesis.

Data in Table XV indicate the preliminary findings that the regression lines of the four subgroups did not depart significantly from the parallel. The F-value of 2.66 necessary for the .05 level of confidence was not approximated.

It was then appropriate to proceed with tests of the null hypothesis related to the main question. An examination of Table XVI reveals that the F-values relative to nine of the criteria did not reach the F-values of 3.05 and 3.06 which were necessary for the .05 level of confidence for the situations involved. In the case of the Watson-Glaser Critical Thinking Appraisal total score, however, the necessary F-values of 4.75 and 4.72 for the .01 level of confidence in the English and history comparisons were surpassed (criteria 10 and 16). Again, the F-value for the history comparison with this instrument was the larger. It may also be noted that the negative trends for the female population persisted.

The null hypothesis could not be rejected for the Cray American History Test, the Cooperative English Expression
TABLE XV

CATEGORICAL VECTOR COEFFICIENTS, CATEGORICAL REGRESSION SLOPES OF ACT COMPOSITE, AND F-VALUE

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Categorical Vector Coefficients</th>
<th>Slope of Regression ACT Composite</th>
<th>Constant</th>
<th>F-value</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male Exper.</td>
<td>Male Control</td>
<td>Female Exper.</td>
<td>Female Control</td>
<td>Male Exper.</td>
</tr>
<tr>
<td>1</td>
<td>-3.56</td>
<td>3.37</td>
<td>4.06</td>
<td>3.17</td>
<td>.41</td>
</tr>
<tr>
<td>2</td>
<td>-2.77</td>
<td>.05</td>
<td>-1.17</td>
<td>-2.48</td>
<td>.17</td>
</tr>
<tr>
<td>3</td>
<td>7.58</td>
<td>2.76</td>
<td>15.59</td>
<td>8.11</td>
<td>-.03</td>
</tr>
<tr>
<td>4</td>
<td>-1.88</td>
<td>-3.35</td>
<td>8.18</td>
<td>-1.52</td>
<td>.16</td>
</tr>
<tr>
<td>5</td>
<td>1.35</td>
<td>1.96</td>
<td>-4.73</td>
<td>-3.14</td>
<td>.01</td>
</tr>
<tr>
<td>6</td>
<td>2.93</td>
<td>-.67</td>
<td>-3.10</td>
<td>-2.64</td>
<td>-.15</td>
</tr>
<tr>
<td>7</td>
<td>2.60</td>
<td>2.20</td>
<td>1.76</td>
<td>-1.74</td>
<td>.03</td>
</tr>
<tr>
<td>8</td>
<td>6.49</td>
<td>.28</td>
<td>-2.17</td>
<td>5.55</td>
<td>-.12</td>
</tr>
<tr>
<td>9</td>
<td>.00</td>
<td>-1.81</td>
<td>.00</td>
<td>-4.53</td>
<td>-.09</td>
</tr>
<tr>
<td>10</td>
<td>19.11</td>
<td>6.75</td>
<td>2.82</td>
<td>0.00</td>
<td>-.30</td>
</tr>
<tr>
<td>11</td>
<td>2.56</td>
<td>.61</td>
<td>3.53</td>
<td>1.70</td>
<td>.01</td>
</tr>
<tr>
<td>12</td>
<td>.00</td>
<td>-6.24</td>
<td>6.04</td>
<td>-4.35</td>
<td>-.15</td>
</tr>
<tr>
<td>13</td>
<td>3.44</td>
<td>4.41</td>
<td>2.61</td>
<td>2.42</td>
<td>.03</td>
</tr>
<tr>
<td>14</td>
<td>7.00</td>
<td>.93</td>
<td>-1.71</td>
<td>7.31</td>
<td>-.12</td>
</tr>
<tr>
<td>15</td>
<td>.00</td>
<td>-1.84</td>
<td>.00</td>
<td>-3.76</td>
<td>-.09</td>
</tr>
<tr>
<td>16</td>
<td>17.17</td>
<td>2.08</td>
<td>-4.80</td>
<td>7.78</td>
<td>-.31</td>
</tr>
<tr>
<td>17</td>
<td>.10</td>
<td>.30</td>
<td>.86</td>
<td>.72</td>
<td>.02</td>
</tr>
<tr>
<td>18</td>
<td>.07</td>
<td>.10</td>
<td>.39</td>
<td>-1.07</td>
<td>.00</td>
</tr>
</tbody>
</table>
### TABLE XVI

**Regression Slope of ACT Composite, Categorical Vector Coefficients, and F-value**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Regression Line Slope</th>
<th>Categorical Vector Coefficients</th>
<th>Constant</th>
<th>F-value</th>
<th>L.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACT Composite</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.19</td>
<td>Male Exper.</td>
<td>.05</td>
<td>1.89</td>
<td>- .59</td>
</tr>
<tr>
<td>2</td>
<td>.11</td>
<td>Male Control</td>
<td>.71</td>
<td>.29</td>
<td>- .08</td>
</tr>
<tr>
<td>3</td>
<td>- .07</td>
<td>Female Exper.</td>
<td>1.93</td>
<td>.00</td>
<td>.59</td>
</tr>
<tr>
<td>4</td>
<td>.06</td>
<td>Female Control</td>
<td>.87</td>
<td>.00</td>
<td>.18</td>
</tr>
<tr>
<td>5</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>- .03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>- .02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>- .06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>- .05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>- .11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>- .07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>- .04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>- .15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Test and the Purdue Attitude Scale, and the research hypothesis was rejected. Relative to the Watson-Glaser Critical Thinking Appraisal, however, the null hypothesis was rejected and the research hypothesis was accepted.

The Eighth Hypothesis

The eighth hypothesis was stated as follows: With sex and pretest score held constant, the mean gain made by the experimental group on each of the above measures will be significantly greater than the mean gain of the control group. Table XVII and Table XVIII (see pages 122 and 123) sequentially present the data relevant to this hypothesis.

As in the analyses of Hypotheses III through VII, the first step in the statistical procedure was to determine whether or not the change in the criterion per unit change in the control variable could be considered the same for the four subgroups (Special Program males, control males, Special Program females, control females). Stated in the null form, the hypothesis for this step was that there were not significant differences in the change in the criterion per unit change in the pretest score for the four subgroups. In short, the null hypothesis of this step may be graphically represented as stating that if the regression lines for the four subgroups were plotted, the lines would not depart significantly from the parallel. It was necessary to make this determination before it was possible to proceed with the main question of interest.
**TABLE XVII**

CATEGORICAL VECTOR COEFFICIENTS, CATEGORICAL REGRESSION SLOPES OF CRITERIA PRETESTS, AND F-VALUE

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Categorical Vector Coefficients</th>
<th>Slope of Regression Criteria Pretests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.84</td>
<td>3.37</td>
</tr>
<tr>
<td>2</td>
<td>-3.55</td>
<td>2.90</td>
</tr>
<tr>
<td>3</td>
<td>-0.16</td>
<td>-1.78</td>
</tr>
<tr>
<td>4</td>
<td>-0.68</td>
<td>1.49</td>
</tr>
<tr>
<td>5</td>
<td>5.61</td>
<td>2.23</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
<td>-3.51</td>
</tr>
<tr>
<td>7</td>
<td>-0.09</td>
<td>-6.99</td>
</tr>
<tr>
<td>8</td>
<td>10.46</td>
<td>1.06</td>
</tr>
<tr>
<td>9</td>
<td>0.82</td>
<td>-1.62</td>
</tr>
<tr>
<td>10</td>
<td>50.32</td>
<td>7.63</td>
</tr>
<tr>
<td>11</td>
<td>5.48</td>
<td>2.83</td>
</tr>
<tr>
<td>12</td>
<td>0.99</td>
<td>-0.55</td>
</tr>
<tr>
<td>13</td>
<td>8.58</td>
<td>4.03</td>
</tr>
<tr>
<td>14</td>
<td>10.12</td>
<td>3.01</td>
</tr>
<tr>
<td>15</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>39.08</td>
<td>7.88</td>
</tr>
<tr>
<td>17</td>
<td>3.44</td>
<td>2.16</td>
</tr>
<tr>
<td>18</td>
<td>2.09</td>
<td>2.00</td>
</tr>
<tr>
<td>Criterion</td>
<td>Regression Line Slope Criteria Pretests</td>
<td>Categorical Vector Coefficients</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-.14</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-.35</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-.34</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-.25</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-.82</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-.78</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-.80</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-.81</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>-.64</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>-.78</td>
<td></td>
</tr>
</tbody>
</table>
An examination of Table XVII reveals that in eight instances it was necessary that this null hypothesis be rejected. The F-value for the .05 level of confidence was 2.67, and for the .01 level of confidence it was 3.91. In these eight instances it was not possible to proceed with the main question of interest.

An examination of Table XVIII reveals that in the analyses of the main question, findings were at or better than the .01 level of confidence for the Watson-Glaser Critical Thinking Appraisal’s subtest of Evaluation of Arguments for the English comparison (4.75).

It should be recalled that the categorical vector coefficients in Table XVIII indicate the differences within groups on the criteria measures for students with the same pretest score on the instruments in question. The comparison was between males in experimental and control groups and females in experimental and control groups. The statistical trends favored the control group on the Crary American History Test and the Purdue Attitude Scale for English, but favored the experimental group on the Cooperative English Expression Test total score. The F-value of 2.13 relative to the Cooperative English Expression Test: Effectiveness, approached but did not reach the required 3.06 value for the .05 level of confidence. The subtest of Inference (criteria 11) of the Watson-Glaser Critical Thinking Appraisal also approached the .05 level with 2.35 as compared to the required 3.05.
In all instances involving the Crary American History Test, the Cooperative English Expression Test, and the Purdue Attitude Scales, the null hypothesis could not be rejected, and the research hypothesis was rejected. With reference to the Watson-Glaser Critical Thinking Appraisal, the null hypothesis could not be rejected with the exception of the subtest of Evaluation of Arguments with the English groups. With this exception the research hypothesis was rejected. In the case of the Evaluation of Arguments for the English groups, however, the null hypothesis was rejected, and the research hypothesis was accepted.
CHAPTER V

SUMMARY, CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS

The problem of this study was to determine the differences in achievement, critical thinking ability and attitude toward subjects of junior college students which could be attributed to two approaches to the teaching of introductory English composition and American history. Five classes in each of these disciplines meeting in a conventional pattern comprised the control groups. The experimental group was comprised of students concurrently enrolled in both of the disciplines participating in a composite of procedures involving coordinated large lecture sessions, small seminars and reduced class contact time combined with guided independent study.

Experimental and control groups were given alternate forms of four criteria instruments before and after the experimental treatment. The instruments were the Crary American History Test, the Cooperative English Expression Test, the Watson-Glaser Critical Thinking Appraisal, and the Purdue Attitude Scale, Part A, Attitude Toward Any Subject. The ACT subtests and composite score as well as pretests of the criteria instruments were used as control variants. The data were subjected to multiple linear regression analyses.
Summary of Findings

1. The mean gain made by the experimental group was not significantly greater than the mean gain of the control group on the Crary American History Test, the Cooperative English Expression Test, and the Purdue Attitude Scale for English. Significant mean gains were made by the experimental group on the Watson-Glaser Critical Thinking Appraisal and the Purdue Attitude Scale for history.

2. With sex held constant, the mean gain made by the experimental group was not significantly greater than the mean gain of the control group on the Crary American History Test, the Cooperative English Expression Test, and the Purdue Attitude Scale for English. Significant mean gains were made by the experimental group on the Watson-Glaser Critical Thinking Appraisal and the Purdue Attitude Scale for history.

3. With sex and ACT English score held constant, the mean gain made by the experimental group was not significantly greater than the control group on three of the four measures. Significant mean gains were made by the experimental group on the Watson-Glaser Critical Thinking Appraisal.

4. With sex and ACT mathematics score held constant, the mean gain made by the experimental group was not significantly greater than the control group on three of the four measures. The measure on which significant gains were made by the experimental group was the Watson-Glaser Critical Thinking Appraisal.
5. With sex and ACT social studies score held constant, the mean gain made by the experimental group was not significantly greater than the mean gain of the control group on three of the four measures. Significant mean gains were made on the Watson-Glaser Critical Thinking Appraisal by the experimental group.

6. With sex and ACT natural science score held constant, the mean gain made by the experimental group on three of the four measures was not significantly greater than the control group. Gains on the Watson-Glaser Critical Thinking Appraisal were significantly greater for the experimental group, however.

7. With sex and ACT composite score held constant, the mean gain made by the experimental group on three of the four measures was not significantly greater than the mean gain of the control group. On the Watson-Glaser Critical Thinking Appraisal, however, the experimental group made significant mean gains.

8. With sex and pretest score held constant, significantly greater mean gains were made by the experimental group on the Watson-Glaser Critical Thinking Appraisal subtest of Evaluation of Arguments as compared to the English control group. There were no significant differences in gains on the other three measures, however.
Conclusions

Within the limitations of this study, the following conclusions are presented:

1. Achievement in the factual content of a discipline is not significantly related to variations in instructional methodology.

2. Attitude changes toward subjects vary from discipline to discipline although the same methods of instruction may be utilized.

3. The ability to think critically may be increased by the deliberate utilization of instructional procedures designed to achieve this goal.

Implications

Based on the conclusions of this study, the following implications are formulated:

1. With the added evidence of this study, teachers may proceed with the utilization of the many methods and techniques at their command with reduced concern for resulting content achievement differences.

2. The evidence of this study may contribute to a further awareness of the need to consider the appropriate relationship of method to subject matter when attitudinal goals are involved in planning.

3. The evidence of this study may encourage teachers to provide for deliberate manipulation of the instructional
environment in order to directly stimulate growth in the processes of critical thinking.

Recommendations

The conclusions, implications, and limitations of this study suggest several recommendations for further research.

1. The present study should be replicated with the use of other appropriate standardized instruments as they become available.

2. Research is needed regarding the appropriateness of various approaches to different subject matter. Studies are needed that are concerned with identification of significant factors such as the discipline's degree of reference to personal problems, to bodies of facts, to acquisition of skills, and to the level of academic sophistication.

3. Additional research should explore the differences of instructors as they relate to the competencies required.

4. Information is needed concerning the importance of differences in students. Correlation studies are needed involving components such as age, degree of emotional and social maturity, personality factors, and scores on socio-personal attitude scales.

5. The relationship of the current social climate to the effectiveness of different instructional methods should be identified. Implications of questions should be explored regarding a society characterized by social and protest
manifestations. Research procedures should attempt to identify characteristics of instructional methods which correlate with characteristics of the society's interaction patterns.

6. Additional research inputs are needed concerning the long-run effects of the two treatments.

7. Research needs to be conducted relevant to effects of different combinations of the methods of instruction described in this study.
APPENDIX A

EXAMPLE OF A STUDY GUIDE FOR THE EXPERIMENTAL GROUP

Tape One—Unity

Perhaps it will help you if you think of the purpose in your topic sentence as the controlling idea. Make certain, then, that your topic sentence contains a key word or group of words that expresses a controlling idea. A controlling idea will help you to limit your subject to one you can deal with more completely in a paragraph and to avoid the kind of broad, general topic sentence that tempts students to include a variety of detail only loosely related to their central idea. Here are some examples of broad, general topic sentences with suggested revisions of them. Each revision sharpens the focus of the original sentence by stressing a controlling idea.

<table>
<thead>
<tr>
<th>Original</th>
<th>Revised</th>
</tr>
</thead>
<tbody>
<tr>
<td>To the Moon, by J. Chapman, is an interesting book.</td>
<td>To the Moon, by J. Chapman, presents an interesting account of the problems of landing a man on the moon.</td>
</tr>
<tr>
<td>Professional ice hockey is exciting.</td>
<td>Professional ice hockey is fiercely competitive.</td>
</tr>
<tr>
<td>Communism is evil.</td>
<td>Communism smothers man's individuality by restricting his freedom of expression.</td>
</tr>
<tr>
<td>Theodore Roosevelt was a good President.</td>
<td>Theodore Roosevelt did much to preserve the natural beauty of America.</td>
</tr>
</tbody>
</table>

In the following sentences, circle the word or words that contain the controlling idea.

Example: A dictionary is a useful book for a college student.

1. Getting an education is a harrowing experience.
2. I have had several unusual experiences with blind dates.
3. Traveling by airplane has several advantages over traveling by train.
4. Good English is clear, appropriate, and vivid.
5. There are three important steps in barbecuing a steak.
6. Campus politics often involves more social than political activity.

Revise the following topic sentences to narrow the focus and stress a dominant idea.

Example: Original: The new Carter automobile is a fine car.
Revision: Superior workmanship, beautiful design, and economic operation make the new Carter a fine automobile.

1. Teen-age marriages are foolish.
2. Abraham Lincoln was a great President.
3. A college education is becoming more essential every year.
5. Motion pictures should be censored. (should not)
6. The test of nuclear weapons should be suspended. (should not)
7. The American automobile needs to be redesigned.
8. Shakespeare is a great playwright.
9. Football is my favorite sport.
10. A good friendship is a worthwhile experience.
BIBLIOGRAPHY

Books


Becker, Samuel L., *Teaching by the Discussion Method*, Iowa City, State University of Iowa, 1958.


College Media Catalogue, Dallas, El Centro Junior College, 1968.


Havighurst, Robert J., American Higher Education in the 1960's, Columbus, Ohio State University Press, 1960.


Johnson, B. Lamar, New Directions for Instruction in the Junior College, Los Angeles, The University of California, 1965.


---


**Articles**


Bane, C. L., "The Lecture vs. the Class Discussion Method of Teaching," *School and Society*, XXI (March, 1925), 300-302.


Corey, Stephen M., "Learning from Lectures vs. Learning from Reading," *Journal of Educational Psychology*, XXV (September, 1934), 459-470.


Hovey, Donald E., Howard E. Gruber, and Glenn Terrell, "Effects of Self-Directed Study on Course Achievement, Retention and Curiosity," Journal of Educational Research, LVI (March, 1963), 346-351.


Research Division of the Coordinating Board of the Texas
College and University System, "Enrollment Projections
Texas Post-High School Education 1968-1985," TACT

Rohrer, John H., "Large and Small Sections in College Classes,"

Ruja, Harry, "Experimenting with Discussion in College Teach-
ing: Survey of Recent Research," Educational Adminis-
tration and Supervision, XXXIX (October, 1953), 321-342.

Schellenberg, J. A., "Group Size as a Factor in Success of
Academic Discussion Groups," Journal of Educational
Sociology, XXXIII (October, 1959), 73-79.

Siegel, Lawrence, F. G. Macomber, and J. F. Adams, "The
Effectiveness of Large Group Instruction at the Univer-
sity Level," Harvard Educational Review, XXIX (Summer,
1959), 216-225.

Master Scale: A Scale to Measure Attitude Toward Any
School Subject," Purdue University Studies in Higher

Smith, H. C., "Team Work in the College Class," Journal of
Educational Psychology, XLVI (May, 1955), 274-288.

Spence, Ralph B., "Lecture and Class Discussion in Teaching
Educational Psychology," Journal of Educational Psy-
chology, XIX (October, 1928), 454-462.

Stephan, Frederick F. and Elliott G. Mishler, "The Distribu-
tion of Participation in Small Groups: An Experimental
Approximation," American Sociological Review, XVII
(October, 1952), 598-608.


Reports


Eells, Walter C., College Teachers and College Teaching: Second Supplement to the Annotated Bibliography Published in 1957, Atlanta, Southern Regional Educational Board, 1962.


Umstattd, James G., Teaching Procedures Used in Twenty-Eight Midwestern and Southwestern Colleges and Universities, Austin, University Cooperative Society, 1954.

Publications of Learned Organizations


Encyclopedia Articles


Tests


Unpublished Materials


