COMPARATIVE EFFECTS OF TWO METHODS OF TEACHING CONCEPTS OF AMERICAN LAW TO HIGH SCHOOL STUDENTS

APPROVED:

Graduate Committee:

Charles J. Cross
Major Professor

Susan W. Casey
Committee Member

Harold Luehrman
Committee Member

Dean of the School of Education

Robert B. Touloung
Dean of the Graduate School

The problem of this study was to compare two approaches to teaching concepts of American law and government as to their relative effects on achievement, attitude, and critical thinking performance of high school students.

The experimental program, designed principally for inquiry-discovery learning, was used in a pilot study of a newly adopted curriculum for a one-semester course in problems of democracy. The materials utilized were the *Justice in Urban America* series from the *Law in American Society* social studies materials. Control classes received the traditional lecture-recitation-textbook type of instruction.

A total of 453 students were included in 18 separate experimental sections distributed over 6 high schools. The control group included 434 students in 12 sections distributed over 6 high schools. Nine of the experimental sections and six of the control sections were randomly selected to receive both pretest and posttest in each of the following areas: achievement, critical thinking, and attitude. The remaining experimental and control sections received posttest only. This resulted in a four-group experimental design.
The measuring instruments utilized were a standardized achievement test entitled *Principles of Democracy Test*, the *Watson-Glaser Critical Thinking Appraisal*, the *California Test of Mental Maturity*, and a ten-concept semantic differential especially constructed for use in this study. Pretests were administered between February 2 and February 17, 1971.

Achievement, critical thinking, and attitude scores were each examined statistically by comparing posttest means of the four groups in all possible combinations. In addition, the two experimental groups were combined into one group and the posttest mean compared to the posttest mean of the combined control groups. Statistical techniques used were analysis of covariance, analysis of variance, and Duncan's New Multiple Range Test.

In four of the five combinations which involved comparing the achievement scores of the experimental groups with those of the control groups, the adjusted means of the control groups were greater than those of the experimental groups. However, in two of these four cases, the differences were not significant at the 0.05 level. In the one remaining comparison, the adjusted mean of the experimental group was greater, but the difference was not significant at the 0.05 level.

The students subjected to the experimental instruction scored significantly higher on the critical thinking test. In each of the five comparisons of experimental groups with control groups, the posttest mean of the experimental group was greater at the 0.01 level of significance.
Posttest attitude scores were significantly more positive for those students subjected to the experimental instruction. In each of the five comparisons involving experimental groups with control groups, the posttest mean of the experimental group was more positive at the 0.01 level of significance.

Conclusions from this study are limited to instructional conditions which parallel those in this study. Under the conditions existing in this study, the following is concluded:

1. As measured by the Principles of Democracy Test, the traditional type of instruction results in higher achievement scores among high school students.

2. As measured by the Critical Thinking Appraisal, the experimental instruction results in higher critical thinking performance.

3. When total score on the semantic differential is considered as a measure of attitude, the experimental instruction results in a more positive attitude on the part of high school students.
COMPARATIVE EFFECTS OF TWO METHODS OF TEACHING CONCEPTS OF AMERICAN LAW TO HIGH SCHOOL STUDENTS

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

Darrell L. Gentry, B. S., M. A. T., M. S.
Denton, Texas
August, 1971
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v</td>
</tr>
</tbody>
</table>

## Chapter

### I. INTRODUCTION 1

- Statement of the Problem
- Purposes of the Study
- Hypotheses
- Background and Significance of the Study
- Definition of Terms
- Limitations
- Basic Assumptions
- Instruments
- Procedures for Collecting Data
- Procedures for Analysis of Data

### II. SURVEY OF THE LITERATURE 25

- New Instructional Programs for Social Studies
- Inquiry and Discovery Techniques
- Critical Thinking
- Achievement
- Attitude
- Summary

### III. EXPERIMENTAL PROCEDURES 54

- Description of the Sample
- Instrumentation
- Experimental Design and Collection of Data
- Procedures for Analysis of Data

### IV. ANALYSIS OF DATA 70

- Achievement
- Critical Thinking
- Attitude
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. SUMMARY AND CONCLUSIONS</td>
<td>99</td>
</tr>
<tr>
<td>Experimental Setting and Design</td>
<td></td>
</tr>
<tr>
<td>Findings Related to Achievement</td>
<td></td>
</tr>
<tr>
<td>Findings Related to Critical Thinking</td>
<td></td>
</tr>
<tr>
<td>Findings Related to Attitude</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td></td>
</tr>
<tr>
<td>Recommendations</td>
<td></td>
</tr>
<tr>
<td>APPENDIX</td>
<td>118</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>119</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Professional Qualification Data on Experimental and Control Teachers</td>
<td>56</td>
</tr>
<tr>
<td>II. Number and Per Cent of Subjects Dropped Completely From the Study Because of Insufficient Data</td>
<td>71</td>
</tr>
<tr>
<td>III. Results of Analysis of Covariance on Achievement Posttest Scores</td>
<td>73</td>
</tr>
<tr>
<td>IV. Results on Critical Thinking Posttest Means of Combined Experimental and Combined Control Groups</td>
<td>78</td>
</tr>
<tr>
<td>V. Results of Analysis of Covariance on Posttest Scores of Group A and Group C</td>
<td>79</td>
</tr>
<tr>
<td>VI. Analysis of Variance for Critical Thinking Posttest Scores of Groups A, B, C, and D</td>
<td>80</td>
</tr>
<tr>
<td>VII. Results of Duncan's New Multiple Range Test on Critical Thinking Posttest Scores for Groups A, B, C, and D</td>
<td>80</td>
</tr>
<tr>
<td>VIII. Results of $t$-test on Composite Attitude Posttest Means of Combined Experimental and Combined Control Groups</td>
<td>85</td>
</tr>
<tr>
<td>IX. Analysis of Variance on Semantic Differential Composite Posttest Scores of Groups A, B, C, and D</td>
<td>86</td>
</tr>
<tr>
<td>X. Results of Duncan's New Multiple Range Test on Semantic Differential Composite Posttest Scores for Groups A, B, C, and D</td>
<td>87</td>
</tr>
<tr>
<td>XI. $t$-test on Semantic Differential Individual Concept Posttest Means of the Combined Experimental and Combined Control Groups</td>
<td>90</td>
</tr>
</tbody>
</table>
Table

XII. Analysis of Variance for Semantic Differential Posttest Scores of Groups A, B, C, and D ............. 92

XIII. Results of Duncan's New Multiple Range Test on Individual Concept Posttest Scores for Groups A, B, C, and D ............. 94
CHAPTER I

INTRODUCTION

There is substantial dissatisfaction and disenchantment among the young people with the current social, economic, and political conditions throughout the world. Dissent, often of a violent nature, is being vividly expressed in the streets, at government offices and establishments, and on college and high school campuses. To summarize the situation:

At this moment in history, the American consensus is in considerable disarray. People are deeply divided over issues of race, poverty, war, and personal values—all of which are entangled in the web of law and are fundamental to the maintenance of American rights, freedoms and responsibilities (8, p. 237).

Various forms of protest such as marches, picketing, and sit-ins are now common mechanisms for expressing concern for these issues. Too often protest actions exceed the bounds of reasonable decorum and result in loss of life and in extensive damage to private and public property. The participants, regardless of their personal position on the issues, frequently show either misunderstanding of or disregard and disrespect for the law, the police, the courts, and the rights of other individuals (8, pp. 237-238). Public education, which has assumed a part of the responsibility for inculcating values and understanding, must also accept a portion of the responsibility for the increase in deviant behavior. Recognizing
their responsibility, educators and members of other professions have sought solutions to these problems. In several instances educators and attorneys have combined their efforts to devise new curricula for grades K-12 which, it is hoped, will instill better understanding of and a more positive attitude toward government under the rule of law (1, 8, 10). These new programs have been tried and evaluated, and there is some evidence of success. However, there is a need for further study and refinement of these programs in relation to their effectiveness upon pupils who live in different geographical locations and come from different cultural and socio-economic backgrounds. This study was proposed as an effort to partially fulfill that need.

Statement of the Problem

The problem of this study was to compare two methods of teaching concepts of American law and government as to their relative effects on achievement, attitude, and critical thinking ability of high school students.

Purposes of the Study

The purpose of this study was to determine if there is a significant difference between a specific inquiry-type approach and the traditional lecture-recitation-textbook approach to teaching concepts of American law and government as each affects the performance of high school students in the following areas: 1) achievement, 2) attitude, and
critical thinking. The inquiry approach studied was utilized in the pilot program for a Problems of Democracy curriculum entitled Law in a Changing Society. The materials used were originally developed through the joint efforts of the Chicago Bar Association and the Chicago Public School System and its teaching staff and are made available through the Law in American Society Foundation.

Hypotheses

To carry out the purposes of this study, the following hypotheses were formulated. For purposes of these hypotheses and this study, experimental group is defined as those classes utilizing the Law in American Society materials; control group is defined as those classes utilizing the traditional lecture-recitation-textbook method. The hypotheses are as follows:

1. When measured by a selected achievement test, there will be no significant difference between the mean posttest achievement scores of the following groups:

   a. The combined experimental groups and the combined control groups.

   b. The experimental group subjected to both pretest and posttest and the control group subjected to both pretest and posttest.

   c. The experimental group subjected to both pretest and posttest and the control group subjected to posttest only.
d. The experimental group subjected to posttest only and the control group subjected to both pretest and posttest.

e. The experimental group subjected to posttest only and the control group subjected to posttest only.

f. The experimental group subjected to both pretest and posttest and the experimental group subjected to posttest only.

g. The control group subjected to both pretest and posttest and the control group subjected to posttest only.

2. There will be a significant difference between the mean posttest scores of certain groups on the Watson-Glaser Critical Thinking Appraisal test as follows:

a. The combined experimental groups will score significantly higher than the combined control groups.

b. The experimental group subjected to both pretest and posttest will score significantly higher than the control group subjected to both pretest and posttest.

c. The experimental group subjected to both pretest and posttest will score significantly higher than the control group subjected to posttest only.

d. The experimental group subjected to posttest only will score significantly higher than the control group subjected to both pretest and posttest.
e. The experimental group subjected to posttest only will score significantly higher than the control group subjected to posttest only.

3. There will be no significant difference between the mean posttest scores of the following groups on the Watson-Glaser Critical Thinking Appraisal test:
   a. The experimental group subjected to both pretest and posttest and the experimental group subjected to posttest only.
   b. The control group subjected to both pretest and posttest and the control group subjected to posttest only.

4. Among certain groups, there will be a significant difference in attitude as reflected by the mean of the posttest composite scores on the ten concepts included on the semantic differential test as follows:
   a. The combined experimental groups will show a significantly more positive attitude than the combined control groups.
   b. The experimental group subjected to both pretest and posttest will show a significantly more positive attitude than the control group subjected to both pretest and posttest.
   c. The experimental group subjected to both pretest and posttest will show a significantly more positive
attitude than the control group subjected to posttest only.

d. The experimental group subjected to posttest only will show a significantly more positive attitude than the control group subjected to both pretest and posttest.

e. The experimental group subjected to posttest only will show a significantly more positive attitude than the control group subjected to posttest only.

5. Among certain groups, there will be no significant difference in attitude as reflected by the mean of the posttest composite scores on the ten concepts included on the semantic differential test as follows:

a. The experimental group subjected to both pretest and posttest and the experimental group subjected to posttest only.

b. The control groups subjected to both pretest and posttest and the control group subjected to posttest only.

6. As measured by the semantic differential, the group composed of all experimental subjects will show a significantly more positive mean posttest score than the group composed of all control subjects on

a. The concept "policemen"

b. The concept "courts of law"

c. The concept "constitutional rights"
d. The concept "United States Supreme Court"

e. The concept "trial by jury"

f. The concept "the United States Constitution"

g. The concept "majority rule"

h. The concept "the Bill of Rights"

i. The concept "political parties"

j. The concept "minority rights"

7. As measured by the semantic differential, the group of experimental subjects receiving both pretest and posttest will show a significantly more positive mean posttest attitude score than the group of control subjects receiving both pretest and posttest on

a. The concept "policemen"

b. The concept "courts of law"

c. The concept "constitutional rights"

d. The concept "United States Supreme Court"

e. The concept "trial by jury"

f. The concept "United States Constitution"

g. The concept "majority rule"

h. The concept "the Bill of Rights"

i. The concept "political parties"

j. The concept "minority rights"

8. As measured by the semantic differential, the group of experimental subjects receiving both pretest and posttest will show a significantly more positive mean posttest attitude
score than the group of control subjects receiving posttest only on

a. The concept "policemen"
b. The concept "courts of law"
c. The concept "constitutional rights"
d. The concept "United States Supreme Court"
e. The concept "trial by jury"
f. The concept "United States Constitution"
g. The concept "majority rule"
h. The concept "the Bill of Rights"
i. The concept "political parties"
j. The concept "minority rights"

9. As measured by the semantic differential, the group of experimental subjects receiving posttest only will show a significantly more positive mean posttest attitude score than the group of control subjects receiving both pretest and posttest on

a. The concept "policemen"
b. The concept "courts of law"
c. The concept "constitutional rights"
d. The concept "United States Supreme Court"
e. The concept "trial by jury"
f. The concept "United States Constitution"
g. The concept "majority rule"
h. The concept "the Bill of Rights"
i. The concept "political parties"

j. The concept "minority rights"

10. As measured by the semantic differential, the group of experimental subjects receiving posttest only will show a significantly more positive mean posttest attitude score than the group of control subjects receiving posttest only on

a. The concept "policemen"

b. The concept "courts of law"

c. The concept "constitutional rights"

d. The concept "United States Supreme Court"

e. The concept "trial by jury"

f. The concept "United States Constitution"

g. The concept "majority rule"

h. The concept "the Bill of Rights"

i. The concept "political parties"

j. The concept "minority rights"

11. As measured by the semantic differential, there will be no significant difference between the mean posttest score of the experimental subjects receiving both pretest and posttest and the mean posttest score of the experimental subjects receiving posttest only on

a. The concept "policemen"

b. The concept "courts of law"

c. The concept "constitutional rights"

d. The concept "United States Supreme Court"

e. The concept "trial by jury"
12. As measured by the semantic differential, there will be no significant difference between the mean posttest score of the control subjects receiving both pretest and posttest and the mean posttest score of the control subjects receiving posttest only on

a. The concept "policemen"
b. The concept "courts of law"
c. The concept "constitutional rights"
d. The concept "United States Supreme Court"
e. The concept "trial by jury"
f. The concept "United States Constitution"
g. The concept "majority rule"
h. The concept "the Bill of Rights"
i. The concept "political parties"
j. The concept "minority rights"

Background and Significance of the Study

For many years in the Western World, the primary function of secondary schools was to provide a classical-liberal education denoted by liberal arts subjects as distinguished from vocational subjects. In about the middle of the nineteenth
century, Herbert Spencer upset tradition by claiming that in emphasizing the humanities, the schoolmen had their priorities reversed—practical rather than liberal studies should have priority (6, p. 32). This resulted in a view of education as education for complete living, an ideal which finally was expressed officially in the Seven Cardinal Principles. The responsibility of the schools thus became that of meeting the needs of pupils in all facets of their lives. This notion of education was pre-eminent until after World War II.

In Western society, where the rate of social change is not constant but accelerating, educators have become somewhat disenchanted by the immensity of the task required by "education for complete living." In an effort to solve the dilemma, there has been a trend back to the classical belief that the primary objective of education should be the cultivation of the rational powers.

The current trend, however, is not so much an abandonment of those goals and objectives prevailing in the earlier part of the twentieth century, but rather a new approach to the attainment of those objectives. In 1961, the Educational Policies Commission of the National Education Association adopted the development of rational powers as the central role of education without entirely discarding the Seven Cardinal Principles (6). Under classical-liberal education, the prevailing thought was that the best method for developing the rational powers was to study and assimilate the knowledge
and information contained in the classical subjects or disciplines. The current view de-emphasizes facts and information as being important within themselves, and strives instead to develop rational powers by stressing an understanding of those underlying theories, laws, and principles of each discipline which have broad generalization and application. In summary, it may be said that the trend since the latter part of the nineteenth century has been from an emphasis on the knowledge and information contained in the classical liberal arts to an emphasis on practical skills and knowledge, and then from the practical to a recent emphasis on the understanding and application of basic laws, theories, and principles with the disciplines acting as the medium for conveying understanding.

The newer approaches to learning, curriculum development, and teaching techniques draw from various sources, but most of their rationale may be illustrated through two theories. One is the Gestalt-field theory of learning, which views learning as a process of gaining or changing insights, outlooks, or thought patterns (2, p. 9). The other is the Taxonomy of Educational Objectives, which holds that types of behavior may be classified into three major categories: cognitive, affective, and psychomotor. Little work has been done in the psychomotor domain, but tentative classifications have been developed in the remaining two. As stated by Bloom et al. and referenced by Clark (6, pp. 40-55), the hierarchy of the cognitive domain in order of increasing complexity of behavior
is 1) knowledge, 2) comprehension, 3) application, 4) analysis, 5) synthesis, and 6) evaluation. The hierarchy of the affective domain is 1) receiving, 2) responding, 3) valuing, 4) organization, and 5) characterization.

Words such as "discovery," "inquiry," "systems," "structure," and "process" are key words in the current approaches which are being used in an attempt to guide the student toward higher levels of behavior. There are many approaches; but if there is one common theme, it is probably that expressed by the Educational Policies Commission of the National Education Association: "The purpose which runs through and strengthens all other educational purposes—the common thread of education—is the development of the ability to think" (7, p. 12).

That critical thinking is a part of this common thread is evident. Hardy summarized critical thinking succinctly in the following statement: "Critical thinking has been described in terms of the ability to recognize valid inferences and unstated assumptions, to reason deductively, and to weigh evidence" (9, p. 26). In speaking of the rational powers of an individual, the Educational Policies Commission states:

These powers involve the process of recalling and imagining, classifying and generalizing, comparing and evaluating, analyzing and synthesizing, and deducing and inferring. These processes enable one to apply logic and the available evidence to his ideas, attitudes and actions, and to pursue better whatever goals he may have (7, p. 5).

A parallel of content between rational powers and critical thinking can be detected when words from Hardy's statement
are paired with words from the statement of the Educational Policies Commission, e.g., inference-inferring; deductively-deducing; weigh evidence-evaluate; recognize-recall. Further, these processes correspond closely to the five areas of ability sampled by the Watson-Glaser Critical Thinking Appraisal test which are inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments (12). Thus it would seem that rational powers and critical thinking are closely related parts of the cognitive domain.

The last line from the above statement of the Educational Policies Commission implies also a relationship between rational powers and attitude. Further evidence of such a relationship is given by the following:

Thus, by critical thinking, he [the student] can deepen his respect for the importance of values and strengthen his sense of responsibility . . . .

Development of the ability to reason can also lead to dedication to the values which inhere in rationality: commitment to honesty, accuracy, and personal reliability; respect for the intellect and for the intellectual life; devotion to the expansion of knowledge (7, pp. 7, 8).

Critical thinking, then, seems to influence not only those types of behavior classified in the cognitive domain but also those types of behavior in the affective domain such as values and attitudes.

The Law in American Society materials, which were developed in Chicago, is a program of instruction especially designed and presented to develop in students those powers discussed above. Jerome S. Bruner has stated: "... we have
not reaped the benefits that might have come from a joining of the efforts of eminent scholars, wise and skillful teachers, and those trained in the fields related to teaching and learning” (3, p. 3). In line with this suggested rationale, the Law in American Society project was begun in 1965 and developed through the joint efforts of the Chicago Bar Association, The Chicago Board of Education, and selected teachers from the Chicago Public Schools (11, p. 1). The program existed from 1966 through 1969 as a federally funded project and since that time it has become a permanent part of the curriculum in the Chicago schools.

The program is designed to instill in students a better understanding of and appreciation for the role of law in American society—especially towards the law, the courts, and the role and responsibilities of citizens in our society. The materials are designed and teachers are trained to stress the inquiry approach to concepts, generalizations, and principles which have broad application to practical and current problems, and to present instruction in an atmosphere of open discussion undominated by the teacher. mediums used are the case-study, mock trials, and simulation.

The program has been evaluated in terms of effects upon students of the Chicago inner-city schools, and the effects have been judged significantly favorable (5). The materials are now being offered for use by other schools, and heretofore
uninvestigated aspects of the materials as well as their effectiveness in different cultural and geographical areas, merit study and evaluation.

The adoption of the Law in American Society materials by a large school district in North Texas for use in a pilot program under a different title--Law in a Changing Society--formed the setting for this study.

Definition of Terms

For the purposes of this study, the following operational definitions were formulated:

Achievement.--Achievement is defined as knowledge and understanding of subject matter content as measured by the Principles of Democracy Test, published by Science Research Associates, Inc., Chicago, Illinois.

Attitude.--Attitude toward a concept is defined as the mean rating on a set of selected evaluative scales composed of bipolar adjectives.

Positive attitude.--A positive attitude toward a concept is defined as a mean rating on the selected evaluative scales which is greater than or equal to 4.5.

Negative attitude.--A negative attitude toward a concept is defined as a mean rating on the selected evaluative scales which is less than or equal to 3.5.
**Neutral attitude.**--A neutral attitude toward a concept is defined as a mean rating on the selected evaluative scales which is greater than 3.5 but less than 4.5.

**Critical thinking ability.**--Critical thinking ability is defined as the ability to make inferences, recognize assumptions, reason deductively, evaluate arguments, and make interpretations as measured by the Watson-Glaser Critical Thinking Appraisal test.

**Experimental groups.**--Experimental groups are defined as those classes or sections which studied the *Law in American Society* materials.

**Control groups.**--Control groups are defined as those classes which received instruction through the traditional lecture-recitation-textbook method and were selected for comparison.

**Limitations**

This study was limited to selected schools within one large city in North Texas and to those subjects assigned to classes utilizing the *Law in American Society* materials and to classes assigned as control groups.

This study was further limited to a situation in which the teacher variable was not controlled within the experimental design. Control classes were not taught by the same teachers who taught the experimental classes, and the teachers assigned
to experimental classes received approximately thirty hours of special training in the inquiry approach to teaching and in the subject matter content of *Law in American Society* materials. A further discussion of this aspect is included in Chapter III.

**Basic Assumptions**

It is assumed that the subjects responded honestly and to the best of their ability on the instruments used to measure achievement, attitude, and critical thinking ability, and to questionnaires and instruments used to collect other needed data.

It is also assumed that the experimental schools, which were selected by the Superintendent of the school system with the help of his staff, are representative of the general cultural, socio-economic, and academic composition of the overall school system insofar as limitations and conditions on such intact-unit selection makes representative sampling feasible.

It is further assumed that the selection of the control schools, which was done by the project director, was accomplished in a professional manner and contained no bias other than that of matching control schools with experimental schools relative to cultural, socio-economic, and academic standing.
Instruments

Three measuring instruments were used in this study. A semantic differential was used to measure attitude; the Watson-Glaser Critical Thinking Appraisal was used to measure critical thinking ability; and the Principles of Democracy test was used to measure achievement. In addition to the brief description in this section, a more detailed discussion of instruments may be found in Chapter III.

The Semantic Differential

The semantic differential used in this study consisted of ten separate concepts. Ten scales, each composed of a seven-step continuum between a pair of bipolar adjectives, were included on each concept. Six of the ten scales were selected to have high loading on the evaluative (attitude) factor. Of the remaining four, two were weighted on the potency factor and two were loaded on the activity factor. The scores on the six evaluative (attitude) scales only were summed to yield an attitude score. The potency and activity scales were included on the test to obscure the purpose of measurement.

Each individual scale was rated from "one" to "seven." A score of "one" represented the extreme negative attitude score; "four" represented a neutral attitude; and "seven" represented the extreme positive attitude. When the six evaluative scales were summed, a score ranging from six to
forty-two on each concept was possible. When scores on each concept were summed, the total attitude score could range from 60 to 420.

**Watson-Glaser Critical Thinking Appraisal**

This is a widely used test developed by Goodwin Watson and Edward M. Glaser. It is designed to measure critical thinking ability by sampling the subject's ability in five areas: inference; recognition of assumptions; deduction; evaluation of arguments; and interpretation. Test reliability is well established, and group norms are available on several group levels. The test comes in two parallel forms—the Ym and Zm—and consists of 100 multiple-response items. The Ym form was used in this study.

**Principles of Democracy Test**

This is a standardized test published by Science Research Associates. It consists of sixty-five multiple-response items covering such areas as constitutional law of the United States; the structure of federal, state, and local government; the judicial, executive, and legislative branches of government; and elections, voting, and ballots. Percentile norms for grades 9 through 12 were established utilizing 7,386 students. Published reliability and validity coefficients were judged acceptable.
Procedures for Collecting Data

Six experimental schools and six control schools, all from one large school system, were used in this study. In the experimental schools, Problems of Democracy classes utilized the Law in American Society materials. Problems of Democracy classes in the control schools studied the usual or traditional curriculum.

A slightly altered version of the Solomon Four-Group Design as described by Campbell and Stanley (4) was used. There was a total of eighteen classes or sections of experimental subjects and twelve classes or sections of control subjects. Nine experimental sections and six control sections were selected, by use of a table of random numbers, to receive both pretests and posttests in achievement, critical thinking, and attitude. The remaining 50 per cent of the experimental sections and the control sections received posttest only. Pretests were administered early in the Spring semester of 1971—between February 2 and February 17. Posttests were administered near the end of the Spring semester—between May 1 and May 15. I. Q. scores were obtained from the students' permanent school records.

Procedures for Analysis of Data

Data processing was performed by the Data Processing Center at North Texas State University. Hypotheses were
tested in the null form and the 0.05 level of significance was arbitrarily selected as the point of rejection or retention of the hypotheses.

The following hypotheses were tested by analysis of covariance: 1a through 1g; and 2b.

The following hypotheses were tested by one-way analysis of variance: 2c; 2d; 2e; 3a; 3b; 4b; 4c; 4d; 4e; 5a; 5b; 7a through 7j; 8a through 8j; 9a through 9j; 10a through 10j; 11a through 11j; and 12a through 12j.

The following hypotheses were tested by the t-test: 2a; 4a; and 6a through 6j.
CHAPTER BIBLIOGRAPHY


CHAPTER II

SURVEY OF THE LITERATURE

In order to present a survey of the literature related to the problem investigated in this study, the relevant literature is organized into five categories: new instructional programs for social studies, inquiry and discovery techniques, critical thinking, achievement, and attitude.

New Instructional Programs for Social Studies

If the extensive changes in physical science and mathematics curricula which occurred in the late 1950's and early 1960's can be considered a curriculum revolution in those areas, then the current revisions in social studies curricula can probably be considered a revolution also. Among the people of the United States there seems to be occurring a shift in interest from technological problems to social problems. As a result, the social sciences appear to be receiving an increased share of attention in the public schools.

Hunt and Metcalf (24) define the goals of social studies as helping students to learn to reflectively examine issues in the closed areas of American culture. They consider closed areas to be any area of belief, behavior, or conflict characterized by relatively large amounts of irrationality,
confusion, and inconsistency. These goals cover a variety of subject areas, perhaps greater in number than those usually included in the realm of social studies. Even when confined to those subjects normally placed in the category of social studies, the number of new curriculum programs is quite large. Knight (26) has compiled an excellent, but nevertheless incomplete, annotated list of new social studies projects. He lists the project name, director, address, summary of conceptual themes, objectives, and grade level of thirty-four projects. The projects are classified under the following disciplines: anthropology, economics, geography, history, political science, social psychology, sociology, and general and interdisciplinary. These, along with other new programs, present a kaleidoscope of content.

In the words of Isadore Starr, "The issues of our streets must become a part of the content of our curriculum" (47, p. 335). Law is an issue of our streets, and it should not be surprising that this topic is receiving increased emphasis at both the secondary and elementary school levels. In an editorial, Social Education supports the contention that the subject of law has been neglected in our social studies (27, pp. 333, 334). McDaniel and Truce (32) have presented a case for teaching "the law" to young students. Johnson and Sublett (25) praise law projects at the high school level and further propose that more law be taught in the upper elementary grades.
Responding to the plea, many of the new social studies programs have included concepts of law. These concepts range from constitutional law (26) and the Bill of Rights (5) to consumer law (4) and tort cases (36). O'neil and O'neil (36) advocate the use of tort cases in the classroom because they feel tort cases are important in everyday living, they are easy to grasp, the facts involved are familiar to most citizens, and the analysis of cases requires no particular knowledge of any field of law. Gerlach (19, pp. 237-239), in building a case for teaching constitutional law and the Bill of Rights, contends that the American adolescent's education concerning the United States Constitution and the Bill of Rights is seriously deficient; public espousal and attachment to these documents is superficial and verbal. He lists as faults with most of the present curriculum: it is treated as required factual material; it is unrelated to current social, political, and economic problems; principles of the Constitution are presented as absolute self-enforcing truths; and teachers are inadequately trained. Aided by such rationale as this, new programs emphasizing law are finding their way into the broad spectrum of the social studies curriculum. The materials investigated in this study emphasize the law and law concepts. The following few pages will deal with a description of several new programs which also emphasize the teaching of law.

The Oregon Bill of Rights project, developed by the Oregon State Bar Association and the Portland Public School
System between 1964 and 1966, consists of a series of case studies on the Bill of Rights. Now published in a text entitled *Liberty and the Law*, the materials were designed to provide a basis for developing analytical and critical thinking about the freedoms guaranteed by the Bill of Rights and to give a sense of the role and function of the courts. The program is primarily discussion oriented, and reportedly has been used successfully with freshmen and sophomores and with slower students (5).

The Harvard Social Studies Project, funded by Harvard University and the United States Office of Education since 1961, has a curriculum designed around three ideas: "(1) the conceptual framework for dealing with public issues, (2) the background, principles, and structure of American constitutionalism, and (3) substantive problem units to which the concepts contained in (1) and (2) can be applied"(19, p. 248). These ideas are based upon the conviction that the analysis of public controversy should command the primary attention of the teaching of social studies in the public schools (26, p. 18).

Another program dealing with constitutionalism is the American Liberties Project of the Hartford Board of Education, Hartford, Connecticut. Through this project materials have been prepared which deal with constitutional cases presented at a level appropriate for use with non-academic students in inner-city schools. Designed for use at the twelfth-grade level, two booklets, *You and Your Civil Liberty and Problems in*
American Liberty, are to be used in conjunction with a directed reading approach. Tests for the materials come with visuals and include transparencies. This is one of the few programs which give special consideration to the less able or less academically inclined student (26, p. 12).

The World Law Fund High School Program in New York City applies the case study approach to international conflict. The materials of this instructional program emphasize the contribution of law to the control of violence in international affairs. It is interesting to note that three well-known entertainment films, Lord of the Flies, Dr. Strangelove, and High Noon, are used to initiate discussions of world order problems (26).

Several efforts toward revision of the social studies curriculum have been made in California. The Statewide Social Studies Committee (10) devised a plan designed to provide students with conceptual tools and processes of inquiry which are necessary for effective organization and interpretation of information. The Committee on Civic Education, University of California, Los Angeles, has developed elementary and secondary materials dealing with situations and case studies of important principles of constitutional democracy (26). These materials generally follow two patterns: an attempt to develop concepts in a situation related closely to student experience; and an attempt to furnish opportunities for students to develop a frame of reference. Two student booklets are available:
Your Rights and Responsibilities as an American Citizen: a Civics Casebook; and Conflicts, Politics, and Freedom. A third California program, The California Bill of Rights Project, developed a source book for teachers based upon the premise that effective teaching requires a functional presentation and not a numerical, non-discriminating catalogue of factual material (19). The materials take a point of view in one sense—the source book accepts the basic values of the Bill of Rights such as freedom of speech and religion and equal protection under the laws—but outside this, the points of view are largely left for the students to develop (45).

A program which contains perhaps less emphasis on law than those mentioned above but nevertheless does not ignore law concepts is the Lincoln Filene Center Program in Research and Development in the Social Studies, Tufts University, Medford, Massachusetts. Workers on this project have developed materials for the elementary grades which focus on intergroup relations and materials for the secondary schools which are broader and cover social studies in general. Law topics, however, receive attention. One of the three basic instructional programs is Inner-City Problems and Prospects, which deals with legal problems as reflected by topic titles contained within, i.e., "The Police: Fact and Fiction" and "Civil Disobedience: a Higher Law?" (26).

The Chicago Law in American Society Project, now also known as the Law in American Society Foundation, was begun in
1965 by the Chicago Bar Association and the Chicago Public Schools (28). The project grew out of a request for the Chicago Bar Association to supply volunteer lawyers to speak in the public schools (38). In the summer of 1966 a pilot program was conducted to test the feasibility of introducing aspects of law and legal reasoning into the public school curricula. This program consisted of a summer institute for teachers. Thirty-one twelfth-grade teachers of contemporary American history attended. The summer institute combined law lectures and education workshops and produced a resource unit for twelfth-grade contemporary American history. The developmental phase was deemed successful, and the program was expanded. Summer institutes have been continued in succeeding years and now the institutes cover four grade-levels of instructional programs: intermediate social studies, junior high school American history, high school civics and government, and high school United States history (28, 38). Published materials developed through the project are now available, and the high school government and civics materials, entitled Justice in Urban America, constitute the materials used by the experimental classes involved in this study.

Most of the programs described above have produced printed materials for classroom use. It should be noted that production of instructional materials has not been limited to printed matter such as textbooks and booklets. Various types of audio-visual teaching materials have come from these programs as
well as from other sources. Hartley (20) has compiled a list of films, tapes, transparencies, and other materials which are available. This list gives the title, nature of the materials, time requirements, and producers of each set of materials.

The discussion above is limited to those new social studies programs which have introduced the study of law into the curriculum. A summary of new programs dealing with topics other than law can be found in the works of Knight (26) and Sanders and Tanck (43).

Inquiry and Discovery Techniques

With knowledge and technology advancing at what appears to be a geometric progression, many scholars and educators have accepted the hypothesis that teaching facts and factual knowledge alone is no longer adequate. Attention has been shifted toward methods, approaches, and techniques designed to more effectively teach students to think. Whether this ability to think is labeled "critical thinking" or called by some other name, it covers a variety of skills. Within the social studies, the objectives outlined for the Civic Education Project of the National Council for Social Studies may be viewed as a small sample of the variety of skills now receiving attention. They assert that in order to create an informed, analytic, committed, and involved citizenry, the development of such intellectual processes as valuing, critical inquiry, decision-making, and problem solving should be the paramount objectives of the social studies (41).
In an effort to attain objectives such as the above, there has been a search to identify and delineate the structure of disciplines, an effort to make that which is taught more meaningful, and an attempt to discover better methods of promoting transfer of learning. Issue-centered classroom discussions and contemporary problems are utilized in an attempt to put subject content into a context relevant to today's society.

Such efforts have resulted in new curriculum programs involving approaches with various labels such as discovery, inquiry, role playing, and simulation. Some programs involve almost all of these approaches; others involve only specific combinations. Unfortunately, this sometimes makes it difficult to isolate or identify precisely the approach being used or investigated.

A point of example is the confusion which Chambers (11) pinpoints concerning discovery and inquiry. He contends that there are three points of confusion: 1) discovery learning is confused with meaningful learning; 2) discovery teaching is confused with discovery learning; and 3) discovery learning is confused with inquiry learning. Relative to the first point of confusion, most studies have been designed to test meaningfulness and discovery on the one hand with rote and didacticism on the other. Rote and meaningful is one dichotomy which refers to use of knowledge by the learner. Discovery and didacticism is a different dichotomy which refers to the
way in which information is learned. Meaningful and discovery do not necessarily have to be paired and then compared to rote and didacticism; all combinations are possible. Learning can be meaningful even when learned under didactic instruction. Concerning the second point of confusion, the fact that one teaches by the discovery method does not necessarily mean that learning occurs in that way. Yet when meaningful learning is measured in experimental studies, the interpretation often includes the assumption that the learning may be attributed to the discovery method. The converse is also true; didactic teaching does not guarantee rote learning. One may be taught by the didactic method but learn by discovery. On the third point, learning skills of inquiry and learning by discovery, Chambers contends, are two different things. Discovery learning is learning something by discovery; inquiry learning is learning how to discover. Many of the effects attributed to learning by discovery can be explained by the student's learning how to discover or how to learn.

In an effort to shed light on the points of confusion, Chambers (11) performed a study designed to control these factors. The study was performed over four levels of discovery and two levels of overlearning. The results showed that overlearning has a considerably more powerful effect on transfer of learning than does discovery. Overlearning, he proposes, seems to be important for transfer of learning not because the principle is apt to be discovered during overlearning,
but because a certain amount of practice is necessary to make the discovered principle available for transfer. Also, Loupe (30) obtained results which indicate it is possible to train one in problem-solving skills and that these skills have transfer value into the inquiry situation. These results should be of interest to Jerome Bruner because he is a proponent of non-specific transfer, i.e., learning not a skill but a general idea which can then be utilized as a basis for recognizing future problems as special cases of the idea originally mastered (8, p. 17).

In addition to the above points of confusion, role playing, sociodrama, and simulation add to the confusion because they are closely related and easily treated as synonymous. Garvey and Garvey (17) distinguish the three as follows: in role playing the student assumes the identity of another to increase his understanding of the other person; sociodrama adds the element of problem-solving; and simulation incorporates both techniques into a game structure.

Despite confusion, programs utilizing new techniques are being introduced and decisions about methodology are being made. The programs range from the use of community resources as field laboratories for real-life experiences in social studies (44) to case studies on the Bill of Rights (5). In a recent seminar composed of historians, law professors, and educators, the participants decided that presentation by case study would be the most effective instructional method and
that students should be presented with vital and contemporary problems concerning the Bill of Rights (37). Decisions such as these are sometimes based upon subjective judgement rather than empirical data or at least not upon conclusive empirical data.

Some research has been done in an effort to develop instruments and methods for measuring how well teachers use the new techniques. Massialas (31) has developed a category system to analyze or evaluate issue-centered classroom discussion. Zevin (53) built a model of the inquiry teacher constructed upon the theoretical works of John Dewey, Jerome Bruner, B. O. Smith, Byron Massialas, and others. He then operationalized it for experimental purposes by adapting Ned Flanders' Interaction Analysis to a description of the inquiry classroom. Use of these and other techniques have yielded evidence that teachers can be trained to use the new inquiry method.

Zevin (53) used his instrument to measure the inquiry methods of teachers participating in the Summer Institute for the Chicago Law in American Society Project. Fifteen teachers enrolled in the Institute were subjected to systematic inquiry training. Ten teachers participated in the Institute but received no special inquiry training. Finally, a control group was matched to the experimental group on the basis of sex, age, experience, race, and location of school. The experimental group demonstrated a significant change in the direction of inquiry, while the other two groups did not. He concluded
that training in the use of the inquiry/discovery method can be effective. He is supported in this belief by the work of Ryan \((42)\) who found that teachers can be taught to use the inquiry method even in short, in-service workshops. The effectiveness of such training programs does have limitations, however, because Dobbins \((15)\) reports that three films developed for use in teaching teachers how to use the "Public Issues Series" of the Harvard Social Studies Project were found to be inadequate for that purpose when used alone.

Research seems to show that training teachers to use the new techniques and approaches is no insurmountable problem. The superiority of these approaches over the traditional methods is not so clear-cut.

**Critical Thinking**

According to Dull \((16)\), critical thinking and problem solving may be used synonymously and the necessary factors in the process are: mastery of sub-skills; correct classroom climate; independent study; and group cooperation. Regardless of the degree of agreement or disagreement with this contention, a survey of the literature on the subject gives indication that there does exist in people an ability labeled critical thinking which can be measured and improved \((49)\). Recent research dealing with the improvement of critical thinking can be roughly grouped into two categories: 1) research dealing with techniques specifically designed to
develop critical thinking and to be used with or without a subject matter medium, and 2) research dealing with instruction in specific subjects or disciplines or with special approaches to instruction wherein concomitant effect on critical thinking performance is measured.

Relative to the first category, Herber (22) conducted a study designed to discover if critical thinking can be taught. He concluded that critical thinking performance can be improved through instruction. Wickman (52) found that critical thinking performance gains are greater simply if learning experiences are carefully selected and organized. Devine (13) says that critical thinking cannot be taught.

Rothstein (40) devised a history course specifically designed to improve critical thinking and compared it to the conventional method of teaching history. He found a significant difference in development of critical thinking in favor of the specially designed course. Lee (29) compared students who studied the traditional high school chemistry with students subjected to chemistry problems specifically designed to teach critical thinking skills and found no significant overall differences in critical thinking performance. The only significant difference was found between those experimental and control subjects who scored one standard deviation or more below the class mean on critical thinking. The difference was significant at the .10 level with the higher scores being attained by the problem-solving group.
Emphasizing instruction in critical thinking instead of subject matter, Morton (34) studied the effect of such instruction on reading improvement or achievement. He found that instruction in critical thinking resulted in significant gains in reading performance. In the same study, he also found that traditional lecture-discussion instruction in critical thinking was superior to television instruction in developing critical thinking.

Within the second category of research, Sorensen (46) compared laboratory-centered and lecture-demonstration-centered patterns of instruction in high school biology and found that the laboratory-centered instruction produced a significant increase in critical thinking performance while the lecture-demonstration-centered technique produced no statistically significant gains.

Ballew (3) compared public school algebra students who studied special discovery exercises to students who studied conventional algebra. He found that those trained in discovery showed a significant improvement in critical thinking performance while those who were trained by the conventional approach did not. Both groups showed gain, however, and even though the improvement of the experimental group was statistically significant and the improvement of the control group was not, the improvement of the experimental group was significantly greater than that of the control group in only one experimental class or section.
Hunkins (23) performed a study in which he sought to determine whether a dominant use of analysis and evaluation questions as defined by Bloom's Taxonomy would effectively stimulate the development of sixth-grade pupils' critical thinking performance when such questions were used with social studies text-type materials. Dominant use of these questions did not produce statistically significant greater effects on critical thinking performance than the traditional approach.

At the college level, Henkel (21) compared two groups of physics students relative to critical thinking as measured by the Watson-Glaser Critical Thinking Appraisal. The experimental group used Physical Science Study Committee (PSSC) materials and discussion and discovery; the control group used a traditional text and the lecture-recitation approach. Both the experimental and control groups contained some day sections and some evening sections. Results showed that all students improved in critical thinking performance, but only one day experimental section showed a statistically significant gain and this gain was not significantly greater than gains registered by other groups.

In a study involving 405 students, Garvey and Seiler (18) compared simulation teaching versus lecture-discussion teaching on students studying international relations. They reported some significant differences but no identifiable pattern in the performance of control and experimental groups on critical thinking tests. The tests used were the Watson-
Turning to a different aspect of critical thinking, Clubok (12) investigated the use of critical thinking techniques in the classroom by twelfth-grade civics teachers in the Detroit Public Schools. He studied forty-eight civics teachers and concluded the following: 1) there is no significant relationship between either total teaching experience or years of teaching civics and the amount of emphasis put on critical thinking instruction by teachers, 2) there is no significant relationship between the teacher's level of education and emphasis on the objective of critical thinking or the use of methods and techniques which foster it, 3) there is no significant relationship between semester-hours in political science possessed by the teacher and emphasis put upon critical thinking instruction in the classroom, 4) there is no significant difference between the emphasis put upon critical thinking by inner-city and outer-city teachers, and 5) most teachers stress critical thinking less than other basic objectives, and most do not use methods and techniques which foster it.

Achievement

Research dealing with the relationship between achievement and critical thinking as well as between achievement and
the newer approaches to teaching such as discovery and simulation, does not seem to show results which give indisputable evidence favoring any particular pattern.

Rothstein (40), using a history course specially designed to improve critical thinking, found no significant difference in the achievement of the two groups when those taking the special course were compared to those studying the conventional curriculum. Herber (22), however, in concentrating upon instruction geared specifically to improve critical thinking performance, found that this training also strengthened vocabulary and reading accuracy and comprehension.

In a study of moral concepts, Milgram (33) investigated the effects of inquiry-discovery instruction versus a teacher-oriented method on 111 fourth-grade and seventh-grade students. The students were posttested on Biblical facts and also on concepts of evil both immediately after instruction and two months later. Both methods effectively increased scores on retention of facts, understanding of moral concepts, and transfer of understanding of moral concepts at both age levels; but no significant difference was found between the two methods.

Richardson (39) studied the use of an inquiry-discovery method of laboratory instruction in college chemistry. The control group used commercially prepared lab manuals; the experimental group used materials prepared by the investigator. The study was conducted over three semesters and involved 200 students. He reported that the experimental group scored
significantly better on a teacher-made achievement test, but showed no significant difference in ACT composite and ACT natural science scores.

Ballew (3) found no significant difference in mathematical achievement when he compared high school algebra students who had been exposed to special discovery exercises to students subjected only to conventional instruction.

Hunkins (23) found that when analysis and evaluative questions as defined by Bloom's Taxonomy were used in sixth-grade social studies, the students showed significantly more achievement gain than students studying the traditional materials. It should be pointed out that in this study achievement was measured by an instrument constructed by the investigator.

A study comparing textbook and simulation approaches in teaching junior high school American history was done by Baker (2). He found a significant difference in the amount of immediate learning but no significant difference in the amount of retentive learning between the two groups. Garvey and Seiler performed a similar study in which they compared simulation teaching to the lecture-discussion method while teaching international relations to 405 high school students. They reported some significant differences but no identifiable pattern in the performance of the control and experimental groups on content examinations (18).
Amidon and Flanders (1) have concluded that generally all types of students learn more working with teachers who emphasize the indirect rather than the direct approach. Even in this area, they report that teacher flexibility appears to be more important as a predictor of teacher success than the either/or concept of direct-indirect influence. Apparently the ability of the teacher to shift behavior from direct to indirect and vice versa in order to meet the particular situation is the more important or effective characteristic.

**Attitude**

John Dewey (14) has elaborated extensively upon the importance of experience in the educational process. His ideas on this topic have become quite popular, and authorities accept experience as being of paramount importance in both the cognitive and affective domains of learning. Attitudes, values, feelings, and sensitivity are words commonly used to describe types of learning classified in the affective domain. One viewpoint on teaching this type of learning is reflected in the words of Taba:

> Feelings, values, and sensitivities are matters that need to be discovered rather than taught. Neither democratic values nor feelings of tolerance can be developed solely by teaching about them. This means that provisions for these objectives must include opportunities for direct experiencing of some sort and materials which affect feelings. . . . This makes the modification of attitudes and feelings one of the most difficult of educational tasks (14, p. 224).
The word "attitude," when used in its broader sense, covers a multitude of connotations as well as connotations toward a multitude of things. Despite its many and intricate ramifications, the import of attitude in the classroom learning situation is widely recognized. Bruner (9, pp. 42-50) gives considerable weight to attitude (i.e., a disposition toward learning) and also to the importance of experience in instilling a predisposition to learn. The specification of experiences which most effectively implant in the individual the predisposition to learn is one of the four major features of his theory of instruction.

Recognizing the importance of attitude, investigators have studied the effects of different techniques and methods upon the attitude of students in various areas of study. Studies related to attitude vary from attitude toward closed-circuit television instruction (7) to the interrelationships between mathematical attitude and mathematical achievement (51). In the area of social studies, Trotter (50) compared a problem-media-dialogue method of teaching American government at the junior college level with the lecture method. Although he found no significant difference in achievement, he reported a more positive attitude on the part of the problem-media-dialogue group toward American government as a college course. At a lower grade level, Nepi (35) used a case study approach in teaching the Bill of Rights to fifth
and sixth graders and reported changes in attitude significant at the .01 and .001 levels in favor of the case study approach.

Blackburn (6) conducted a study for the purpose of testing experimentally three methods of instruction in social problems to determine which method was most effective in modifying undesirable attitudes that affect adjustment. The three techniques studied were bibliography, group discussion, and sociodrama. The subjects used in the study were sixth-grade students. The number of subjects was 120. He reported sociodrama as the most effective, group discussion next, and bibliography least effective.

Summary

Among authorities, the consensus seems to be that concepts of American law should be taught at the secondary and elementary school levels and that this area has, in the past, been neglected.

Many new social studies programs are being implemented which include the study of law. Most of the programs are developed through the joint efforts of some combination of different agencies. State education agencies, large-city boards of education, local and state bar associations, and colleges or universities are the usual participants. Often the projects are partly financed by federal funds.

Most of the programs are designed around certain approaches to instruction such as inquiry, discovery, simulation, role
playing, and case study. Teacher training seems to be a prerequisite to effective use of the techniques involved in these approaches. The training of teachers in these areas seems to be effective and feasible, but also expensive.

The superiority of the different instructional techniques and approaches over traditional methods lacks conclusive empirical evidence. Reported research on the topic is abundant, but the results are often contradictory. One study yields significant differences; another reports no significant difference. In addition to a variety of experimental designs with differing amounts of controls, interpretation is further complicated by the large number of different combinations of factors investigated by the reported studies. There are several instructional techniques and approaches which are popular, and they are measured against a variety of objectives. The pre-eminence of any technique, method, or approach in attaining a given objective is not clear-cut. This appears to be true in relation to critical thinking, achievement, and attitude.
CHAPTER BIBLIOGRAPHY


CHAPTER III

EXPERIMENTAL PROCEDURES

Description of the Sample

This study was conducted in a large metropolitan school district in North Texas. In an effort to improve the problems of democracy curriculum, a pilot program was implemented within the school district in which certain selected schools would utilize the Law in American Society study materials. Six schools were selected by the superintendent of the school district to participate in the pilot project. The schools were arbitrarily selected to give representation to the different geographical, cultural, and socio-economic areas of the city. This resulted in seven teachers being assigned to teach experimental materials in eighteen sections or classes.

Because it would create extra class preparations on the teachers involved, it was deemed not feasible to assign both experimental and control classes to the same teacher. To solve the problem, control classes were obtained by selecting for each experimental school a matching control school which had a student body that paralleled as closely as possible the experimental school in cultural and socio-economic background and in academic standing. The selection of control schools was made by the project director with the consultation of his
staff. Within the six matching schools selected, six teachers were assigned to a total of twelve problems of democracy classes or sections. These classes studied the traditional materials and constituted the control group.

All teachers for the experimental and control sections were certified teachers of established qualifications. However, teachers for the experimental sections received approximately thirty hours of special training in concepts of American law and in the inquiry method of classroom instruction. In addition, each teacher of an experimental group had one or more licensed attorneys assigned to him as a consultant. The attorneys were available to visit classrooms and answer questions. These factors suggest that instruction in experimental classes and in control classes may not have been of equivalent quality. In order to determine if there were other factors which created a substantial difference between the professional training of the experimental teachers and that of the control teachers, data concerning the following items was collected: number of years of college training; the highest earned degree; the number of years teaching experience; and the number of semester-hours in social science. The data is reported in Table I.

The data in Table I shows that the two groups of teachers are not equivalent in amount of qualifications in those areas
TABLE I
PROFESSIONAL QUALIFICATION DATA ON EXPERIMENTAL AND CONTROL TEACHERS

<table>
<thead>
<tr>
<th>Teacher No.</th>
<th>Years of College</th>
<th>Highest Degree</th>
<th>Years of Teaching Experience</th>
<th>Semester Hours in Social Science</th>
<th>Teacher No.</th>
<th>Years of College</th>
<th>Highest Degree</th>
<th>Years of Teaching Experience</th>
<th>Semester Hours in Social Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.0</td>
<td>B.A.</td>
<td>13</td>
<td>80</td>
<td>8</td>
<td>4.0</td>
<td>B.A.</td>
<td>3</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>5.5</td>
<td>M.S.</td>
<td>9</td>
<td>90</td>
<td>9</td>
<td>5.0</td>
<td>B.A.</td>
<td>5</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>6.0</td>
<td>M.A.</td>
<td>10</td>
<td>40</td>
<td>10</td>
<td>7.0</td>
<td>M.S.</td>
<td>7</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>7.0</td>
<td>M.S.</td>
<td>4</td>
<td>84</td>
<td>11</td>
<td>6.0</td>
<td>B.A.</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>4.5</td>
<td>B.S.</td>
<td>5</td>
<td>67</td>
<td>12</td>
<td>4.5</td>
<td>B.A.</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>5.2</td>
<td>M.S.</td>
<td>32</td>
<td>35</td>
<td>13</td>
<td>4.5</td>
<td>B.A.</td>
<td>6</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>5.5</td>
<td>M.S.</td>
<td>30</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group Means</td>
<td>5.5</td>
<td></td>
<td>14.7</td>
<td>61.7</td>
<td>5.2</td>
<td></td>
<td></td>
<td>10.7</td>
<td>41.5</td>
</tr>
</tbody>
</table>

surveyed. In those areas surveyed, the teachers of the experimental materials have more training.

No special criterion was used in assigning students to either the experimental or control classes. The students in all sections enrolled for problems of democracy classes through the usual procedures. They were included in the study because
they happened to attend a school and enrolled in a class which was selected for this study.

Instrumentation

The Semantic Differential

The semantic differential technique developed by Charles E. Osgood was used to measure student attitude. This technique has been frequently used for measuring attitude (2, 6, 8, 9).

Test-retest reliability data collected over six concepts on six evaluative or attitude scales, after realignment according to a constant evaluation direction, yielded test-retest coefficients ranging from .87 to .93 with a mean $r$ of .91 (10, p. 192).

Osgood et al. also determined reliability coefficients and established validity while comparing semantic differential scores with Thurstone Scale scores. The scores on the respective instruments were compared on three concepts: the Church, capital punishment, and the Negro. Test-retest reliability coefficients for the evaluative scales on the semantic differential were .83, .91, and .87 respectively; for the Thurstone scores the corresponding reliability coefficients were .81, .78, and .87. The correlation coefficients between semantic differential scores and Thurstone scores on the first test were .74 on the concept "church," .81 on "capital punishment," and .82 on "the Negro." Corresponding scores on the retest were .76, .77, and .81.
respectively. Osgood has concluded: "... whatever the Thurstone scales measure, the evaluative factor of the semantic differential measures just about as well" (10, p. 194).

In a further study of validity, Osgood compared the semantic differential with a fourteen-item Guttman-type scale. Attitude scores on the semantic differential were obtained by summing over the three evaluative scales used (good-bad, fair-unfair, valuable-worthless). The rank order correlation between the two instruments was highly significant—rho = .78, p < .01 (10, p. 194).

In constructing the semantic differential used in this study, a survey questionnaire was utilized to select those key concepts for which student attitude was to be measured. Fourteen concepts were selected from the experimental materials and fourteen concepts selected from the curriculum guide for the traditional materials. These items constituted a check list on the questionnaire whereby respondents could check those concepts which they felt were most appropriate. Space was provided for respondents to enter concepts which were not listed but which they thought to be more applicable. The questionnaire was submitted to seven attorneys who were acting as consultants on the project, to the seven teachers assigned to teach experimental sections, and to the six teachers assigned to teach the control sections. Each respondent selected five concepts from the experimental materials and five concepts from the control materials. The total number
of selections for each concept was determined, and the five concepts from the experimental materials which received the most votes and the five concepts from the control materials which received the most votes were included on the semantic differential. The five concepts selected from the experimental materials were: 1) policemen, 2) courts of law, 3) constitutional rights, 4) United States Supreme Court, and 5) trial by jury. The five concepts selected from the traditional materials were: 1) United States Constitution, 2) majority rule, 3) the Bill of Rights, 4) political parties, and 5) minority rights.

Each concept was accompanied by ten separate scales. Each scale was composed of a seven-step continuum between a pair of bi-polar adjectives. The bi-polar adjective scales were selected according to Osgood's criteria (10, pp. 76-85). Six of the ten scales were selected to have high loading on the evaluative (attitude) factor. When scores were summed over the six scales, the total score for each concept could range from six to forty-two, six being the most unfavorable, twenty-four neutral, and forty-two most favorable. This provided eighteen degrees of intensity of attitude in each direction.

The remaining four scales were selected from Osgood's potency and activity factors. They were included to obscure somewhat the purpose of measurement and they were not included in the scoring. A sample concept from the semantic differential
is included in Appendix A. The scales which were scored in measuring attitude are denoted by an asterisk.

**Watson-Glaser Critical Thinking Appraisal**

The Watson-Glaser Critical Thinking Appraisal is an instrument designed to measure critical thinking abilities. The test consists of the following five subtests:

1. **Inference.**—Samples the ability to discriminate among degrees of truth or falsity of inferences drawn from the given data.

2. **Recognition of assumptions.**—Samples ability to recognize unstated assumptions or presumptions which are taken for granted.

3. **Deduction.**—Samples ability to reason deductively from given statements or premises and to recognize the relation of implication between propositions.

4. **Evaluation of arguments.**—Samples the ability to distinguish between arguments which are strong and relevant and those which are weak or irrelevant.

5. **Interpretation.**—Samples the ability to weigh evidence and to distinguish between generalizations drawn from given data (3, p. 328).

The test comes in two parallel forms—Ym and Zm. There are 100 items on each form. Hovland (7, p. 798) reports reliability coefficients from .79 to .84. Validity studies were conducted by having high school science teachers identify
those students who appeared markedly able or markedly poor in ability to reason accurately and to think logically. The test distinguished significantly between the two groups. Correlations of from .33 to .52 between the teacher's ratings and total scores on the test were obtained over four different classes. Similarly, with groups of 15 research chemists, 12 biologists, 18 engineers, and 15 accountants, there was a significant difference between those who were rated by their supervisor as being in the top and bottom halves of the group. However, the correlation in this case reportedly was not very impressive. In a further study of validity, a correlation of .70 between the Watson-Glaser Critical Thinking Appraisal and the Terman-McNemar Test of Mental Maturity was obtained.

Practice effects are relatively slight. Hovland also reports that "the average improvement for a group of secondary students retested after a week was only 0.6 points, as compared with the 6-point difference in scores... for alternate forms of the test" (7, p. 798).

The test is not without criticism. An example is one offered by Hill:

In the inference subtest they [the authors] feel it is necessary to require use of "certain commonly accepted knowledge or information which practically every person knows." This is a loophole which makes it possible to question the key to certain items, depending on what one considers to be commonly accepted knowledge (5, p. 796).

In conclusion, Hill states that the criticisms should not be allowed to obscure the over-all value of the instrument.
Hovland (7, p. 796) is less critical, and states that the test has found useful application both for selection purposes and evaluation of programs of instruction.

**Principles of Democracy Test**

When new study materials and new subject content are introduced into an established course of study, measurement of achievement using a standardized test presents the danger of penalizing the student who has been instructed only in the new materials. The alternative is to use a test designed for the new materials, and this penalizes the student instructed only in the established materials. The achievement test used in this study is the **Principles of Democracy Test**, a standardized test published by Science Research Associates, Inc., Chicago, Illinois. It was arbitrarily selected because the test items appeared to fit the course content of both the experimental and control classes better than other available standardized tests. This, it is hoped, helped to equalize the penalization of students in both groups.

The test consists of sixty-five multiple response items which are divided into the following areas: 1) constitutional and fundamental law of the United States, 2) elections, voting, and ballots, 3) congress, 4) the president and executive branch, 5) judiciary and law enforcement, 6) national government powers, 7) state and local government, 8) Declaration of Independence and Articles of Confederation, 9) civil liberties and the Bill of Rights (4, p. 1253).
The established percentile norms for the test were based on the scores of 7,386 students in grades 9 through 12 from 20 schools in 13 states. Reliability and validity studies were based on four representative samples of 400 students, one sample for each grade from 9 through 12. Reliability coefficients were computed by both the split-half technique and the Kuder-Richardson Formula 20. Values ranged from .85 to .89. Correlation coefficients to establish concurrent validity were determined by correlation with a variety of tests such as portions of the Iowa Test of Educational Development, the California Test of Mental Maturity, the Terman-McNemar Test of Mental Ability, and portions of the SRA Tests of Educational Ability (4, p. 1244). The coefficients ranged from .33 to .83 (1, p. 1243).

One serious criticism of the test was expressed by Haefner (4, p. 1244). He contends that ten of the sixty-five items do not clearly test knowledge but are cast in the form of opinion or judgement. He feels that such items would be more effective in an achievement test if they tested for knowledge rather than for the judgement of the student as to what ought to be done in a hypothetical situation.

Experimental Design and Collection of Data

Nine of the eighteen experimental sections and six of the control sections were selected by use of a table of random numbers. The sections thus selected were administered both
pretest and posttest in each of the following areas: 1) achievement, 2) attitude, and 3) critical thinking. The remaining nine experimental sections and six control sections received posttest only in each of the three areas. This created an experimental design which may be depicted as follows:

\[
0_1 \ X_a \ 0_2 \\
0_3 \ X_b \ 0_4 \\
\quad \ X_a \ 0_5 \\
\quad \ X_b \ 0_6
\]

Where 0 = an observation (testing)

\[X_a = \text{treatment (experimental instruction)}\]

\[X_b = \text{treatment (traditional instruction)}\]

Four different groups were thus provided: a group of experimental students who received both pretests and posttests, a group of experimental students who received posttests only, a group of control students who received both pretests and posttests, and a group of control students who received posttests only.

A total of 477 students were enrolled in the experimental classes and 371 enrolled in the control classes. Drop-outs, transfers, and absences resulted in insufficient data on nineteen experimental subjects and twenty-eight control subjects. These subjects were dropped from the study. The attrition rates were thus 3.98 per cent for the experimental group, 7.55 per cent for the control group, and 5.54 per cent for the combined groups.
In addition to the above loss, scores on all three instruments were not available for a considerable number of students. Because of absences, scores were often available on one or two of the instruments but missing on the remaining one(s). In such cases, students were not dropped completely from the study. Since the statistical treatment of scores on one instrument was performed independently from the scores on other instruments, it was decided to use the data when present for a particular instrument even though the data on other instruments was missing. Data processing procedures thus included all subjects supplying complete data on any given instrument in the analysis of data on that instrument. Subjects were discarded only when data was absent or incomplete for the instrument scores being analyzed. As a result, the computer read-out showed different N's for each set of instrument scores. This fact is discussed further in Chapter IV.

The problems of democracy classes involved in this study were one-semester courses. Pretests on all three instruments were administered between February 2 and February 17, 1971. Posttests were administered between May 1 and May 15, 1971.

As a matter of procedural policy, the California Test of Mental Maturity is administered to all ninth-grade students within the district and to all move-ins who enroll after completion of the ninth grade. The test is administered again to these students at the beginning of their senior year.
The scores from this test were used as covariates in the statistical treatment of data. The scores were obtained from records at the school district's central administration building or from the student's cumulative file at the school which he attended. In cases where more than one score appeared in the records, the latest score was used.

Procedures for Analysis of Data

Data processing was performed by the Data Processing Center at North Texas State University. The 0.05 level of significance was arbitrarily selected as the point of rejection when hypotheses were tested in the null form.

For purposes of brevity in discussion, the four groups of subjects are defined as follows:

Group A.—Experimental sections which received both pretests and posttests.

Group B.—Experimental sections which received posttests only.

Group C.—Control sections which received both pretests and posttests.

Group D.—Control sections which received posttests only.

In the remainder of this chapter, the individual groups are referred to as groups A, B, C, and D. Hypotheses are referred to by their number. A specific statement of hypotheses may be found in Chapter I and in Chapter IV.
Achievement

Hypotheses 1a through 1g were tested by performing analysis of covariance on posttest achievement scores for all possible combinations of groups A, B, C, and D along with analysis of covariance between groups A and B combined and groups C and D combined. In the comparison of group A with group C, both I. Q. and pretest scores were used as covariates. In the remaining cases, I. Q. scores served as the covariate.

Critical Thinking

Hypothesis 2a was tested by performing a t-test for a significant difference between the critical thinking posttest mean of groups A and B combined and the critical thinking posttest mean of groups C and D combined.

Hypothesis 2b was tested by performing analysis of covariance on the critical thinking posttest scores of group A and group C. Pretest scores on critical thinking were used as the covariate.

Hypothesis 2c, 2d, 2e, and 3a and 3b were tested by performing analysis of variance on critical thinking posttest scores for all combinations between groups A, B, C, and D.

Attitude

Hypothesis 4a was tested by performing a t-test for a significant difference between the semantic differential
posttest mean of groups A and B combined and the posttest mean of groups C and D combined.

Hypotheses 4b, 4c, 4d, 4e, and 5a and 5b were tested by performing analysis of variance on the semantic differential posttest scores of groups A, B, C, and D.

Hypotheses 6a through 6j were tested by performing a t-test for a significant difference between the posttest mean of groups A and B combined and the posttest mean of groups C and D combined on each of the individual concepts included on the semantic differential.

Hypothesis 7a through 7j; 8a through 8j; 9a through 9j; 10a through 10j; 11a through 11j; and 12a through 12j were tested by performing analysis of variance for a significant difference between the posttest means of groups A, B, C, and D on each of the individual concepts included on the semantic differential.
CHAPTER BIBLIOGRAPHY


CHAPTER IV

ANALYSIS OF DATA

The results of statistical analysis of test data concerning achievement, critical thinking, and attitude are presented in this chapter. Data was processed by the Data Processing Center at North Texas State University.

A total of 848 students were originally enrolled in the classes involved in this study. Forty-seven subjects were dropped from the study because of insufficient data. Data was not obtained because students dropped out, transferred, or were absent during testing. The number of subjects dropped from each group and the per cent loss is given in Table II.

In many cases students were present when some of the tests were administered but absent when others were administered. This created a loss of data in addition to that described in Table II. In order to keep this loss to a minimum, test scores were used whenever possible. For example: if a student's scores were available on the achievement test but missing on the attitude and critical thinking instruments, the student was not dropped completely from the study; his achievement scores were included in the processing of achievement test data. The computer program was designed to include all usable data in a given operation and exclude subjects
TABLE II
NUMBER AND PER CENT OF SUBJECTS DROPPED COMPLETELY FROM THE STUDY BECAUSE OF INSUFFICIENT DATA

<table>
<thead>
<tr>
<th>Category</th>
<th>Number Originally Enrolled</th>
<th>Number Dropped</th>
<th>Number Remaining</th>
<th>Per Cent Lost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental subjects scheduled to receive pretest and posttest</td>
<td>222</td>
<td>16</td>
<td>206</td>
<td>7.2</td>
</tr>
<tr>
<td>Experimental subjects scheduled to receive posttest only</td>
<td>255</td>
<td>3</td>
<td>252</td>
<td>1.2</td>
</tr>
<tr>
<td>Control subjects scheduled to receive pretest and posttest</td>
<td>197</td>
<td>24</td>
<td>173</td>
<td>12.2</td>
</tr>
<tr>
<td>Control subjects scheduled to receive posttest only</td>
<td>174</td>
<td>4</td>
<td>170</td>
<td>2.3</td>
</tr>
<tr>
<td>Total</td>
<td>848</td>
<td>47</td>
<td>801</td>
<td>5.5</td>
</tr>
</tbody>
</table>

from the study when data needed for the particular operation was missing or incomplete. For this reason, the data processing read-out sheet showed different N's for the same group of subjects depending upon the particular set of scores being analyzed. This accounts for the inconsistent size of particular group N's shown in the tables within this chapter.

The experimental design used in this study resulted in the following four groups or classifications of subjects:
A) experimental subjects who received both pretest and posttest, B) experimental subjects who received posttest only,
control subjects who received both pretest and posttest, and D) control subjects who received posttest only. In Chapter III these groups were defined as groups A, B, C, and D respectively. This manner of designating groups will also be utilized in this chapter.

Achievement

The instrument used to measure achievement was the Principles of Democracy Test published by Science Research Associates, Inc., Chicago, Illinois. The instrument contains sixty-five multiple response items and permits scores with a possible range from zero to sixty-five. The statistical technique used to test for a significant difference between the posttest performance of the different groups was analysis of covariance. This technique was applied individually to all possible combinations of groups A, B, C, and D and also to the mean of groups A and B combined and the mean of groups C and D combined. In the comparison of the experimental group receiving both pretest and posttest with the control group receiving both pretest and posttest (i.e., group A with group C), both I. Q. scores and pretest scores were used as covariates. The remaining comparisons involved one or more groups not receiving pretest; therefore I. Q. scores alone were used as the covariate. The test used to obtain I. Q. scores was the California Test of Mental Maturity, which is administered each year to all ninth- and
twelfth-graders in the school district. The results of the analysis of covariance are depicted in Table III.

TABLE III

RESULTS OF ANALYSIS OF COVARIANCE ON ACHIEVEMENT POSTTEST SCORES

<table>
<thead>
<tr>
<th>Groups Compared</th>
<th>Adjusted Group Means</th>
<th>N</th>
<th>F</th>
<th>P</th>
<th>S or Not S</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A+B) with (C+D)</td>
<td>(A+B): 33.6</td>
<td>(A+B): 399</td>
<td>5.145</td>
<td>0.0222</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>(C+D): 35.2</td>
<td>(C+D): 263</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A with C</td>
<td>A: 35.5</td>
<td>A: 164</td>
<td>2.729</td>
<td>0.0956</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>C: 36.9</td>
<td>C: 114</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A with D</td>
<td>A: 34.9</td>
<td>A: 172</td>
<td>0.391</td>
<td>0.5397</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>D: 34.2</td>
<td>D: 139</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B with C</td>
<td>B: 32.5</td>
<td>B: 227</td>
<td>14.988</td>
<td>0.0003</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>C: 36.4</td>
<td>C: 124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B with D</td>
<td>B: 32.8</td>
<td>B: 227</td>
<td>0.590</td>
<td>0.5507</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>D: 33.5</td>
<td>D: 139</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A with B</td>
<td>A: 36.3</td>
<td>A: 172</td>
<td>2.799</td>
<td>0.0911</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>B: 34.8</td>
<td>B: 227</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C with D</td>
<td>C: 33.8</td>
<td>C: 124</td>
<td>6.437</td>
<td>0.0114</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>D: 31.0</td>
<td>D: 139</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data in Table III relates to hypotheses 1a through 1g. These hypotheses, as stated in Chapter I, are restated as follows:

When measured by a selected achievement test, there will be no significant difference between the mean posttest achievement scores of the following groups:

a. The combined experimental groups and the combined control groups.
b. The experimental group subjected to both pre-test and posttest and the control group subjected to both pretest and posttest.

c. The experimental group subjected to both pre-test and posttest and the control group subjected to posttest only.

d. The experimental group subjected to posttest only and the control group subjected to both pretest and posttest.

e. The experimental group subjected to posttest only and the control group subjected to posttest only.

f. The experimental group subjected to both pre-test and posttest and the experimental group subjected to posttest only.

g. The control group subjected to both pretest and posttest and the control group subjected to posttest only.

Results pertaining to hypothesis la are given in row one of Table III. $F = 5.145$, and $p = 0.0222 < 0.05$. Hypothesis la is therefore rejected. There was a significant difference between the adjusted mean achievement score of the combined experimental groups and that of the combined control groups. The adjusted mean of the combined control groups was significantly greater at the 0.05 level.

Results pertaining to hypothesis lb are shown in row two of Table III. $F = 2.729$, and $p = 0.0956 > 0.05$. Hypothesis
lb is therefore retained. The difference between the adjusted mean score of the experimental group subjected to both pretest and posttest and that of the control group subjected to both pretest and posttest was not significant at the 0.05 level.

Results pertaining to hypothesis 1c are given in row three of Table III. \( F = 0.391, \) and \( p = 0.5397 > 0.05. \) Hypothesis 1c is therefore retained. The difference between the adjusted mean score of the experimental group subjected to both pretest and posttest and that of the control group subjected to posttest only was not significant at the 0.05 level.

Results pertaining to hypothesis 1d are shown in row four of Table III. \( F = 14.988, \) and \( p = 0.0003 < 0.05. \) Hypothesis 1d is therefore rejected. There was a significant difference between the adjusted mean score of the experimental group subjected to posttest only and that of the control group subjected to both pretest and posttest. The adjusted mean of the control group was significantly greater at the 0.05 level.

Results pertaining to hypothesis 1e are shown in row five of Table III. \( F = 0.590, \) and \( p = 0.5507 > 0.05. \) Hypothesis 1e is therefore retained. The difference between the adjusted mean score of the experimental group subjected to posttest only and that of the control group subjected to posttest only was not significant at the 0.05 level.
Results pertaining to hypothesis If are shown in row six of Table III. $F = 2.799$, and $p = 0.0911 > 0.05$. Hypothesis If is therefore retained. The difference between the adjusted mean score of the experimental group subjected both to pretest and posttest and that of the experimental group subjected to posttest only was not significant at the 0.05 level.

Results pertaining to hypothesis Ig are shown in row seven of Table III. $F = 6.437$, and $p = 0.0114 < 0.05$. Hypothesis Ig is therefore rejected. There was a significant difference between the adjusted mean score of the control group, subjected to both pretest and posttest and that of the control group subjected to posttest only. The adjusted mean of the control group subjected to both pretest and posttest was significantly greater at the 0.05 level.

Interpretation of the data and results presented above and conclusions appertaining thereto are discussed in Chapter V.

Critical Thinking

The instrument used to measure critical thinking performance was the Watson-Glaser Critical Thinking Appraisal, published by Harcourt, Brace & World, Inc., New York, New York. The instrument contains 100 multiple response items and permits scores with a possible range from 0 to 100.
Hypotheses 2a through 2e and hypotheses 3a and 3b pertain to critical thinking performance. These hypotheses as stated in Chapter I are as follows:

Hypothesis 2: There will be a significant difference between the mean posttest scores of certain groups on the Watson-Glaser Critical Thinking Appraisal test as follows:

a. The combined experimental groups will score significantly higher than the combined control groups.

b. The experimental group subjected to both pretest and posttest will score significantly higher than the control group subjected to both pretest and posttest.

c. The experimental group subjected to both pretest and posttest will score significantly higher than the control group subjected to posttest only.

d. The experimental group subjected to posttest only will score significantly higher than the control group subjected to both pretest and posttest.

e. The experimental group subjected to posttest only will score significantly higher than the control group subjected to posttest only.

Hypothesis 3: There will be no significant difference between the mean posttest scores of the following groups on the Watson-Glaser Critical Thinking Appraisal test:

a. The experimental group subjected to both pretest and posttest and the experimental group subjected to posttest only.
b. The control group subjected to both pretest and posttest and the control group subjected to posttest only.

Hypothesis 2a was tested by performing a $t$-test for a significant difference between the critical thinking mean posttest scores of the combined experimental groups and those of the combined control groups. The results of the $t$-test are shown in Table IV.

**TABLE IV**

**RESULTS ON CRITICAL THINKING POSTTEST MEANS OF COMBINED EXPERIMENTAL AND COMBINED CONTROL GROUPS**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>397</td>
<td>60.6</td>
<td>13.7</td>
<td>9.183</td>
<td>0.0001</td>
</tr>
<tr>
<td>Control</td>
<td>268</td>
<td>50.4</td>
<td>14.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results show that $t = 9.183$, and $p = 0.0001 < 0.05$. The computer read-out gave 0.0000 as the value of $p$. To avoid confusion and the implication of zero probability, all such values of $p$ have been rounded off and show a one in the fourth position to the right of the decimal.

Hypothesis 2a, which is stated directionally, is therefore accepted in the form stated. There was a significant difference between the mean posttest score of the combined experimental groups and that of the combined control groups.
The mean of the combined experimental groups was significantly greater at the 0.01 level.

Hypothesis 2b was tested by performing analysis of covariance on the critical thinking posttest scores of the experimental group which received both pretest and posttest and the control group which received both pretest and posttest. Critical thinking pretest scores served as the covariate. The results are shown in Table V.

**TABLE V**

RESULTS OF ANALYSIS OF COVARIANCE ON POSTTEST SCORES OF GROUP A AND GROUP C

<table>
<thead>
<tr>
<th>Groups Compared</th>
<th>Adjusted Group Means</th>
<th>N</th>
<th>F</th>
<th>p</th>
<th>S or Not S</th>
</tr>
</thead>
<tbody>
<tr>
<td>A with C</td>
<td>A: 60.91</td>
<td>164</td>
<td>68.05</td>
<td>0.0001</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>C: 44.29</td>
<td>144</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results show that $F = 68.05$, and $p = 0.0001 < 0.05$. Hypothesis 2b, which is stated directionally, is therefore accepted. There was a significant difference between the mean posttest score of the experimental group which received both pretest and posttest and that of the control group which received both pretest and posttest. The mean score of the experimental group was significantly greater at the 0.01 level.

Hypotheses 2c, 2d, 2e, 3a, and 3b were tested by performing one-way analysis of variance on the critical thinking posttest scores. Results are shown in Table VI.
TABLE VI
ANALYSIS OF VARIANCE FOR CRITICAL THINKING POSTTEST SCORES OF GROUPS A, B, C, AND D

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Variance Estimate</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>18799</td>
<td>3</td>
<td>6266</td>
<td>31.52</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>131405</td>
<td>661</td>
<td>199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>150204</td>
<td>664</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since analysis of variance showed a difference among posttest means of the four groups which was significant beyond the 0.05 level, Duncan's New Multiple Range Test was utilized to identify the groups for which the difference in means was significant. The results are presented in Table VII.

TABLE VII
RESULTS OF DUNCAN'S NEW MULTIPLE RANGE TEST ON CRITICAL THINKING POSTTEST SCORES FOR GROUPS A, B, C, AND D

<table>
<thead>
<tr>
<th>Groups Compared</th>
<th>Ranked Means</th>
<th>Mean Difference</th>
<th>Range Products 0.01 Level</th>
<th>S or Not S 0.01 Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A with D</td>
<td>63.0 - 49.4</td>
<td>13.6</td>
<td>4.39</td>
<td>S</td>
</tr>
<tr>
<td>A with C</td>
<td>63.0 - 51.6</td>
<td>11.4</td>
<td>4.58</td>
<td>S</td>
</tr>
<tr>
<td>A with B</td>
<td>63.0 - 59.0</td>
<td>4.0</td>
<td>3.70</td>
<td>S</td>
</tr>
<tr>
<td>B with D</td>
<td>59.0 - 49.4</td>
<td>9.6</td>
<td>3.95</td>
<td>S</td>
</tr>
<tr>
<td>B with C</td>
<td>59.0 - 51.6</td>
<td>7.4</td>
<td>4.12</td>
<td>S</td>
</tr>
<tr>
<td>C with D</td>
<td>51.6 - 49.4</td>
<td>2.2</td>
<td>4.47</td>
<td>NS</td>
</tr>
</tbody>
</table>
In Table VII, if the mean difference exceeds the range product value, the means are significantly different at the 0.01 level. The data processing read-out sheet furnished data for both the 0.05 and 0.01 levels of significance. The 0.05 level was arbitrarily selected as the criterion for rejecting or retaining hypotheses, but since all comparisons which showed a significant difference at the 0.05 level also showed a significant difference at the 0.01 level, the data for the higher level was reported in Table VII.

Hypothesis 2c stipulated that the experimental group subjected to both pretest and posttest would score significantly higher than the control group subjected to posttest only. From Table VII, the mean difference for the comparison of group A with group D is 13.6 and the range product is 4.39. Hypothesis 2c, as stated directionally, is therefore accepted. There was a significant difference between the mean posttest scores of the two groups. The mean of group A was significantly greater at the 0.01 level.

Hypothesis 2d stipulated that the experimental group subjected to posttest only would score significantly higher than the control group subjected to both pretest and posttest. The mean difference for the comparison of group B with group C is 7.4 and the range product is 4.12. Hypothesis 2d, as stated directionally, is therefore accepted. There was a significant difference between the mean posttest scores of the two groups. The mean of group B was significantly greater at the 0.01 level.
Hypothesis 2e stipulated that the experimental group subjected to posttest only would score significantly higher than the control group subjected to posttest only. The mean difference for the comparison of group B with group D is 9.6 and the range product is 3.95. Hypothesis 2e, as stated directionally, is therefore accepted. There was a significant difference between the mean posttest scores of the two groups. The mean of group B was significantly greater at the 0.01 level.

Hypothesis 3a stipulated that there would be no significant difference between the mean posttest score of the experimental group subjected to both pretest and posttest and that of the experimental group subjected to posttest only. The mean difference for the comparison of group A with group B is 4.0 and the range product is 3.70. Hypothesis 3a, in the null form, is therefore rejected. There was a significant difference between the mean posttest scores of the two groups. The mean of the group which received both pretest and posttest was significantly greater at the 0.01 level.

Hypothesis 3b stipulated that there would be no significant difference between the mean posttest scores of the control group subjected to both pretest and posttest and the control group subjected to posttest only. As shown in Table VII, the mean difference for the comparison of group C with group D is 2.2 and the product range is 4.47 at the 0.01 level. At the 0.05 level, the mean difference was 2.2 and the range
product was 3.40. Hypothesis 3b is therefore retained. The difference in the mean scores of groups C and D was not significant at the 0.05 level.

Attitude

The instrument used to measure attitude was a semantic differential. The semantic differential included ten concepts. Ten scales, each composed of a seven-step continuum between a pair of bipolar adjectives, were included on each concept. Each scale permitted the respondent to rate the concept on a seven-point spread between the positive adjective and the related negative adjective. Six of the ten scales were selected to measure attitude. Only the scores on these six scales were utilized. The remaining four scales were included to help obscure the purpose of measurement and were not scored. With a range from one to seven on each scale and six scales on each concept, the possible score for each concept had a range from six to forty-two. When scores on each concept were summed, the total score on the entire test had a possible range from 60 to 420. A sample sheet showing one concept from the semantic differential may be found in Appendix A.

Hypotheses 4a through 4e and 5a and 5b pertain to attitude as measured by composite scores on the semantic differential. The hypotheses from Chapter I are restated as follows:
Hypothesis 4: Among certain groups, there will be a significant difference in attitude as reflected by the mean of the posttest composite scores on the ten concepts included on the semantic differential as follows:

a. The combined experimental groups will show a significantly more positive attitude than the combined control groups.

b. The experimental group subjected to both pretest and posttest will show a significantly more positive attitude than the control group subjected to both pretest and posttest.

c. The experimental group subjected to both pretest and posttest will show a significantly more positive attitude than the control group subjected to posttest only.

d. The experimental group subjected to posttest only will show a significantly more positive attitude than the control group subjected to both pretest and posttest.

e. The experimental group subjected to posttest only will show a significantly more positive attitude than the control group subjected to posttest only.

Hypothesis 5: Among certain groups, there will be no significant difference in attitude as reflected by the mean of the posttest composite scores on the ten concepts included on the semantic differential test as follows:
a. The experimental group subjected to both pre-test and posttest and the experimental group subjected to posttest only.

b. The control group subjected to both pretest and posttest and the control group subjected to posttest only.

Hypothesis 4a was tested by performing a t-test on the mean composite posttest scores of the combined experimental and combined control groups. The results are presented in Table VIII.

**TABLE VIII**

RESULTS OF t-TEST ON COMPOSITE ATTITUDE POSTTEST MEANS OF COMBINED EXPERIMENTAL AND COMBINED CONTROL GROUPS

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>418</td>
<td>333.9</td>
<td>39.3</td>
<td>7.106</td>
<td>0.0001</td>
</tr>
<tr>
<td>Control</td>
<td>296</td>
<td>310.0</td>
<td>50.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results show $t = 7.106$, and $p = 0.0001 < 0.05$. Hypothesis 4a, as stated directionally, is therefore accepted. There was a significant difference between the mean posttest score of the combined experimental groups and that of the combined control groups. The mean score of the combined experimental groups was significantly greater beyond the 0.01 level.
Hypotheses 4b through 4e and 5a and 5b were tested by performing one-way analysis of variance on the composite attitude posttest scores. The results are shown in Table IX.

**TABLE IX**

**ANALYSIS OF VARIANCE ON SEMANTIC DIFFERENTIAL COMPOSITE POSTTEST SCORES OF GROUPS A, B, C, AND D**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Variance Estimate</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>102976</td>
<td>3</td>
<td>34325</td>
<td>17.63</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>1382463</td>
<td>710</td>
<td>1947</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1485439</td>
<td>713</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since analysis of variance showed a difference among post-test means of the four groups which was significant beyond the 0.05 level ($F = 17.6, p = 0.0001$), Duncan's New Multiple Range Test was utilized to identify the groups for which the difference in means was significant. The results are presented in Table X.

In Table X, if the mean difference for a particular comparison exceeds the range product value, the means are significantly different. All comparisons which showed a significant difference at the 0.05 level also showed a significant difference at the 0.01 level; therefore, the data for the higher level was reported in Table X.

Hypothesis 4b stipulated that the experimental group subjected to both pretest and posttest would show a
TABLE X

RESULTS OF DUNCAN'S NEW MULTIPLE RANGE TEST ON SEMANTIC DIFFERENTIAL COMPOSITE POSTTEST SCORES FOR GROUPS A, B, C, AND D

<table>
<thead>
<tr>
<th>Groups Compared</th>
<th>Ranked Means</th>
<th>Mean Difference</th>
<th>Range Products 0.01 Level</th>
<th>S or Not S 0.01 Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A with C</td>
<td>334.3 - 305.9</td>
<td>28.4</td>
<td>13.7</td>
<td>S</td>
</tr>
<tr>
<td>A with D</td>
<td>334.3 - 313.8</td>
<td>20.5</td>
<td>12.9</td>
<td>S</td>
</tr>
<tr>
<td>A with B</td>
<td>334.3 - 333.5</td>
<td>0.7</td>
<td>11.2</td>
<td>NS</td>
</tr>
<tr>
<td>B with C</td>
<td>333.5 - 305.9</td>
<td>27.6</td>
<td>12.6</td>
<td>S</td>
</tr>
<tr>
<td>B with D</td>
<td>333.5 - 313.8</td>
<td>19.8</td>
<td>11.8</td>
<td>S</td>
</tr>
<tr>
<td>D with C</td>
<td>313.8 - 305.9</td>
<td>7.8</td>
<td>13.2</td>
<td>NS</td>
</tr>
</tbody>
</table>

significantly more positive attitude than the control group subjected to both pretest and posttest. From Table X, the mean difference for the comparison of group A with group C is 28.4 and the range product is 13.7. Hypothesis 4b, as stated directionally, is therefore accepted. There was a significant difference between the mean posttest scores of the two groups. The mean posttest score of group A was significantly more positive at the 0.01 level.

Hypothesis 4c stipulated that the experimental group subjected to both pretest and posttest would show a significantly more positive attitude than the control group subjected to posttest only. From Table X, the mean difference for the comparison of group A with group D is 20.5 and the range
product is 12.9. Hypothesis 4c, as stated directionally, is therefore accepted. There was a significant difference between the mean posttest scores of the two groups. The mean posttest score of the experimental group was significantly more positive at the 0.01 level.

Hypothesis 4d stipulated that the experimental group subjected to posttest only would show a significantly more positive attitude than the control group subjected to both pretest and posttest. The mean difference for the comparison of group B with group C is 27.6 and the range product is 12.6. Hypothesis 4d, as stated directionally, is therefore accepted. There was a significant difference between the mean posttest scores of the two groups. The mean posttest score of group B was significantly more positive at the 0.01 level.

Hypothesis 4e stipulated that the experimental group subjected to posttest only would show a significantly more positive attitude than the control group subjected to posttest only. The mean difference for the comparison of group B with group D is 19.8, and the range product is 11.8. Hypothesis 4e, as stated directionally, is therefore accepted. There was a significant difference between the mean posttest scores of the two groups. The mean of group B was significantly more positive at the 0.01 level.

Hypothesis 5a stipulated that there would be no significant difference between the posttest mean attitude score of the experimental group which received both pretest and posttest
and that of the experimental group which received posttest only. For the comparison of group A with group B, the mean difference is 0.7 and the range product is 11.2 at the 0.01 level. The corresponding values at the 0.05 level, which are not shown in the table, were 0.7 and 8.5 respectively. At both levels the mean difference is less than the range product. Hypothesis 5a, in the null form, is therefore retained. The difference between the means of the two groups was not significant at the 0.05 level.

Hypothesis 5b stipulated that there would be no significant difference between the posttest mean attitude score of the control group which received both pretest and posttest and that of the control group subjected to posttest only. The mean difference for the comparison of group C with group D is 7.8 and the range product is 13.2 at the 0.01 level. Corresponding values at the 0.05 level were 7.8 and 10.1 respectively. At both levels the range product exceeds the mean difference. Hypothesis 5b, in the null form, is therefore retained. The difference between the means of the two groups was not significant at the 0.05 level.

Hypotheses 6a through 6j required statistical analysis of posttest scores for each individual concept. The mean individual concept score of groups A and B combined and that of groups C and D combined were subjected to $t$-test. The results are shown in Table XI.
TABLE XI

**t-TEST ON SEMANTIC DIFFERENTIAL INDIVIDUAL CONCEPT POSTTEST MEANS OF THE COMBINED EXPERIMENTAL AND COMBINED CONTROL GROUPS**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Experimental N = 418</th>
<th>Control N = 296</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD*</td>
<td>Mean</td>
<td>SD*</td>
</tr>
<tr>
<td>Policemen</td>
<td>32.4</td>
<td>6.38</td>
<td>29.7</td>
<td>7.68</td>
</tr>
<tr>
<td>Courts of law</td>
<td>32.6</td>
<td>5.66</td>
<td>30.4</td>
<td>6.77</td>
</tr>
<tr>
<td>Constitutional rights</td>
<td>38.1</td>
<td>4.86</td>
<td>34.3</td>
<td>6.80</td>
</tr>
<tr>
<td>United States Supreme Court</td>
<td>34.3</td>
<td>5.56</td>
<td>31.4</td>
<td>6.87</td>
</tr>
<tr>
<td>Trial by jury</td>
<td>34.6</td>
<td>5.91</td>
<td>31.4</td>
<td>7.55</td>
</tr>
<tr>
<td>United States Constitution</td>
<td>37.2</td>
<td>4.81</td>
<td>34.1</td>
<td>6.39</td>
</tr>
<tr>
<td>Majority rule</td>
<td>30.6</td>
<td>7.64</td>
<td>28.9</td>
<td>8.15</td>
</tr>
<tr>
<td>The Bill of Rights</td>
<td>37.6</td>
<td>4.87</td>
<td>34.2</td>
<td>6.88</td>
</tr>
<tr>
<td>Political parties</td>
<td>25.5</td>
<td>9.04</td>
<td>25.8</td>
<td>8.23</td>
</tr>
<tr>
<td>Minority rights</td>
<td>31.0</td>
<td>7.54</td>
<td>29.9</td>
<td>7.90</td>
</tr>
</tbody>
</table>

*SD* = Standard deviation

Hypotheses 6a through 6j were stated in Chapter I as follows:

Hypothesis 6: As measured by the semantic differential, the group composed of all experimental subjects will show a significantly more positive mean posttest score than the group composed of all control subjects on:
a. The concept "policemen"
b. The concept "courts of law"
c. The concept "constitutional rights"
d. The concept "United States Supreme Court"
e. The concept "trial by jury"
f. The concept "the United States Constitution"
g. The concept "majority rule"
h. The concept "the Bill of Rights"
i. The concept "political parties"
j. The concept "minority rights"

The data in Table XI show that the value of $t$ for each of the first eight concepts is significant at a point beyond the 0.05 level. Hypotheses 6a through 6h, as stated directionally, are therefore accepted. The difference between the posttest means of the two groups was significant. For each concept "a" through "h," the mean score of the experimental group was significantly more positive at the 0.01 level.

For the concept "political parties," $t = -0.55$ and $p = 0.5892$. For the concept "minority rights," $t = 1.89$ and $p = 0.0553$. Hypotheses 6i and 6j, as stated directionally, are therefore rejected at the 0.05 level. Mean posttest scores of the experimental group were not more positive at the selected level of significance.

Hypotheses 7a through 12j were tested by analysis of variance. The results are shown in Table XII.
<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Variance Estimate</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept a:</td>
<td>1811</td>
<td>3</td>
<td>603.8</td>
<td>12.65</td>
<td>0.0001</td>
</tr>
<tr>
<td>Between</td>
<td>33896</td>
<td>710</td>
<td>47.7</td>
<td>12.65</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>35707</td>
<td>713</td>
<td>710.8</td>
<td>47.7</td>
<td>12.65</td>
</tr>
<tr>
<td>Total</td>
<td>39720</td>
<td>713</td>
<td>710.8</td>
<td>12.65</td>
<td>0.0001</td>
</tr>
<tr>
<td>Concept b:</td>
<td>972</td>
<td>3</td>
<td>323.9</td>
<td>8.57</td>
<td>0.0001</td>
</tr>
<tr>
<td>Between</td>
<td>26820</td>
<td>710</td>
<td>37.8</td>
<td>8.57</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>27792</td>
<td>713</td>
<td>713.9</td>
<td>37.8</td>
<td>8.57</td>
</tr>
<tr>
<td>Total</td>
<td>39572</td>
<td>713</td>
<td>713.9</td>
<td>8.57</td>
<td>0.0001</td>
</tr>
<tr>
<td>Concept c:</td>
<td>2714</td>
<td>3</td>
<td>904.5</td>
<td>27.65</td>
<td>0.0001</td>
</tr>
<tr>
<td>Between</td>
<td>23228</td>
<td>710</td>
<td>32.7</td>
<td>27.65</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>25942</td>
<td>713</td>
<td>713.9</td>
<td>32.7</td>
<td>27.65</td>
</tr>
<tr>
<td>Total</td>
<td>39662</td>
<td>713</td>
<td>713.9</td>
<td>27.65</td>
<td>0.0001</td>
</tr>
<tr>
<td>Concept d:</td>
<td>1461</td>
<td>3</td>
<td>487.1</td>
<td>12.92</td>
<td>0.0001</td>
</tr>
<tr>
<td>Between</td>
<td>26771</td>
<td>710</td>
<td>37.7</td>
<td>12.92</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>28232</td>
<td>713</td>
<td>713.9</td>
<td>37.7</td>
<td>12.92</td>
</tr>
<tr>
<td>Total</td>
<td>39992</td>
<td>713</td>
<td>713.9</td>
<td>12.92</td>
<td>0.0001</td>
</tr>
<tr>
<td>Concept e:</td>
<td>1864</td>
<td>3</td>
<td>621.3</td>
<td>13.91</td>
<td>0.0001</td>
</tr>
<tr>
<td>Between</td>
<td>21711</td>
<td>710</td>
<td>44.7</td>
<td>13.91</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>23575</td>
<td>713</td>
<td>713.9</td>
<td>44.7</td>
<td>13.91</td>
</tr>
<tr>
<td>Total</td>
<td>38986</td>
<td>713</td>
<td>713.9</td>
<td>13.91</td>
<td>0.0001</td>
</tr>
<tr>
<td>Concept f:</td>
<td>1797</td>
<td>3</td>
<td>599.0</td>
<td>19.72</td>
<td>0.0001</td>
</tr>
<tr>
<td>Between</td>
<td>21567</td>
<td>710</td>
<td>30.4</td>
<td>19.72</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>23364</td>
<td>713</td>
<td>713.9</td>
<td>30.4</td>
<td>19.72</td>
</tr>
<tr>
<td>Total</td>
<td>39424</td>
<td>713</td>
<td>713.9</td>
<td>19.72</td>
<td>0.0001</td>
</tr>
<tr>
<td>Concept g:</td>
<td>742</td>
<td>3</td>
<td>247.3</td>
<td>4.02</td>
<td>0.0077</td>
</tr>
<tr>
<td>Between</td>
<td>43677</td>
<td>710</td>
<td>61.5</td>
<td>4.02</td>
<td>0.0077</td>
</tr>
<tr>
<td>Within</td>
<td>44420</td>
<td>713</td>
<td>713.9</td>
<td>61.5</td>
<td>4.02</td>
</tr>
<tr>
<td>Total</td>
<td>49997</td>
<td>713</td>
<td>713.9</td>
<td>4.02</td>
<td>0.0077</td>
</tr>
<tr>
<td>Concept h:</td>
<td>2120</td>
<td>3</td>
<td>706.6</td>
<td>21.09</td>
<td>0.0001</td>
</tr>
<tr>
<td>Between</td>
<td>23783</td>
<td>710</td>
<td>33.5</td>
<td>21.09</td>
<td>0.0001</td>
</tr>
<tr>
<td>Within</td>
<td>25903</td>
<td>713</td>
<td>713.9</td>
<td>33.5</td>
<td>21.09</td>
</tr>
<tr>
<td>Total</td>
<td>54686</td>
<td>713</td>
<td>713.9</td>
<td>21.09</td>
<td>0.0001</td>
</tr>
<tr>
<td>Concept i:</td>
<td>131</td>
<td>3</td>
<td>43.6</td>
<td>0.57</td>
<td>0.6370</td>
</tr>
<tr>
<td>Between</td>
<td>54038</td>
<td>710</td>
<td>76.1</td>
<td>0.57</td>
<td>0.6370</td>
</tr>
<tr>
<td>Within</td>
<td>54169</td>
<td>713</td>
<td>713.9</td>
<td>76.1</td>
<td>0.57</td>
</tr>
<tr>
<td>Total</td>
<td>56449</td>
<td>713</td>
<td>713.9</td>
<td>0.57</td>
<td>0.6370</td>
</tr>
<tr>
<td>Concept j:</td>
<td>272</td>
<td>3</td>
<td>90.6</td>
<td>1.53</td>
<td>0.2041</td>
</tr>
<tr>
<td>Between</td>
<td>42046</td>
<td>710</td>
<td>59.2</td>
<td>1.53</td>
<td>0.2041</td>
</tr>
<tr>
<td>Within</td>
<td>42318</td>
<td>713</td>
<td>713.9</td>
<td>59.2</td>
<td>1.53</td>
</tr>
<tr>
<td>Total</td>
<td>45046</td>
<td>713</td>
<td>713.9</td>
<td>1.53</td>
<td>0.2041</td>
</tr>
</tbody>
</table>
Inspection of the data in Table XII reveals only two concepts for which the groups A, B, C, and D did not show a significant difference between posttest means. For concept "i," $F = 0.57$ and $p = 0.6370 > 0.05$; for concept "j," $F = 1.53$ and $p = 0.204 > 0.05$. Hypotheses 6i, 6j, 7i, 7j, 8i, 8j, 9i, 9j, 10i, and 10j, as stated directionally, are therefore rejected. Hypotheses 11i, 11j, 12i, and 12j, as stated in the null form, are retained. The mean concept score of the four groups under consideration showed no significant difference on the concepts "political parties" and "minority rights."

For all other concepts, the values of $p$ were less than 0.05 indicating $F$ as significant at the 0.05 level. Duncan's New Multiple Range Test was performed on the data for these concepts to identify those groups which differed significantly. The results are shown in Table XIII.

Hypotheses 7a through 7j stipulated that on each concept the mean attitude posttest score of the experimental group subjected to both pretest and posttest would be significantly more positive than that of the control group which received both pretest and posttest. As stated previously, hypotheses 7i and 7j were rejected. No significant difference was found between any of the four groups on these particular concepts. The results of Duncan's New Multiple Range Test, shown in Table XIII, reveal that group A showed a significantly more positive score than group C on each of the remaining concepts. Hypotheses 7a through 7h, as stated directionally, are
### TABLE XIII

RESULTS OF DUNCAN'S NEW MULTIPLE RANGE TEST
ON INDIVIDUAL CONCEPT POSTTEST SCORES
FOR GROUPS A, B, C, AND D

<table>
<thead>
<tr>
<th>Concept</th>
<th>Ranked Means</th>
<th>Mean Difference</th>
<th>Range Products</th>
<th>S or Not S 0.05 Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B - C</td>
<td>33.1 - 28.8</td>
<td>4.3</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>B - D</td>
<td>33.1 - 30.5</td>
<td>2.6</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>B - A</td>
<td>33.1 - 31.4</td>
<td>1.7</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>A - C</td>
<td>31.4 - 28.8</td>
<td>2.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>A - D</td>
<td>31.4 - 30.5</td>
<td>0.9</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>D - C</td>
<td>30.5 - 28.8</td>
<td>1.7</td>
<td>1.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept</th>
<th>A - C</th>
<th>32.8 - 29.8</th>
<th>3.0</th>
<th>1.5</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A - D</td>
<td>32.8 - 30.9</td>
<td>1.9</td>
<td>1.4</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>A - B</td>
<td>32.8 - 32.5</td>
<td>0.3</td>
<td>1.2</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>B - C</td>
<td>32.5 - 29.8</td>
<td>2.7</td>
<td>1.4</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>B - D</td>
<td>32.5 - 30.9</td>
<td>1.6</td>
<td>1.3</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>D - C</td>
<td>30.9 - 29.8</td>
<td>1.1</td>
<td>1.4</td>
<td>NS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept</th>
<th>A - C</th>
<th>38.6 - 33.5</th>
<th>5.1</th>
<th>1.4</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A - D</td>
<td>38.6 - 35.1</td>
<td>3.5</td>
<td>1.3</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>A - B</td>
<td>38.6 - 37.7</td>
<td>0.9</td>
<td>1.1</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>B - C</td>
<td>37.7 - 33.5</td>
<td>4.2</td>
<td>1.3</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>B - D</td>
<td>37.7 - 35.1</td>
<td>2.6</td>
<td>1.6</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>D - C</td>
<td>35.1 - 33.5</td>
<td>1.6</td>
<td>1.3</td>
<td>S</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept</th>
<th>A - C</th>
<th>34.3 - 31.1</th>
<th>3.2</th>
<th>1.5</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A - D</td>
<td>34.3 - 31.6</td>
<td>2.7</td>
<td>1.4</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>A - B</td>
<td>34.3 - 34.2</td>
<td>0.1</td>
<td>1.2</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>B - C</td>
<td>34.2 - 31.1</td>
<td>3.1</td>
<td>1.3</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>B - D</td>
<td>34.2 - 31.6</td>
<td>2.6</td>
<td>1.2</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>D - C</td>
<td>31.6 - 31.1</td>
<td>0.5</td>
<td>1.4</td>
<td>NS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept</th>
<th>A - C</th>
<th>34.8 - 30.8</th>
<th>4.0</th>
<th>1.6</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A - D</td>
<td>34.8 - 32.0</td>
<td>2.8</td>
<td>1.5</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>A - B</td>
<td>34.8 - 34.4</td>
<td>0.4</td>
<td>1.3</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>B - C</td>
<td>34.4 - 30.8</td>
<td>2.6</td>
<td>1.5</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>B - D</td>
<td>34.4 - 32.0</td>
<td>2.4</td>
<td>1.4</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>D - C</td>
<td>32.0 - 30.8</td>
<td>1.2</td>
<td>1.5</td>
<td>NS</td>
</tr>
</tbody>
</table>
**TABLE XIII—Continued**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Ranked Means</th>
<th>Mean Difference</th>
<th>Range Products 0.05 Level</th>
<th>0.05 Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Concept f:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B - C</td>
<td>37.3 - 33.4</td>
<td>3.9</td>
<td>1.3</td>
<td>S</td>
</tr>
<tr>
<td>B - D</td>
<td>37.3 - 34.7</td>
<td>2.6</td>
<td>1.2</td>
<td>S</td>
</tr>
<tr>
<td>B - A</td>
<td>37.3 - 37.1</td>
<td>0.2</td>
<td>1.1</td>
<td>NS</td>
</tr>
<tr>
<td>A - C</td>
<td>37.1 - 33.4</td>
<td>3.7</td>
<td>1.5</td>
<td>S</td>
</tr>
<tr>
<td>A - D</td>
<td>37.1 - 34.7</td>
<td>2.4</td>
<td>1.2</td>
<td>S</td>
</tr>
<tr>
<td>D - C</td>
<td>34.7 - 33.4</td>
<td>1.3*</td>
<td>1.3**</td>
<td>S</td>
</tr>
</tbody>
</table>

| **Concept g:** |              |                |                          |            |
| A - C   | 30.7 - 28.0  | 2.7            | 1.9                      | S          |
| A - D   | 30.7 - 29.7  | 1.0            | 1.8                      | NS         |
| A - B   | 30.7 - 30.5  | 0.2            | 1.5                      | NS         |
| B - C   | 30.5 - 28.0  | 2.5            | 1.7                      | S          |
| B - D   | 30.5 - 29.7  | 0.8            | 1.6                      | NS         |
| D - C   | 29.7 - 28.0  | 1.7            | 1.8                      | NS         |

| **Concept h:** |              |                |                          |            |
| A - C   | 37.8 - 33.8  | 4.0            | 1.4                      | S          |
| A - D   | 37.8 - 34.6  | 3.2            | 1.3                      | S          |
| A - B   | 37.8 - 37.5  | 0.3            | 1.1                      | NS         |
| B - C   | 37.5 - 33.8  | 3.7            | 1.3                      | S          |
| B - D   | 37.5 - 34.6  | 2.9            | 1.2                      | S          |
| D - C   | 34.6 - 33.8  | 0.8            | 1.3                      | NS         |

*1.30286 before rounding off.

**1.25722 before rounding off.

therefore accepted. The scores of the experimental group subjected to both pretest and posttest were significantly more positive at the 0.05 level than those of the control group subjected to both pretest and posttest.

Hypotheses 8a through 8j stipulated that the group of experimental subjects which received both pretest and posttest would show a significantly more positive posttest mean attitude score than the group of control subjects which received posttest only on each of the concepts "a" through "j."
Hypotheses 8i and 8j were rejected on the basis of analysis of variance results. Results of Duncan's New Multiple Range Test show that there was a significant difference in the post-test mean scores of group A and those of group D for the concepts b, c, d, e, f, and h. The differences for concepts a and g were not significant. Hypotheses 8b, 8c, 8d, 8e, 8f, and 8h, as stated directionally, are accepted. The scores of the experimental group were significantly more positive at the 0.05 level. Hypotheses 8a and 8g, as stated directionally, are rejected. The difference in posttest mean scores of group A and group D on the concept "policemen" and the concept "majority rule" did not differ significantly at the 0.05 level.

Hypotheses 9a through 9j stipulated that on each concept the mean posttest score of the experimental group which received posttest only would be significantly more positive than the scores of the control group which received both pre-test and posttest. Inspection of the data in Table XIII shows that group B scored significantly higher than group C on each concept except "i" and "j" which failed to meet the initial analysis of variance test. Hypotheses 9a through 9h, as stated directionally, are therefore accepted. There was a significant difference in the mean posttest scores of group B and group C. The mean score of group B was significantly more positive at the 0.05 level on concepts "a" through "h."

Hypotheses 10a through 10j stipulated that the group of experimental subjects receiving posttest only would score
significantly more positive on each concept than the group of control subjects subjected to posttest only. Table XIII shows that the difference in posttest means of these two groups was not significantly different at the 0.05 level for concept "g." Hypothesis 10g is rejected. The mean score of group B was more positive than that of group D, but the difference was not significant at the 0.05 level. On each of the remaining seven concepts, the differences between the means of groups B and D were significant at the 0.05 level. Therefore hypotheses 10a, 10b, 10c, 10d, 10e, 10f, and 10h are accepted in the form stated. The mean score of group B was significantly more positive at the 0.05 level on these concepts.

Hypotheses 11a through 11j stated that there would be no significant difference between the mean posttest score of the experimental group which received both pretest and posttest and that of the experimental group which received posttest only. Table XIII shows that there was a significant difference between the means of groups A and B on one concept. The mean of group B was significantly more positive than that of group A for the concept "policemen." Hypothesis 11a is therefore rejected. On all other concepts the difference in mean scores of the two groups were not significant at the 0.05 level. In addition to hypotheses 11i and 11j which were previously retained, hypotheses 11b through 11h are also retained.
Hypotheses 12a through 12j stipulated there would be no significant difference between the mean posttest score of group C and that of group D on each concept included on the semantic differential. Table XIII shows that the scores for these two groups differed significantly at the 0.05 level on concepts "a," "c," and "f." For each of these concepts, the score of group D was significantly greater than the score of group C. Hypotheses 12a, 12c, and 12f are therefore rejected. The difference in mean scores on the remaining concepts was not significant at the 0.05 level. Hypotheses 12b, 12d, 12e, 12g, and 12h are therefore retained.
CHAPTER V

SUMMARY AND CONCLUSIONS

Experimental Setting and Design

The purpose of this study was to determine if there is a significant difference between a specific inquiry-type approach and the traditional lecture-recitation-textbook approach to teaching concepts of American law and government as each affects the performance of high school students in the areas of achievement, critical thinking, and attitude.

The setting for this study was provided when a large school district in North Texas adopted some new materials to be used in problems of democracy classes. The specific textual materials adopted were entitled Justice in Urban America. They are a part of the Law in American Society materials, a special social studies curriculum developed through the Chicago Law in American Society Project—a project sponsored jointly by the Chicago Board of Education and the Chicago Bar Association. The materials consist of a series of student booklets and teacher manuals which deal with practical and current problems through the mediums of case-study, mock trials, and simulation. The materials were designed to stress the inquiry and discovery approaches to concepts, principles, and generalizations which have broad application to current social and legal problems.
As a pilot study program, the materials were selected for use by seven teachers in six different high schools representing different geographic areas within the school district. For the purposes of experimental study, six control schools were arbitrarily selected in which the student body matched as closely as possible that of the experimental schools in cultural and socio-economic background and in academic standing.

A total of 458 students were included in the experimental classes. These students constituted eighteen separate classes or sections distributed over six high schools. The eighteen sections were taught by seven teachers who received special training in the instructional techniques required for the experimental materials. The control classes included 343 students in 12 sections or classes distributed over 6 high schools. The 12 sections were taught by 6 teachers who normally are assigned to teach problems of democracy.

Through use of a table of random numbers, nine of the eighteen experimental sections and six of the twelve control sections were selected to receive both pretest and posttest in each of the following areas: achievement, critical thinking, and attitude. The remaining nine experimental sections and six control sections received posttest only. This created a four-group experimental design composed of: A) experimental subjects who received both pretest and posttest, B) experimental subjects who received posttest only, C) control
subjects who received both pretest and posttest, and D) control subjects who received posttest only.

The Principles of Democracy Test, published by Science Research Associates, Inc., Chicago, Illinois, was administered as a measure of achievement. The Watson-Glaser Critical Thinking Appraisal, published by Harcourt, Brace & World, Inc., New York, was used to measure critical thinking performance. The instrument used to measure attitude was a semantic differential especially constructed for this study.

Performance on each of the above instruments was examined statistically by making all possible comparisons between posttest mean scores for each of the four groups or categories of students. In addition, the scores of both experimental groups were combined and the posttest mean compared with the posttest mean of the combined control groups.

Findings Related to Achievement

The statistical technique used to compare achievement scores was analysis of covariance. In the comparison of the experimental group which received both pretest and posttest with the control group which received both pretest and posttest, both I. Q. scores and pretest scores were utilized as covariates. In all other comparisons on each achievement, I. Q. was measured by the California Test of Mental Maturity. The findings, as related to hypotheses 1a through 1g, are as follows:
1. Hypothesis 1a stipulated that there would be no significant difference between the mean posttest achievement score of the combined experimental groups and that of the combined control groups. This hypothesis was rejected. The adjusted mean of the combined control groups was greater at the 0.05 level of significance.

2. Hypothesis 1b stipulated that there would be no significant difference between the mean posttest achievement score of the experimental group which received both pretest and posttest and that of the control group which received both pretest and posttest. This hypothesis was retained. The adjusted mean of the control group was greater, but the difference was not significant at the 0.05 level.

3. Hypothesis 1c stipulated that there would be no significant difference between the mean posttest achievement score of the experimental group which received both pretest and posttest and that of the control group which received posttest only. This hypothesis was retained. The adjusted mean of the experimental group was greater, but the difference was not significant at the 0.05 level.

4. Hypothesis 1d stipulated that there would be no significant difference between the mean posttest achievement score of the experimental group which received posttest only and that of the control group which received both pretest and posttest. This hypothesis was rejected. The adjusted mean of the control group was greater at the 0.05 level of significance.
5. Hypothesis 1e stipulated that there would be no significant difference between the mean posttest achievement score of the experimental group which received posttest only and that of the control group which received posttest only. This hypothesis was retained. The adjusted mean of the control group was greater, but the difference was not significant at the 0.05 level.

6. Hypothesis 1f stipulated that there would be no significant difference between the mean posttest achievement score of the experimental group which received both pretest and posttest and that of the experimental group which received posttest only. This hypothesis was retained. The adjusted mean of the experimental group which received both pretest and posttest was greater, but the difference was not significant at the 0.05 level.

7. Hypothesis 1g stipulated that there would be no significant difference between the mean posttest achievement score of the control group which received both pretest and posttest and that of the control group which received posttest only. This hypothesis was rejected. The adjusted mean of the control group which received both pretest and posttest was greater at the 0.05 level of significance.

From the results discussed above, two trends pertaining to achievement seem to be distinguishable. First, although the results are not completely in favor of the control groups, the traditional type of instruction appears to result in
achievement scores as measured by the Principles of Democracy Test which, when statistically adjusted by using I. Q. scores as the covariate, give larger adjusted group means. This implies a need for evaluation of the content of the instrument in terms of the educational objectives of the course. Second, the groups which received both pretest and posttest in most instances scored higher than the corresponding groups which received posttest only. This suggests the possibility of some interaction between pretest and treatment or between pretest and posttest. It is possible that such interaction could influence posttest scores. This is one plausible explanation of the one case where the adjusted mean of the experimental group exceeded that of the control group.

Findings Related to Critical Thinking

1. Hypothesis 2a stipulated that the mean posttest critical thinking score of the combined experimental groups would be significantly greater than that of the combined control groups. This hypothesis, as stated directionally, was accepted. Results of the t-test showed that the posttest mean of the combined experimental groups was significantly greater beyond the 0.01 level.

2. Hypothesis 2b stipulated that the mean posttest critical thinking score of the experimental group which received both pretest and posttest would be significantly greater than that of the control group which received both
which received both pretest and posttest. This hypothesis, as stated directionally, was accepted. Results of analysis of covariance showed that the adjusted mean of the experimental group was significantly greater beyond the 0.01 level.

3. Hypothesis 2c stipulated that the mean posttest critical thinking score of the experimental group which received both pretest and posttest would be significantly greater than that of the control group which received posttest only. This hypothesis, as stated directionally, was accepted. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly greater.

4. Hypothesis 2d stipulated that the mean posttest critical thinking score of the experimental group which received posttest only would be significantly greater than that of the control group which received both pretest and posttest. This hypothesis, as stated directionally, was accepted. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly greater at the 0.01 level.

5. Hypothesis 2e stipulated that the mean posttest critical thinking score of the experimental group which received posttest only would be significantly greater than that of the control group which received posttest only. This hypothesis was accepted. Results of analysis of variance and
Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly greater at the 0.01 level.

6. Hypothesis 3a stipulated that there would be no significant difference between the mean posttest critical thinking score of the experimental group which received both pretest and posttest and that of the experimental group which received posttest only. This hypothesis was rejected. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the mean of the group which received both pretest and posttest was significantly greater at the 0.01 level.

7. Hypothesis 3b stipulated that there would be no significant difference between the mean posttest critical thinking score of the control group which received both pretest and posttest and that of the control group which received posttest only. This hypothesis, as stated in the null form, was retained. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the mean of the group which received both pretest and posttest was greater, but the difference was not significant at the 0.05 level.

The results from analysis of critical thinking scores thus give evidence that the experimental treatment promoted substantially greater performance in this area than did the traditional lecture-recitation-textbook type of instruction. In each case which involved comparison of an experimental group with a control group, the experimental group scored significantly higher.
As in the case of achievement, there is also evidence that posttest scores tend to be higher when students have received pretest. In this case the difference between the groups receiving both pretest and posttest and those receiving posttest only was not significant, but those who received pretest scored higher giving some indication of a possible pretest-posttest or pretest-treatment interaction.

Findings Related to Attitude

Composite Attitude Scores

Hypotheses 4a through 4e and 5a and 5b dealt with over-all attitude as reflected by the composite scores on the semantic differential. Posttest scores on all ten concepts were summed to give a composite or total score for each respondent. The mean composite scores for each of the four groups were compared for all possible combinations of the individual groups and also for the combined experimental and combined control groups. The findings are as follows:

1. Hypothesis 4a stipulated that the mean posttest composite score of the combined experimental groups would be significantly more positive than that of the combined control groups. This hypothesis, as stated directionally, was accepted. Results of the t-test showed that the posttest mean of the combined experimental group was significantly more positive beyond the 0.01 level.
2. Hypothesis 4b stipulated that the mean posttest composite score of the experimental group which received both pretest and posttest would be significantly more positive than that of the control group which received both pretest and posttest. This hypothesis, as stated directionally, was accepted. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly more positive at the 0.01 level.

3. Hypothesis 4c stipulated that the mean posttest composite score of the experimental group which received both pretest and posttest would be significantly more positive than that of the control group which received posttest only. This hypothesis, as stated directionally, was accepted. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly more positive at the 0.01 level.

4. Hypothesis 4d stipulated that the mean posttest composite score of the experimental group which received posttest only would be significantly more positive than that of the control group which received both pretest and posttest. This hypothesis, as stated directionally, was accepted. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly more positive at the 0.01 level.
5. Hypothesis 4e stipulated that the mean posttest composite score of the experimental group which received posttest only would be significantly more positive than that of the control group which received posttest only. This hypothesis, as stated directionally, was accepted. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly more positive at the 0.01 level.

6. Hypothesis 5a stipulated that there would be no significant difference between the mean posttest composite score of the experimental group which received both pretest and posttest and that of the experimental group which received posttest only. This hypothesis was retained. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the mean of the group which received both pretest and posttest was greater, but the difference was not significant.

7. Hypothesis 5b stipulated that there would be no significant difference between the mean posttest composite score of the control group which received both pretest and posttest and that of the control group which received posttest only. This hypothesis was retained. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the mean of the group which received posttest only was greater, but the difference was not significant.
Individual Concept Scores

1. Hypotheses 6a through 6j stipulated that the mean posttest score of the combined experimental groups would be significantly more positive than that of the combined control groups for each of the ten concepts on the semantic differential. Results of the t-test showed that the posttest mean of the combined experimental groups was significantly more positive on the following eight concepts: policemen, courts of law, constitutional rights, United States Supreme Court, trial by jury, United States Constitution, majority rule, and the Bill of Rights. Hypotheses 6a through 6h, as stated directionally, were therefore accepted. There was no significant difference between posttest means on two concepts—political parties and minority rights. Hypotheses 6i and 6j, as stated directionally, were therefore rejected.

2. Hypotheses 7a through 7j stipulated that the mean posttest score of the experimental group which received both pretest and posttest would be significantly more positive than that of the control group which received both pretest and posttest for each of the ten concepts on the semantic differential. Results of analysis of variance and Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly more positive on the same eight concepts as those listed in hypothesis six above. Hypotheses 7a through 7h, as stated directionally, were therefore accepted. On two concepts—political parties and minority
rights—the difference was not significant. Hypotheses 7i and 7j were therefore rejected.

3. Hypotheses 8a through 8j stipulated that the mean posttest score of the experimental group which received both pretest and posttest would be significantly more positive than that of the control group which received posttest only for each of the ten concepts on the semantic differential. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly more positive on the following six concepts: courts of law, constitutional rights, United States Supreme court, trial by jury, United States Constitution, and the Bill of Rights. Hypotheses 8b, 8c, 8e, and 8h, as stated directionally, were therefore accepted. The differences between posttest means were not significant on four concepts: policemen; majority rule; political parties; and minority rights. Hypotheses 8a, 8g, 8i, and 8j, as stated directionally, were therefore rejected.

4. Hypotheses 9a through 9j stipulated that the mean posttest score of the experimental group which received posttest only would be significantly more positive than that of the control group which received both pretest and posttest for each of the ten concepts on the semantic differential. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly more positive on the
following eight concepts: policemen, courts of law, constitutional rights, United States Supreme Court, trial by jury, United States Constitution, majority rule, and the Bill of Rights. Hypotheses 9a through 9h, as stated directionally, were therefore accepted. The difference between mean posttest scores were not significant on two concepts—political parties and minority rights. Hypotheses 9i through 9j, as stated directionally, were therefore rejected.

5. Hypotheses 10a through 10j stipulated that the mean posttest score of the experimental group which received posttest only would be significantly more positive than that of the control group which received posttest only for each of the ten concepts on the semantic differential. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that the posttest mean of the experimental group was significantly more positive on the following seven concepts: policemen, courts of law, constitutional rights, United States Supreme Court, trial by jury, United States Constitution, and the Bill of Rights. Hypotheses 10a, 10b, 10c, 10d, 10e, 10f, and 10h, as stated directionally, were therefore accepted. The differences were not significant on three concepts: majority rule, political parties, and minority rights. Hypotheses 10g, 10i, and 10j, as stated directionally, were therefore rejected.

6. Hypotheses 11a through 11j stipulated that there would be no significant difference between the mean posttest
score of the experimental group which received both pretest and posttest and that of the experimental group which received posttest only for each of the ten concepts on the semantic differential. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that there was no significant difference on the following nine concepts: courts of law, constitutional rights, United States Supreme Court, trial by jury, United States Constitution, majority rule, the Bill of Rights, political parties, and minority rights. Hypotheses 11b through 11j, in the null form, were therefore retained. On one concept—policemen—the mean of the experimental group which received posttest only was significantly greater than that of the experimental group which received both pretest and posttest. Hypothesis 11a was therefore rejected.

7. Hypotheses 12a through 12j stipulated that there would be no significant difference between the mean posttest score of the control group which received both pretest and posttest and that of the control group which received posttest only for each of the ten concepts on semantic differential. Results of analysis of variance followed by Duncan's New Multiple Range Test showed that there was no significant difference on the following concepts: courts of law, United States Supreme Court, trial by jury, majority rule, the Bill of Rights, political parties, and minority rights. Hypotheses 12b, 12d, 12e, 12g, 12h, 12i, and 12j, in the null form, were therefore retained. On three concepts—policemen, constitutional rights, and the
United States Constitution—the means of the control group which received posttest only were significantly more positive than those of the control group which received both pretest and posttest. Hypotheses 12a, 12c, and 12f were therefore rejected.

In summary, the results give evidence that highly favor the experimental treatment as producing more positive attitude scores. This was true for the composite or total score on the semantic differential, and was also true in eight out of ten cases when scores in individual concepts were analyzed.

Conclusions

The teacher variable was not controlled in this study. Teachers assigned to the experimental sections received special training and also had considerably more professional training and experience than the teachers assigned to control groups. In addition, the selection of experimental and control samples was not random. Experimental and control sections were selected arbitrarily, and the students involved were selected because they enrolled in these sections through normal enrollment procedures. This procedure perhaps lends some degree of randomness, and the use of a covariate helps to improve equivalency of groups. However, the assumption of complete equivalency between groups is not warranted.

For these reasons, conclusions are limited to the results and conditions of this study. Although the results may
have implications for other populations and for merits of the experimental curriculum investigated, conclusions and inferences are limited to the specific program which was studied.

Based upon the results of this study, the following may be concluded:

1. As measured by the Principles of Democracy Test, the traditional type of instruction, in general, results in higher achievement scores among high school students.

2. As measured by the Watson-Glaser Critical Thinking Appraisal, the experimental instruction results in higher critical thinking performance among high school students.

3. The experimental instruction results in a more positive attitude among high school students toward eight of the ten concepts included on the semantic differential.

4. When the total score on the semantic differential is considered as a measure of attitude, the experimental instruction results in a more positive attitude on the part of high school students.

Recommendations

Based upon the results obtained from this study, the following recommendations are proposed for school systems which are planning to adopt new instructional programs for use in social studies classes:

1. Adoption of the experimental instructional program investigated in this study is recommended for school systems
which desire to emphasize the development of critical thinking and a positive attitude toward those concepts included on the semantic differential. The evidence favoring traditional instruction over the experimental program in the area of achievement is less conclusive than the evidence favoring the experimental program in the areas of critical thinking and attitude. If adoption of the experimental program should result in less subject matter knowledge on the part of the student, it is likely that the loss would be compensated for by improvement in the areas of critical thinking and attitude.

2. It is recommended that educational objectives relative to achievement be clearly defined and an instrument chosen which will measure the degree to which the objectives are obtained. Problems of democracy curricula are by nature relatively flexible, and the subject matter content in traditional courses varies considerably. When new content and a new approach are introduced, the problem is compounded. Selection of an achievement test which is equally fair to both experimental and control groups becomes very difficult. The type of knowledge or achievement desired is perhaps the fundamental question. Agreement on this type of question is not always easy to attain. If knowledge of subject matter content is deemed worthy of measure, then the instrument for measurement should be carefully selected.

In situations similar to that encountered in this study, a plausible solution might be to construct an instrument.
A large number of test items might be collected from teachers of both the experimental and control materials and an instrument constructed by randomly selecting an equal number of items from each group. With equal numbers of test items coming from both types of materials, the penalization of students in one or the other of the groups should be diminished. Validity and reliability of the instrument could be established by administration to a pilot study group; or, in cases where the instrument is to be used in successive years, by performing the necessary statistical treatment on scores from the first administration of the test.

3. It is further recommended that statistical evaluation of new programs, if possible, include an evaluation performed during the Fall semester. If the program necessitates administration of test near the end of the Spring semester, the test should be administered as early as feasible. This recommendation is based upon the fact that in this study, a high rate of absences near the end of the Spring semester resulted in missing test data on a substantial number of students. If within a particular school or school system the absence rate is not great near the end of the semester, this recommendation may not be applicable.

4. Whenever feasible, control classes should be conducted within the same school as the experimental classes with students being assigned to the control or experimental sections in a random manner.
APPENDIX A

SAMPLE PAGE: SEMANTIC DIFFERENTIAL

Rate the following:

Policemen

Awful _____ _____ _____ _____ _____ _____ Nice*
Fast _____ _____ _____ _____ _____ Slow
Important ____ ____ ____ ____ ____ Unimportant*
Bad ____ ____ ____ ____ ____ Good*
Light ____ ____ ____ ____ ____ Heavy
Fair ____ ____ ____ ____ ____ Unfair*
Valuable ____ ____ ____ ____ ____ Worthless
Cruel ____ ____ ____ ____ ____ Kind*
Weak ____ ____ ____ ____ ____ Strong
Passive ____ ____ ____ ____ ____ Active

*Scales scored to measure attitude.
BIBLIOGRAPHY

Books


Articles


Devine, Thomas G., "Can We Teach Critical Thinking?" Elementary English, XLI (February, 1964), 154-156.


Reports


Test Department, Harcourt, Brace & World, Inc., Watson-Glaser Critical Thinking Appraisal References, a selected list of research studies, books, and articles related to the use of the Critical Thinking Appraisal and to the general concepts of critical thinking, New York, Harcourt, Brace & World, Inc., not dated.

Unpublished Materials


