LEVELS OF QUESTIONING USED BY STUDENT TEACHERS AND ITS EFFECT ON PUPIL ACHIEVEMENT AND CRITICAL THINKING ABILITY

APPROVED:

Graduate Committee:

Major Professor

Frank Halstead

Minor Professor

Committee Member

Committee Member

Dean of the College of Education

Dean of the Graduate School

**Problem:** The effect of questioning from student teachers on pupil achievement and critical thinking ability.

**Sub-problem:** The effect of feedback from university supervisor to student teachers on the latters' convergent and divergent questioning.

**Purposes** were to determine the effect of student teachers' questioning on secondary social studies students' achievement as measured by the *Iowa Tests of Educational Development* and the *Sequential Tests of Educational Progress*, and on critical thinking ability as measured by the *Watson-Glaser Critical Thinking Appraisal*; and the effect of feedback to student teachers concerning how they ask convergent and divergent questions.

An experimental and a control group design was used. The experimental group consisted of eight student teachers and 258 public school students, while the control group consisted of eight student teachers and 263 public school students. The student teachers were randomly assigned to the control and experimental groups. Each of them taught in
a metropolitan secondary school. They taught a subject in the social studies.

Three research hypotheses were formulated. The level of significance for acceptance was selected as .05.

Hypotheses One and Two stated that for students in the experimental group, the adjusted group means on two post-tests of achievement and a post-test of critical thinking ability would be significantly higher than those for students in the control group. Pupils were administered pre- and post-tests. Between testing, experimental student teachers received twelve hours of training in questioning tactics. Questioning patterns of both groups were coded once weekly for eight weeks, using the Observation Schedule and Record 5V. Experimental student teachers received feedback from codings.

Analysis of covariance tested the hypotheses. No significant difference was found on the ITED or STEP tests. A significant difference existed on the Critical Thinking Appraisal, but it was opposite to the direction hypothesized. Control students were superior in critical thinking ability. Hypotheses One and Two were rejected.

Conclusions: Pupils taught by trained student teachers cannot be expected to show significant gain in social studies achievement or critical thinking ability, as measured by the tests used.
Hypothesis Three stated that for student teachers in the experimental group there would be a significantly greater frequency of high-level, divergent questions asked during the eight weeks between pre- and post-testing. Codings of both groups were subjected to chi-square tests. A significantly greater number of divergent questions were asked by experimental student teachers. Hypothesis Three was accepted.

Conclusion: Through training and feedback, student teachers may increase divergent questions. Total number of questions can increase, intensifying classroom verbal interaction. The experimental student teachers increased their divergent questions, producing a decrement in critical thinking of their pupils.

Recommendations for further research were presented.
LEVELS OF QUESTIONING USED BY STUDENT TEACHERS AND ITS EFFECT ON PUPIL ACHIEVEMENT AND CRITICAL THINKING ABILITY

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

Charles Garrett Beseda, B. F. A., M. ED.
Denton, Texas
August, 1972
TABLE OF CONTENTS

LIST OF TABLES

Chapter

I. INTRODUCTION

Statement of the Problem
Statement of a Sub-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. A REVIEW OF THE LITERATURE AND RELATED RESEARCH

A History of the Development of Research on Classroom Observation Systems
A History of the Development and Modifications of the Observation Schedule and Record (OSoAR) in Research Studies of Teacher Effectiveness
A Summary of the Classification of Types of Questions Asked by Teachers As Identified and Assessed in Observational Studies
A Summary of Research Related to the Use of Questions in Teaching
A Summary of Research Relevant to This Study on the Use of Questions in Teaching
A Review of the Analysis of Questioning Tactics As Reported in Writings Other Than Research Studies
A Review of Some Current Teaching Methods Textbooks Related to Questioning in Teaching and Specialized Training Programs Designed to Change Teachers' Questioning Behavior
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Statement of the Relationship of the Reported Research to This Study</td>
<td>102</td>
</tr>
<tr>
<td>III. METHODS AND PROCEDURES</td>
<td>102</td>
</tr>
<tr>
<td>Selection and Description of the Groups</td>
<td></td>
</tr>
<tr>
<td>Description of Instruments</td>
<td></td>
</tr>
<tr>
<td>Description of the Study</td>
<td></td>
</tr>
<tr>
<td>Description of Training in Questioning Techniques</td>
<td></td>
</tr>
<tr>
<td>Procedures for Treating the Data</td>
<td></td>
</tr>
<tr>
<td>IV. STATISTICAL TREATMENT AND ANALYSIS OF THE DATA</td>
<td>128</td>
</tr>
<tr>
<td>Statistical Treatment</td>
<td></td>
</tr>
<tr>
<td>Analysis of the Data</td>
<td></td>
</tr>
<tr>
<td>V. SUMMARY, FINDINGS, CONCLUSIONS, FACTORS BEARING UPON THE FINDINGS, AND RECOMMENDATIONS FOR FURTHER RESEARCH</td>
<td>148</td>
</tr>
<tr>
<td>Summary, Findings, and Conclusions</td>
<td></td>
</tr>
<tr>
<td>Factors Which Bear Upon the Findings and Recommendations for Further Research</td>
<td></td>
</tr>
<tr>
<td>APPENDIX</td>
<td>166</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>170</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Distribution of Student Teachers to Supervising Teachers by Group and Subject Area</td>
<td>105</td>
</tr>
<tr>
<td>II. Social Studies Students' Criterion and Control Variable Means for the Iowa Tests of Educational Development, Test 5</td>
<td>129</td>
</tr>
<tr>
<td>III. Analysis of Covariance of Experimental and Control Social Studies Students' Achievement Performance on the Iowa Tests of Educational Development, Test 5</td>
<td>130</td>
</tr>
<tr>
<td>IV. Social Studies Students' Criterion and Control Variable Means for the Sequential Tests of Educational Progress--Social Studies</td>
<td>132</td>
</tr>
<tr>
<td>V. Analysis of Covariance of Experimental and Control Social Studies Students' Achievement Performance on the Sequential Tests of Educational Progress--Social Studies</td>
<td>133</td>
</tr>
<tr>
<td>VI. Social Studies Students' Criterion and Control Variable Means for the Watson-Glaser Critical Thinking Appraisal</td>
<td>134</td>
</tr>
<tr>
<td>VII. Analysis of Covariance of Experimental and Control Social Studies Students' Achievement Performance on the Watson-Glaser Critical Thinking Appraisal</td>
<td>135</td>
</tr>
<tr>
<td>VIII. Observed and Expected Frequencies for a One-Sample Chi-Square Analysis of Divergent and Convergent Questions Asked by Student Teachers</td>
<td>137</td>
</tr>
<tr>
<td>IX. Responses of Student Teachers to Questions Concerning the Supervising Teachers' Influence on Their Classroom Questioning Behavior</td>
<td>139</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

The student teaching internship in teacher training is considered by a large majority of the participants to be the most practical and rewarding phase of their education. Effective supervision of student teaching interns is therefore paramount during this experience.

One way to determine what constitutes effective supervision of student teachers and to examine ways of improving this role, as well as that of the teaching act, is to research the functions of the college supervisor and student teachers in the classroom setting.

Certainly there is no more obvious approach to research on teaching than direct observation of the behavior of teachers while they teach and while the pupils learn. Yet, it is a rare study indeed that includes any formal observation at all (20, p. 247).

Review of the literature since Medley and Mitzel made the above statement in 1963 reveals a trend toward more direct and systematic classroom observation (1, 4, 8, 9, 12, 15, 16-25, 27, 29, 32-35, 39). But exactly what kinds of teacher and pupil behavior does the observer look for in the classroom?

Since the 1930’s, when researchers in education became interested in analyzing classroom interaction, the kinds of
behaviors observed range from Anderson's analysis of the integrative-dominative behavior of teachers through the autocratic-democratic dichotomy presented by Lewin, Lippitt, and White, to Withall's classroom climate measured by a category system, Robert Bales' interaction process analysis, Ryans' teacher characteristics, focusing on the teacher as an information-processing system, Flanders' interaction analysis system of teacher's indirect and direct influence (1, pp. 2-3), to Medley's learning environment (22), and currently to various modifications of these approaches. Simon and Boyer in a two-volume anthology of classroom observation instruments report on seventy-nine such systems in use at the time of their publication in 1970 (32), with others appearing almost daily. Neujahr (25, p. 221) indicates that in the last ten years researchers have systematically observed classes from elementary school through college, using observers in the room, photographs, motion pictures, audio tapes, and video tapes.

A look at the procedures and instruments used to find out what kinds of teacher-pupil behaviors result in thinking and learning abilities discloses that many of them place a strong emphasis on teacher questioning techniques. This emphasis is not altogether surprising, because "over the years educators have advocated that one way to stimulate thinking among pupils is the effective use of questions ..." (12, p. 697). Aschner states that "One might even
say the teacher is a professional question maker" (2, p. 44).
Crump believes that "A classroom without questions is as
hard to conceive as Santa Claus without children or Inde-
pendence Day without flags" (7, p. 657). Yet Hunkins says
that "despite effective questioning being advocated and
equated with effective teaching, the past half-century has
provided little empirical research on questions and ques-
tioning" (12, p. 697).

Research into student teachers' levels of questioning,
effects of training in questioning procedures on student
teachers, and how feedback from the supervisors can affect
the types of questions student teachers ask is still needed.
What effect such strategies have on student achievement may
yield answers which will more validly define a necessary
function of college supervisors and add to the knowledge of
how various types of questions affect learning.

Statement of the Problem

The problem of this study was the effect of levels of
questioning, used by secondary social studies student teach-
ers, on pupil achievement and critical thinking ability.

Statement of a Sub-Problem

A sub-problem of this study was the effect of feedback
from the university supervisor to student teachers, on the
latters' use of convergent and divergent levels of ques-
tioning.
Purposes of the Study

The purposes of this study were

1. To determine the effect of levels of questioning used on secondary public school students in social studies, as measured by (a) their achievement scores, and (b) their critical thinking ability;

2. To determine the effect of feedback to student teachers on their patterns of asking convergent and divergent questions, as measured by coding frequencies of each type on an Observation Schedule and Record form;

3. To draw conclusions from the findings and develop implications concerning levels of questioning used by teachers and the use of feedback from college supervisors to student teachers.

Hypotheses

For purposes of this study, the following hypotheses were formulated:

1. For students in the experimental group, the adjusted group means on two post-tests of achievement would be significantly higher than those for students in the control group.

2. For students in the experimental group, the adjusted group mean on a post-test of critical thinking ability would be significantly higher than that for students in the control group.
3. For student teachers in the experimental group, there would be a significantly greater frequency of high-level, divergent questions asked during the eight weeks between pre-testing and post-testing than there would be for student teachers in the control group for the same period.

Background and Significance of the Study

Soar reports that Medley and Mitzel, as late as 1959, after reviewing the status of research which related ratings of teacher effectiveness and learning of pupils, found that results were essentially negative (34, p. 1). He goes on to say in another article, “indeed, almost all the research to that point [1959] could be summarized by the general hypothesis 'nothing makes any difference’” (35, p. 9).

Fortunately, the situation has changed in the ensuing years. Significant advances have been made in systematic classroom observation techniques. In the last decade, research has pointed out some characteristics of effective teaching which are being accepted by educators (4). Central to many of these studies is the role teachers' questions and questioning skills play in the learning process. "Teaching involves questioning. It is virtually impossible to think of teaching (over a period of time) that does not involve questioning. Indeed, questioning is the teacher's chief means of directing or channeling discourse" (13, p. 250). Many studies indicate the frequency of questions used daily
by teachers. About one-third of classroom talk is questioning, and teachers ask about 86 per cent of the questions (3).

At a time when educators are calling for developing in students the ability to "think critically," current research on questioning clearly shows that the predominant emphasis of teachers' questions is on searching out knowledge of facts (2, 7, 8, 10, 11, 13, 15, 36, 37).

Although teachers of most subjects in the curriculum may come under criticism for asking low-level, fact-recalling types of questions, the instructors in the social studies have come under particular attack by researchers.

Questions posed in the social studies classroom for over half a century have been recognized as emphasizing memory as the most important cognitive operation. . . . Yet, during this period, the attention of the social studies has been focused repeatedly on admonitions to foster pupils' critical thinking and, especially in recent years, discovery procedures (8, p. 21).

Davis and Tinsley (8) conducted a study of forty-four secondary student teachers teaching social studies in 1966. They used Sanders' categories (30) of Bloom's Taxonomy in the Cognitive Domain (5) to measure student teachers' levels of questioning. They found that teachers "... asked more "memory" questions than all others combined" (8, p. 23). They concluded that

... courses in ... student teaching could incorporate a component dealing specifically with questions, their cognitive emphases, and candidates' ability to vary their use of questions in classroom discourse (8, pp. 25-26).
In the spring semester of 1967 these same researchers used sixty-seven secondary student teachers to determine the cognitive nature of the questions social studies teachers plan and compose to guide class discussion and to use on tests (37). As in their 1966 study, Tinsley and Davis found

The emphasis on memory and evaluation manifest in the questions by these student teachers suggests a lack of attention to planning for questions that stimulate other high-level cognitive abilities (37, p. 62).

Hunkins (12) recently completed a study in which he sought to determine how various types of questions would affect pupils' thinking. He used analysis and evaluation questions in social studies text-type materials to see if such items would effectively stimulate the development of sixth-grade pupils' critical thinking. He found no significant differences between groups using high-level text-type questions and those using low-level knowledge-type questions. However, since questions were in written form only, and of the textbook type, Hunkins contended that

... pupils might have been affected adversely because of lack of teacher interaction. Pupils are accustomed to the teacher's playing a dominant role in the classroom. Yet, in this investigation, the teacher served only to coordinate the pupils' use of the materials (12, p. 703).

Since asking the right kinds of questions is identified as an important skill for effective teachers and since research indicates "... that most teachers' questions are low on the cognitive-emphasis scale..." (12, p. 697),
justification can be given for conducting further study on levels of questioning as they relate to achievement and critical thinking ability of students. The study reported herein, therefore, was an attempt to provide additional knowledge in the area of inquiry by using high school students and their student teachers as the objects of the feedback procedures.

Definition of Terms

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

1. College Supervisor was the university School of Education faculty member who observed, coded, and worked directly with the student teachers during their internship.

2. Feedback was the information obtained by student teachers from the college supervisor concerning their verbal behavior while teaching, particularly the levels-of-questions of the convergent and divergent types which the experimental student teachers exhibited in the classroom.

3. Learning Environment constituted characteristics of teachers' verbal behavior while interacting with students as defined by and coded on an observation record, OScAR 5V.

4. Levels of Questioning were established by the hierarchy developed by Bloom (5), formulated by Sanders (30), and modified by Davis and Tinsley (8) and identified as the basis for categorizing questions asked by student teachers.
The questions were dichotomized, with Level One, memory questions, classified as convergent. Other substantive types of questions requiring mental processes on the part of the students above the recall level were classified as divergent. This dichotomy provided for more precise coding on the OSCAR 5V, a description of which appears on pages ten and eleven.

5. Student Achievement was defined as scores made by public school students on two standardized social studies tests, the Iowa Tests of Educational Development, Test 5 and the Sequential Tests of Educational Progress—Social Studies.

6. Student Critical Thinking Ability was defined as scores made by public school students on the post-test of the Watson-Glaser Critical Thinking Appraisal.

7. Supervising Teachers were the public school teachers to whom the student teachers were assigned and in whose classes they interned.

Limitations

This study was limited to senior high school students taking social studies courses in a large urban public school system in the North Texas area, and the student teachers who were teaching them.

The implications of the findings of this investigation are limited to the area of social studies under the conditions stipulated herein.
Basic Assumptions

For purposes of this study it was assumed that

1. Any feedback from supervising teachers to student teachers concerning the latters' levels of questioning would be infrequent and would not significantly affect the student teachers' questioning behavior.

2. Skill in asking high-level, divergent questions is one requisite to effective teaching.

3. Student teachers, students, and supervising teachers in the experimental group would not differ significantly with respect to any customary characteristics for grouping from student teachers, students, and supervising teachers in the control group.

Instruments

For purposes of this study the following measurement instruments were utilized:

1. Coding of classroom verbal interaction and levels of questioning was accomplished by using the Observation Schedule and Record, Form 5, Verbal (OScAR 5V), latest revision of the OScAR system, March 1968, by Medley, Schluck, and Ames (22).

The OScAR 5V manual lists eighteen categories which are to be used in the coding: four for pupil utterances and fourteen for teacher utterances. Six of the fourteen categories are specifically concerned with types and levels of
teacher questions (22, pp. 11-12). These question-types are summarized on the OScAR 5V form as being either convergent or divergent. A total of sixty-eight verbal events can be recorded on the OScAR 5V (22, p. 10).

Coding is based on classifying behaviors as to how they appear to the observer, not how they are supposed to appear; the coder is to be primarily concerned with teacher utterances (22, pp. 3, 6).

Mouly in the Encyclopedia of Educational Research, Fourth Edition, writing on observation schedules, states that "probably the best known and most adequate classroom observation is the Observation Schedule and Record (OScAR) developed by Medley and Mitzel . . ." (24, p. 1146). He further adds "... that it provides data amenable to processing by fairly adequate statistical procedures" (24, p. 1147).

2. Achievement scores on pre- and post-tests were obtained from public school students through use of the following standardized instruments in social studies:

a. The Sequential Tests of Educational Progress--Social Studies, Forms A and B, Grades 10-12 (31). According to McLendon in reviewing this instrument in Buros' The Sixth Mental Measurements Yearbook, "The STEP tests in social studies continue without peer, indeed almost without available counterparts, as the leading standardized series of skill tests in social studies" (6. n. 1974).
Oliver reviewed this instrument, also, stating that the series was developed on the assumption

... that the focus of education is upon the development of critical skills and understandings rather than upon teaching only the facts of lesson material and that success in education is to be measured in terms of student's ability to apply school-learned skills to the solution of "new problems" (6, p. 1225).

Reliability correlations on the A form range from .84 to .93.

b. The Iowa Tests of Educational Development, Test 5,
   Ability to Interpret Reading Materials in the Social Studies, Forms X-4 and Y-4, Grades 9-12 (14). Page reviewed the ITED (6, pp. 49-51), stating that the battery has been used to test one and a half million students. He reports predictive validity correlations ranging from the .40's to the .70's between the Iowa Test series composite scores and rank in high school graduating class, average high school grades in specific courses, high school grade-point averages, and college freshmen grades. "Within-grade split-half reliabilities for all tests are reported, ranging in the .80's and .90's with reliabilities for the composite reaching a striking .98 or .99" (6, p. 50).

Test Five consists of representative reading selections taken from social studies textbooks and references, from magazine and newspaper articles on social problems, and from the literature of the social studies in general (14, p. 20). There are eight descriptive passages with four-option
multiple-choice questions following each passage for a total of eighty items. The manual says of Test Five.

The student can not obtain a high score merely by assimilating the ideas presented in the passages. Rather, he must evidence the ability to "read between the lines," to see the implications of the ideas presented, and to evaluate the author's approach and handling of the topic (14, p. 20).

3. To measure critical thinking ability of students taught by student teachers, pre- and post-test scores were obtained by use of the Watson-Glaser Critical Thinking Appraisal, Forms YM and ZM, Grades 9-12 (38). This instrument of 100 multiple-choice items measures inference, recognition of assumptions, deduction, interpretation and evaluation of arguments, and skill at reaching conclusions—abilities which the authors describe as making up a definition of critical thinking (38, p. 10).

Split-half reliability coefficients range from .85 to .87 on Form YM and from .77 to .83 on Form ZM. Validity ranges from correlations of .56 to .79 between subtests and total tests.

4. A questionnaire designed for this study was used to obtain student teacher responses as to the nature and amount of feedback they received from their supervising teachers and to the college supervisor's techniques of supervision. (See Appendix.) Items eight through twelve of the "A" part of the questionnaire were included to find out if the supervising teachers of the student teachers in either
group discussed levels of questioning with their student teachers and whether the student teachers felt that their own patterns of questioning were changed as a result of what their supervising teachers said or did.

Procedures for Collecting Data

Permission was obtained from the Assistant Superintendent for Instruction, the Director of Research, the Director of Teacher Education, and the Consultant for Secondary Social Studies of the public school system in which the study was conducted.

Senior high school students who were taking courses in social studies and who were being taught by student teachers during the spring semester of 1971 were the primary subjects of this study. A total sample of 538 students was used. Sixteen student teachers from a School of Education in a North Texas area university, enrolled in the student teaching course for the spring semester of 1971, were the subjects of the coding and feedback procedures used by the college supervisor. No control was allowed by the public school officials over the assignment of student teachers to the various schools. The public school students, therefore, could not be matched in the experimental and control groups. They were equated statistically, as will be explained in further detail later in this chapter. The student teachers were randomly divided into two equal groups of
eight in the control group and eight in the experimental group. This procedure resulted in a distribution of 275 students in the control group and 263 in the experimental group.

Pre-tests in the achievement of understanding social studies concepts and critical thinking ability were administered by the student teachers under the direction of the college supervisor to all public school students in the study the third week after the student teachers entered the schools and just prior to the time they actually started teaching. Upon completion of administering the pre-tests, each student teacher's verbal behavior in both groups was coded once a week over an eight-week period by the college supervisor, using the OSCAR 5V.

The week immediately following the pre-testing, which was the fourth week of the student teachers' internship, the college supervisor started observing the student teachers in both groups weekly and coding their classroom verbal behavior on the Observation Schedule and Record 5V. Feedback from the OSCAR 5V codings was given to the experimental student teachers, while the control student teachers received standard feedback. The feedback conferences were both on a one-to-one basis and in small groups with the supervising teachers and the public school students not present. These conferences were conducted separately for the control and the experimental groups.
Beginning with the fifth week of their internship, the student teachers in the experimental group were given training by the college supervisor in the skill of questioning. Levels of questions and concepts as formulated by Sanders (30) from Bloom's *Taxonomy* (5) and modified by Davis and Tinsley (8) were used in the training sessions. Other sources used were Aschner (2), Hyman, Chapter 9 (13, pp. 127-255), and the University of Utah Hierarchical Schema of Mental Processes (11). After each visitation, the college supervisor provided feedback to student teachers on their techniques and patterns of asking questions. Information for the feedback came from codings on the OSCAR 5V which showed frequency of convergent and divergent questions asked by the experimental student teachers. The OSCAR 5V forms were shown to and interpreted for them. Particular emphasis was placed on levels of questioning used by the student teachers and an attempt was made by the college supervisor to relate student teachers' questioning practices to the training they received, so as to help them improve their inquiry techniques during the balance of the study, especially in the area of increasing the number of high-level, divergent questions they asked.

Weekly visitations by the college supervisor continued with all student teachers through the last week of their internship at the end of the spring semester. Coding with
OScAR 5V was done at each visit. Frequency tallies from the OScAR 5V codings of high-level, divergent questions and low-level, convergent questions were obtained for each student teacher in both groups once a week during the eight weeks they were teaching, between pre- and post-testing.

At the end of the investigation all public school students in both groups were given post-tests in social studies achievement, using equivalent forms of the same instruments employed for the pre-tests and an equivalent form of the Watson-Glaser Critical Thinking Appraisal.

Student teachers in both groups were asked to respond to the questionnaire on supervision at the conclusion of the study. They were not identified by name but each questionnaire was coded so that it could be recognized as being either from a control group student teacher or an experimental group student teacher.

Procedures for Analysis of Data

To test Hypotheses One and Two, the research data was subjected to the following analysis:

The design of this study made use of student samples in an actual school setting, making it impossible, from the standpoint of getting public school administrative approval, to manipulate the students for experimental purposes. "Intact" student groups had to be used. For this reason, an analysis of covariance as described by Popham (26, pp. 221-
256) and Roscoe (28, pp. 254-263) was used to adjust the experimental and control groups of public school students to make them statistically equal.

The pre-test scores were the control or relevant variables and the covariant in each case. The post-test scores were the criteria or dependent variables. A significance level of .05 was required for rejection of the null hypotheses.

The statistical treatment of analysis of covariance was accomplished by the I.B.M. Computing Center of North Texas State University.

Roscoe explains the rationale for using covariance:

The analysis of covariance is a blending of regression and the analysis of variance, which permits statistical rather than experimental control of variables. The result is equivalent to matching the various experimental groups with respect to the variable or variables being controlled (28, p. 254).

Hypothesis Three was tested by using the chi-square test of independence. A significance level of .05 was required for rejection of the null hypothesis. The variables of convergent versus divergent levels of questions between control and experimental groups were tested to determine if they were related. These levels were identified by frequency tallies for both groups on the OSCAR 5V form with separate sums for convergent and divergent questions being computed for each group.
Roscoe says that "chi-square tests of independence are extremely useful statistical procedures for determining whether two nominal (or higher level) measures are related" (28, p. 196). He explains that

Chi-square tests of independence are an extension of the chi-square approximation of the multinomial to the situation in which a double classification system is used and the expected frequencies are derived from sample data (28, p. 197).

The chi-square procedure involves contrasting sample frequencies actually falling within certain categories with those which might be expected on the basis of the hypothetical distribution (26, p. 277). If a marked difference exists between the observed frequencies in each category and the frequencies expected to fall in each category on the basis of chance or a previously established distribution, then the chi-square test will yield a numerical value large enough to be interpreted as statistically significant.
CHAPTER BIBLIOGRAPHY


11. Frontiers of Thinking (author not given), A Faculty Publication of the William M. Stewart School, University of Utah, Salt Lake City, Utah (no copyright date).


"The 1960's and perhaps the 1970's in the history of American teacher education probably will be remembered as decades triggering intensified examination, criticism, and experimentation," predicted Hermanowicz (52, p. 3) in 1967, when he spoke at the annual Association for Student Teaching meeting in Chicago, Illinois. What he stated five years ago is coming to pass. Both pre-service and in-service teachers are being examined, evaluated, and used in experimentation, the most promising of which appears to be systematic observation of classroom teacher and student verbal behaviors as they relate to the learning process. Of the verbal behaviors being examined and treated, that of teacher questioning, has received a great deal of attention in the last couple of years. The purpose of this chapter is to present reports of research dealing with the development and use of observation systems; to examine studies of the verbal questioning behaviors of teachers; to review literature dealing with questioning strategies in teaching methods textbooks and specialized programs to change teachers' questioning behavior; and to consider how previous research points to the importance of this study.
The review of related literature was concerned with the following areas:

1. A history of the development of research on classroom observation systems

2. A history of the development and modifications of the Observation Schedule and Record (OSeAR) in research studies of teacher effectiveness

3. A summary of the classification of types of questions asked by teachers as identified and assessed in observational studies

4. A summary of research related to the use of questions in teaching

5. A summary of research relevant to this study on the use of questions in teaching

6. A review of the analysis of questioning tactics as reported in writings other than research studies

7. A review of some current teaching methods textbooks related to questioning in teaching and specialized training programs designed to change teachers' questioning behavior

8. A statement of the relationship of the reported research to this study

A History of the Development of Research on Classroom Observation Systems

That "the teaching-learning process is a highly complicated and often badly misunderstood phenomenon which has
been a source of speculation and study throughout recorded time" (30, p. 69) is evident when one examines the literature on the teaching process. Neujahr says that "For a long time this research centered on teacher characteristics and their relation to pupil learning outcomes. Results were contradictory and inconclusive" (86, p. 221). Soar points out that in 1957 Sister Mary Long, in reviewing research studies on predicting teacher efficiency, said, "... if all of the preceding twenty years of research were wiped out overnight it would make virtually no difference" (108, p. 116). In the same paragraph, however, Soar speaks in a more positive and promising tone: "But the nature of this literature has changed considerably since 1960, and a central aspect of the change appears to be the increasing use of systematic observation" (108, p. 116).

One of the earliest attempts to systematically observe classroom behavior was made by Horn in 1914 (73, pp. 254-255; 85, pp. 3-4) when he utilized a classroom seating chart and coded students' behavior by marking circles and squares on the chart. In 1928 Puckett expanded on Horn's system by adding fourteen symbols to represent student behavior (73, p. 254; 85, p. 4). A. S. Barr in 1929 reported his study, which was the first of any magnitude bringing together the greatest variety of behavior data to that time for distinguishing between effective and ineffective teachers (73, pp. 258-260; 83, p. 1146). Wrightstone in 1934-35 used the
seating chart technique and developed a series of categories to record interaction of the total class or individual student responses (73, pp. 255-256; 85, p. 4).

Bellack (9) and Murray (85) have traced the development of observation systems in the areas of verbal and non-verbal interaction, specific subject matter and general content areas, and teacher practices, from 1939 to 1970. Anderson's integrative-dominative verbal behavior research in 1939 (85, p. 4) concluded that the teacher's behavior pattern influenced the behavior of pupils in various ways, and that this teacher behavior could be successfully recorded. Lewin, Lippett and White in 1943 (85, p. 5) found that the strategy of leadership altered the social climate of interaction in boys' club groups. In 1949 Withall (85, p. 5) developed a seven-category system to measure social climate of the classroom, while Bales in 1950 became the first individual to time observations at set intervals, using small groups (85, p. 5).

Flanders (35), in 1959, drew heavily upon Withall for his Interaction Analysis System of ten categories to measure direct and indirect teacher influence, while Hough (55; 85, p. 5) devised a sixteen-category system, in 1964, fashioned after Flanders' system. Ober, in 1968, modified Flanders' instrument and created a nine-category system called the Reciprocal Category System (85, p. 5).
Non-verbal behavior has been measured by Galloway in 1968 and Kounin in 1970 (85, p. 5) who was concerned with classroom management activities (9, p. 12).

Systems have been designed for observing specific subject matter areas by Wright and Proctor in 1961 (9, p. 6) and Hernandez in 1970 (53) in the field of mathematics; Massialas in 1969 (85, p. 6) for social issues; and Oliver and Shaver in 1966 (9, p. 12; 85, p. 6; 87) in the social studies.

A number of observation forms have been developed to make operational the concepts developed by Bloom (14), especially in the area of identifying cognitive levels of teachers' questions. They include the Florida Taxonomy of Cognitive Behavior, in 1968, by Brown, Ober, and Soar (85, p. 6); the Teacher Question Inventory, 1964, by Harris and McIntyre (27, 50, 82); the Teacher-Pupil Question Inventory by Davis and Tinsley in 1966 (26, 27); Clegg's use of the Bloom-Sanders' Taxonomy (97) in 1967 and 1969 with elementary level teachers (11, 20, 21, 27, 59); the Questioning Strategies Observation System by Morse, 1968 (26, 27, 81, 82); a protocol that operationally defined the six levels of Bloom's Taxonomy in a "grammar of the interrogative" developed by McCartin in 1969 (21, pp. 5-9); Rogers' Teacher Oral Question Observation Schedule, 1969 (93); and the Price-Belland Question Analysis System developed by Ann and John Belland and Thomas Price in 1970 (11).
Hill in 1969 (85, p. 6) used the figure-ground concept of perceptual psychology to develop the **Content Analysis System** based on the earlier work of Hough and Duncan (55). Brown, 1968, recorded teacher practices on the **Teacher Practices Observation Record** based on John Dewey's philosophy of education (85, p. 7). He also developed a **Personal Beliefs Inventory** and a **Teachers Practices Inventory** which can be used together to assess the consistency between a teacher's beliefs and his actual teaching practices (85, p. 7).

In the past decade numerous systems have been developed and used to assess teacher effectiveness (1-6, 8-11, 13, 23, 27, 35, 41, 50, 53, 55, 73, 79, 81, 82, 84, 85, 94, 102-106).

Mouly considers observational studies to be the most fundamental procedure to gather scientific data (83, p. 1145) and Biddle classifies the procedures involved in behavioral observation techniques used in recent years into four categories (12, pp. 21-23): (1) **participant observation**, (2) **categorical check lists**, (3) **specimen records**, and (4) **electronic recording** of behavior. Rosenshine describes four major uses of observational systems in the evaluation of the educative process (94, p. 288): (1) assessing the variability of classroom behavior either within or between instructional programs, (2) assessing the agreement between classroom behavior and specified instructional criteria, (3) describing what occurred in implementing
instructional materials, and (4) determining relationships between classroom behavior and instructional outcomes. Bellack states that differences among existing observation systems can be identified in terms of "(1) dimensions of classroom behavior to be classified, (2) type of observational schedule devised, (3) observer's frame of reference for coding, (4) unit of behavior to be used in coding, and (5) range of applicability" (9, p. 4).

Observational systems can be classified into two kinds according to Bellack (9, pp. 4-5), Medley and Mitzel (73, pp. 298-303), Murray (85, p. 3), and Rosenshine (94, pp. 280-281). They are the sign systems and the category systems. The former usually consist of a large number of items referring to concrete, specific behaviors. The observer is to record these behaviors if they occur during a period of observation. The category system limits observation to specified dimensions of classroom behavior, providing a set of categories into which each unit of observed behavior is classified. Such systems are usually made up of a smaller number of items at a higher level of abstraction and, therefore, demand a higher degree of inference by the observer (9, p. 7).

A basic problem in working with any type of observation system is in "... specifying the unit of behavior that is to be used as the basis for coding" (9, p. 9). Some developers such as Flanders have specified an
arbitrary time unit wherein the observer is required to make a record of what behaviors occurred during the specified time (9, pp. 9-10). Time units range from Flanders' three seconds to entire class periods. Instead of time units, other researchers have specified analytic units such as Meux and Smith's "episode" (9, p. 11; 79) and the "interchange" suggested by Medley and his associates with the Observation Schedule and Record, 5V (76, p. 7).

Objections to direct observation systems have been made by some critics, who feel that the presence of the observer or recording equipment in the classroom is so distracting that the observed behavior cannot be regarded as "typical" or "normal" (9, p. 15). The work of experienced observers such as Biddle and Adams (9, pp. 15-16), Meux and Smith (9, p. 16), and Hughes, Flanders and Bellack (9, pp. 15-16) indicate, however, such is not the case. Heyns and Lippitt (54) made an extensive review of systematic observation techniques in social psychological research in 1954. They reported that experienced users of observation systems share the common feeling that "... observers have very little effect if any. This belief is shared by experimenters who have worked in a wide variety of situations and with many kinds of subjects" (54, p. 399).

Another problem in working with observation systems is in establishing their reliability and validity. Medley and Mitzel describe procedures for developing reliability
coefficients and reliability estimations as well as techniques for statistically treating observations to test research hypotheses (73, pp. 307-325). Bellack suggests researchers give more precise definitions to the categories and the units of analysis, formulate clearly stated coding rules for the guidance of coders and carefully train observers and coders to increase reliability of observational systems (9, pp. 17-18).

A final, more recent problem in using observation systems for educational research has resulted from the rapid growth of numerous observational instruments in the last few years. Neujahr says that "... no one satisfactory way has been found for describing what takes place in the classroom" (86, p. 224); "... there is no common acceptance of a model to guide research" (86, p. 225). He suggests that common terms need to be agreed upon, that researchers must gain knowledge as to what affects the attainment of particular goals with particular students, that more comparative studies of existing research need to be conducted, and that researchers should work more closely within the existing systems rather than devising their own observational systems for every new research project (86, p. 228).

Despite the problems connected with the use of observational systems in research, Soar contends that:

There are now a series of findings that are both very convincing and very heartening. Instead of the approximately forty years of research
that . . . produced essentially nothing in the way of identifying the nature of good teaching, there are now a number of conclusions that can be accepted with considerable hope that we are beginning to know what kinds of things a teacher ought to do in the classroom to produce growth in pupils (108, p. 117).

Abramson (1) believes that observation systems make possible increased technology to a wide variety of educational problems such as school program evaluation and teacher preparation program evaluation. Lindvall (69) affirms that data from observational procedures have been used in experiments designed to change teacher behavior along certain dimensions. "This type of study should be emphasized to an even greater extent, particularly when it is coupled with an investigation of the effects on pupils of this changed teacher behavior" (69, p. 17).

Biddle predicts that greater use will be made of behavioral observation during the seventies. "Indeed, it may be said that progress in understanding teacher competence depends upon such methods" (12, p. 23).

Because this study attempted to identify and examine the effects of one aspect of teacher competency, systematic observation in the classroom was utilized. To record teacher and pupil verbal interaction, a form of the Observation Schedule and Record (OSCAR) by Medley and Mitzel (75) was chosen as the observational instrument. An account of its conception, development, modifications, and uses in research follows in the next section.
A History of the Development and Modifications of the Observation Schedule and Record (OScAR) in Research Studies of Teacher Effectiveness

The Observation Schedule and Record (OScAR) system by Medley and Mitzel (75) is a direct descendent of the Withall technique. Following is a brief history of its origin and development to the present form, OScAR 5V, as used in this study.

Modifying and combining items constructed by Cornell, Lindvall, and Saupe in their Classroom Observation Code Digest (1952) and Withall's Social-Emotional Climate Index (1949), Medley and Mitzel (73, 75) redefined the categories in both instruments in simpler terms to increase observer accuracy by reducing the difficulty of the judgments required. They also designed the OScAR to be used by a single observer visiting a classroom alone rather than having several observers in the class together (73, pp. 278-280; 75). Thirdly, they put seventy-one items into the schedule. These items were grouped into several sections and included signs for "teacher lectures," "teacher answers pupil's questions," and "pupil talks to group" (9, p. 7). The authors wanted as many significant aspects of what goes on in a classroom as possible with no concern for their relationship to any dimension or scale.

The original OScAR was printed on a card and was used with forty-nine beginning elementary teachers (75, p. 87). They were observed for thirty minutes at each visit, twice a
day for ten weeks. After the initial round of visits, the first form was revised and after the final visits this modified schedule was subjected to statistical analysis for reliability and validity verification (70; 73, pp. 281-282; 74; 75). Factors which emerged were Emotional Climate—the amount of hostility observable in a classroom, with a high score indicating rare hostile reactions; Verbal Emphasis—the degree to which verbal activities predominate; and Social Organization—the amount of social grouping and pupil autonomy in a class (70, pp. 268-269; 73, pp. 282-283; 74, p. 241). Medley said that reliabilities on these dimensions ranged from .77 to .90 which "... may be taken as indicating that the scales did measure stable differences in teacher behavior" (70, p. 268).

Five measures of teacher effectiveness—reading growth, group problem-solving skill, pupil-teacher rapport, supervisors' ratings, and teachers' self-ratings—were tested for intercorrelations within schools (74). Reliabilities ranged from .605 to .916 on the items used to mark behaviors (73, p. 282). Medley and Mitzel determined that

The five measures of effectiveness were found to center around two distinct aspects of effectiveness. Supervisory ratings and pupils' reactions to their teachers appeared to reflect the teacher's ability to get along with children; teachers' self-ratings and measures of pupil gains (in reading and social skill) appeared to reflect effectiveness in stimulating pupils to learn to read (74, p. 245).
They concluded that "... relatively untrained observers using an instrument like OScAR can develop reliable information about differences in classrooms of different teachers" (75, p. 91), and that observations made with the Observation Schedule and Record "... can contribute to the solution of many important problems having to do with the nature of effective teaching" (75, p. 92). These conclusions were to be tested by others who used and modified the original form.

Wilk and his associates (73, p. 283; 116) in 1960 used OScAR III, a modification of the first schedule, to study thirty-six student teachers by seeing whether admissions data when the interns entered the teacher training program could predict classroom performance. They found that admissions data were not significantly related to classroom observations of teacher dominance but teacher integrating behaviors could be predicted from certain items of admissions data (116, p. 314). Results from OScAR III codings produced some significant findings. The student teachers tended to teach their classes in traditional ways with very little grouping and a great deal of emphasis on verbal materials and pupil recitation. Teachers in the upper elementary grades used a greater variety of materials, talked more to the pupils, and asked more questions than did the student teachers in the lower elementary grades (73, p. 283).

Bowers and Soar (18; 73, pp. 283-286) in 1961 used OScAR 2a to compare on the three dimensions recorded by the
instrument, teachers who had laboratory experience in human relations, to teachers who did not have this training. Personality tests were administered to the teachers and the experimental group attended summer workshops in human relations. The researchers found that there was a correlation between scores on the Minnesota Teacher Attitude Inventory and Emotional Climate and pupil-teacher rapport as recorded on OScAR. There was also a correlation between the Verbal Emphasis dimension of OScAR and scales on the Minnesota Multiphasic Personality Inventory (73, pp. 284-285).

Virginia Morrison (73, p. 283; 80) studied seventy-one secondary student teachers in a laboratory school in 1961 to find out what kind of classroom behavior of teachers and pupils would be rewarded if a system of merit rating were adopted. She used a modified version of the OScAR, which she called the Revised Observer Schedule and Record (ROScAR). She found that student teachers who were rated well by their supervisors and pupils elicited educationally desirable responses from their pupils and that they showed relatively high amounts of positive diversified behaviors, methods, activities, and uses of materials. From this study Medley and Mitzel said of OScAR ratings, "Again it is clear that emotional climate is an important factor in supervisors' ratings of teacher competence, along with other variables which also seem to be related to pupil-teacher rapport" (73, p. 283).
Revised OScARS 3d, e, and f, consisting of 170 items, were used by Schueler, Gold, and Mitzel (70, 99, 100) along with closed-circuit television and 216 video tapes of fifty-four elementary student teachers. Observers coded behaviors on OScAR from the video tapes. The results reported in 1962 "... indicated improvement of student teachers in general but did not establish important differences among the experimental methods employed" (99, p. 362). They concluded that supervisors preferred a combination of in-person observation and kinescopes (100, p. 123).

Two important findings from this study are

3. The cooperating teacher, the class, and environmental factors in the classroom are of critical importance in the experience of the student teacher. ... ... ... ... ... 

5. It is possible to develop an observation schedule to describe and to evaluate teacher and pupil behavior in quantitative terms for later analysis ... (100, p. 123).

In 1964 Soar (107) conducted a study of the development of vocabulary and reading skills in elementary pupils in relation to the climate and control existing in the classroom. He used Flanders' Interaction Analysis and a revised form of OScAR to code classroom climate and verbal behavior. He hypothesized that reading and vocabulary learning would be hampered by stress in the classroom and that stress would be created both by direct teacher control and a negative emotional climate. Soar found, however, that emotional climate did not produce differences in reading growth, but indirect
teaching produced greater growth than did direct (107, p. 248).

Flint (37) in 1965 used OSCAR 3d to determine if a correlation existed between the verbal behavior of student teachers and their cooperating teachers. During an eight-week period, she recorded observations in six elementary grades. At the half-way point, student teachers switched cooperating teachers but remained at the same grade level. Flint's conclusions were that the verbal behavior of student teachers changed significantly during their internship. They became more supportive, less repeating and less accepting. The frequency of teacher initiated statements increased. She found a high positive relationship between the questioning behavior patterns of the student teachers and their cooperating teachers. Relationships were also found to exist between these two groups and other types of verbal behavior and procedural matters.

In 1966 four studies were reported in which a form of the Observation Schedule and Record was used to code classroom behavior (51, 71, 72, 78).

Herman (51, pp. 27-29) assessed teacher and pupil verbal behavior in fifth grade classes of above-average, average, and below-average pupils. He found that teachers of the above-average group were more indirect. They used more praise, questions, and acceptance of pupil ideas than they used lecture. The teachers of the average group
were neutral; that is, their verbal patterns were one
direct statement for each indirect one. Teachers of the
below-average group were the most direct in their verbal
patterns.

Merrifield, Phillips, and Davis (78) observed 115
secondary student teachers on three occasions approximately
one month apart. Five factors which were independent as-
pects of teacher and pupil classroom behavior were isolated.
They were (1) seat work, (2) affection, (3) teacher non-
verbal support of the learner, (4) teacher verbal support
of the learner, and (5) teacher-total-talk, with emphasis on
problem-structuring. Results indicated that teachers
assigned greater amounts of seat work as the quarter prog-
ressed and that teacher-total-talk increased from the early
to the middle portions of the quarter but declined toward
the end, while seat work continued to increase slowly (78,
p. 3).

In 1966 Medley (71) reviewed the study by Schueler,
Gold, and Mitzel (100) and indicated that a number of dimen-
sions which were not accurately measured by the OSCAR used
in that study could be measured if better categories were
devised. This led Medley and his associates at the City
University of New York to modify the OSCAR III into an in-
strument they called OSCAR 4v, this time dealing only with
verbal behaviors (71, p. 49). It contained fifty different
categories and the coding procedures were refined to require only a Yes-No decision on the part of the observer. Medley described **OScar 4V**:

The types of behaviors discriminated in the system look in general like things one should know about a teacher. Statements are classified according to whether they relate to motivation, management, or subject matter; questions are classified according to whether they call for a prescribed answer, offer the pupil a chance to originate his own answer, or ask him to discuss a previous pupil comment. Information is obtained about how complex the content of a lesson is, how intense the hostility or supportiveness of the teacher, the extent to which the teacher dominates the pupils, what kind of pupil behavior he encourages, how he uses criticism and objectives feedback, and so on (71, p. 49).

Medley and Hill (72) used **OScar 4V** and the Flanders’ Interaction Analysis technique with seventh student teachers in 1966 to compare the relationships between the two systems. The authors did not report the findings and a review of the literature since the completion of the study failed to reveal a follow up report.

In 1968 Ray (92), using a revision of Morrison’s **ROsCAR** (80), labeled **ROsCAR II**, examined the relationships of observer feedback, conveyed through conference procedures, to accompanying changes in the teaching behavior of thirty elementary interns. Ray held conferences with the control and experimental groups. He showed them the results of the coding of their verbal and non-verbal behaviors. In the control group he made no interpretations. The experimental group was shown frequencies of occurrences by categorical
behaviors and a graphic representation of their scores. An interpretation and discussion was conducted with the experimental subjects. Ray found no significant changes in separate categories of teacher behavior for student teachers in the experimental group. They did, however, show mean scores above the grand mean in all categories measured, and they showed the greatest improvement in over-all performance.

Further studies using OScAR 4V revealed its weaknesses and led Medley and his associates (76) in 1968 to refine it still more. The result was the Observation Schedule and Record, Form 5, Verbal. This latest revision, still to record only verbal behaviors, was designed so that teacher classroom behaviors could be classified on how they appear and not on how they were supposed to appear. Thus, the observer is not required to make inferences on behaviors (76, p. 3). OScAR 5V uses elementary binary discriminations concerned with teacher verbal behaviors (76, pp. 3-4). There are eighteen categories in all, four for pupil utterances and fourteen for teacher utterances. Sixty-eight different events—thirteen kinds of statements and fifty-five kinds of interchanges—form the basis for inferences to be made about classroom environment from OScAR 5V codings (76, p. 10).

Smoot (105) altered OScAR 5V to sixteen categories on an instrument he called the Laboratory Observation and
Record (LCScAR) for his study in 1968. He used teacher candidates in a teaching laboratory phase of a secondary education course. During an eight-week period, three groups taught micro-lessons in the laboratory setting. Two of the groups underwent three and a half hours of training in using the LCScAR. After each teaching session students in all three groups received standard feedback consisting of peer and instructor evaluations. One of the groups trained in LCScAR received additional feedback from the form following each laboratory lesson. Smoot found that the teacher candidates who received training in the Laboratory Observation and Record demonstrated teaching behaviors different from those which were exhibited by teacher candidates who did not receive this training.

To find out if eighth and eleventh grade mathematics and social studies teachers exhibited different verbal teaching behaviors, Kysilka (66) in 1969 used OScAR SY to code twenty-four teachers who had at least one year of teaching experience. She found that eighth grade teachers used more directing statements than did eleventh grade teachers. Mathematics teachers asked more convergent questions, more procedural-positive questions, used more directing and describing statements, and talked significantly more than did social studies teachers. On the other hand, social studies teachers rejected student responses more frequently, asked more divergent questions and used more
desisting statements than did mathematics teachers. Social studies teachers also asked a significantly greater proportion of divergent questions than convergent questions. Kysilka concluded that students in social studies classes used more non-substantive statements, but that they volunteered substantive information more frequently than did students in mathematics classes.

Morse (81) developed a Questioning Strategies Observation from OScAR 5W in 1968 and used it to compare the effects of laboratory and non-laboratory instruction on questioning behaviors of beginning teacher candidates. Further discussion of Morse's investigation follows in the section dealing with research on questioning relevant to this study.

Friedman and Bowers (40), using OScAR 4W protocols in 1969, investigated the extent to which pre-school, kindergarten, and first grade pupils imitate a rewarding teacher's verbal style when talking among themselves. They found that girls imitated the teacher more than did the boys and that imitation increased with grade level. Friedman and Bowers concluded that

The extension of the OScAR 4W observational technique has yielded promising results in measuring student as well as teacher verbal behavior. . . . While a student's contact with his teacher lacks the intensity of his contacts with his parents or peers, there is every reason to believe that the direct observation of student-teacher interactions can contribute to a broader conception of classroom learning (40, p. 10).
Medley and his associates at the Educational Testing Service (77) recently developed another type of observation instrument called the Personal Record of School Experiences (PROSE) for recording events in the life of a child in school settings as they occur. The items are relevant to preschool children. The authors first used the PROSE form in the 1969-70 school year for a longitudinal study of disadvantaged children. At present the study is still in progress and the PROSE form is in the developmental stages.

The foregoing account of the OSCAR systems of classroom observation has attempted to offer proof of its use in scientifically measuring classroom behaviors and justification for its use in the present study to determine if training teacher candidates produces changes in their behavior and results in improved pupil achievement and alteration of their critical thinking abilities.

A Summary of the Classification of Types of Questions Asked by Teachers As Identified and Assessed in Observational Studies

An area of verbal teacher behavior which is currently being focused upon more and more by observation systems and classroom research is that of questioning strategies. In order to study how teachers ask questions, some sort of classification system or systems had to be devised. Gall (42) indicates at least eleven classification systems have been developed in recent years designed to identify types
and levels-of-questions. Among these are: Adams, 1964 (2); Aschner, 1961 (7); Bloom, 1956 (14); Clements, 1964 (22); Gallagher, 1965 (43); Guszak, 1967 (49); Sanders, 1966 (97); and Schreiber, 1967 (98).

Bloom's and Gallagher's systems consist of a limited number of general categories which can be used to classify questions irrespective of content (42, p. 708). These systems can, therefore, be used to investigate issues such as different types of questions used in various school curricula. They are of limited use if the researcher wants more detailed descriptions of questions asked in a specific context.

Clements' schema (22) classifies questions asked by art teachers as they talk with students about their artwork. Guszak created the Reading-Comprehension Question-Response Inventory specifically for analyzing questions that teachers ask elementary school reading groups (49). Schreiber's system (98) classified social science questions and contains curriculum specific categories.

Gall states, "Most of the question-classification systems are composed almost entirely of categories based on the type of cognitive process required to answer the question" (42, p. 708). Many of them use Bloom's Taxonomy as the basis for the classifications. "It appears that Bloom's Taxonomy best represents the commonalities that exist among the systems" (42, p. 710).
Bloom (14, pp. 15-16) and Gall (42, p. 710) indicate one major weakness of a system which attempts to analyze the cognitive process to questioning is that such a process is based on inferences. It is not always possible to know whether a student answered a question by using a high-level cognitive process, such as analysis or synthesis, or by using a low-level process of recall. One solution to this problem suggested by Gall (42, p. 710) is for the researcher to control the lesson material on which the teacher bases the questions. The questions could then be classified as recall or higher-cognitive depending on whether the answer was given directly in the assignment.

According to Gall (42, p. 710), a few important educational objectives, such as the types of questions which teachers ask to test students' recall of information and to develop their critical thinking processes, are all that the existing taxonomies classify. He suggests that there are other worthwhile question types which need to be classified. These include questions which cue students to improve on an initially weak response to a previous question, questions which create a discussion atmosphere, questions which stimulate the student's sense of curiosity and inquiry, and questions which guide students' learning of a problem-solving, behavioral or affective skill (42, p. 710).

Another limitation of existing classification systems says Gall is their design in which they primarily
... investigate the types of question which teachers actually use in the classroom, not the types of question which teachers should use" (42, pp. 710-711). He says the focus should be on the educational purposes which good questions serve, not on the clarity of phrasing of "good" questions. Gall feels that further research is needed on teachers' "follow-up" questions and use of questions in discussion since such techniques could have a significant impact on student learning in classroom teaching situations (42, p. 712).

A Summary of Research Related to the Use of Questions in Teaching

No teacher, professional educator, or even lay person making an occasional visit to a classroom would challenge the contention that instructors spend much of their time asking students questions. Aschner says that ". . . the teacher is a professional question maker" (7, p. 44) and that asking questions ". . . is one of the basic ways by which the teacher stimulates student thinking and learning" (7, p. 44; 42, p. 707). Crump believes that questions posed by both teachers and students ". . . are foremost among the stimuli which trigger thinking and thus sets the tone of cognition" (24, p. 657). The widely used and accepted Flanders System for Interaction Analysis includes asking questions as one of the ten major dimensions for studying teachers' behavior (36; 42, p. 707).
Yet Gall, in his recent (1970) and very comprehensive paper on the state of research knowledge in the area of classroom questioning, says that "Granting the importance of questions in teaching, researchers still do not know much about them" (42, p. 707). He states that educators, including Aschner (7), Hunkins (57-59), and others, generally agree that teachers should emphasize the development of students' skill in critical thinking rather than in learning and recalling facts. "Yet, research spanning more than a half-century indicates that teachers' questions have emphasized facts" (42, p. 712).

The earliest and probably the first serious study on teachers' questioning practices was made by Stevens (110) in 1912. She found, in her four year study, that two-thirds of the teachers' questions required direct recall of textbook information (15, p. 5; 42, p. 712; 110). On the average, teachers asked two to four questions per minute (110, p. 16), with a mean number of 395 questions per day (42, p. 707). Still, Stevens did not find what she termed "good" questions being asked by the teachers (15, p. 5; 110).

Haynes' study (42, p. 712) in 1935 showed that 77 per cent of teachers' questions, in sixth grade history classes, called for factual answers and only 17 per cent were judged to require students to think. Similarly, Corey (42, p. 712) made a study in 1940 classifying all questions asked by teachers, in a one week period, in a laboratory high school.
He found 71 per cent to be factual and 29 per cent to require a thoughtful answer.

In the last several years studies have indicated that questioning practices are essentially unchanged. In 1960 Floyd (38) investigated the oral questioning activity of thirty primary school teachers. Forty-two per cent of the questions were memory types (38; 42, p. 712; 59, p. 2). About 20 per cent of the questions required thoughtful responses (42, p. 712), and they were employed only slightly more than 5 per cent of the time (38; 59, p. 2). Floyd concluded that teachers' oral questions seemed to be used primarily to check factual recall and not to stimulate thinking. Guszak (49) found similar percentages of fact and thought questions asked in elementary reading classrooms in his 1967 study.

Gallagher and Aschner, in 1963, developed a system containing the categories of cognitive-memory, convergent thinking, divergent thinking, evaluative thinking, and routine to classify questions asked by teachers of gifted secondary students (44, pp. 186-190). As Gallagher reported in 1965, in a final account of the initial study (43), more than half of the questions asked by teachers were of the lowest level. In 1966 and 1967 Davis and Tinsley (26; 27, p. 715; 112) grouped questions asked by social studies student teachers. As was the case with the Gallagher-Aschner study, Davis and Tinsley found that "More than half of the
questions asked by both groups were judged to test students' recall of facts" (42, p. 712).

Developing a seven-category system for analyzing questions, Adams (2), in 1964, categorized 3,628 questions. He found that social studies teachers used significantly more memory questions and significantly fewer evaluative, clarifying, and neutral questions than did English teachers. Senior high school social studies teachers used significantly more logical reasoning questions and significantly fewer memory questions than did junior high school social studies teachers.

Two studies reported in 1965 relate to classroom questioning procedures of teachers. Amidon and Giammatteo (4) wanted to find out if there were verbal patterns of behavior characteristic of superior teachers. Findings revealed that the superior teachers dominated their classrooms less, used indirect verbal behavior more, and used direction-giving and criticism less than did the normative group of teachers. About 12 per cent more student participation was evidenced in the classes of superior teachers.

Flanders' now classic study of mathematics and social studies teacher influence, pupil attitudes, and achievement (35) revealed among other findings that "... asking of questions and the giving of information can account for 70 to 80 per cent of what the teacher says..." (35, p. 40). As concerns social studies teachers, Flanders found that the
most indirect ones "... ask more questions and stimulate more student talk" (35, p. 82). He concluded that "... significant differences in achievement support the generalization that the teaching methods we have called indirect produce more achievement" (35, p. 108).

In 1963 and 1965 Bellack and his associates (10) analyzed the linguistic behavior of teachers and students in fifteen social studies classes. They found that teachers dominated the verbal activities of the classrooms and that an instructional pattern existed. The pattern started with the teacher asking a question which a pupil answered, and then the teacher, in turn, reacted or rated the pupil's response (10, pp. 93-96). Gage and Unruh (41) likened this pattern to the system of programmed instruction. "In short, live teaching and programmed instruction seem to have in common a basic cycle or rhythm" (41, p. 5).

A six-year longitudinal study conducted by Seagren, Gilchrist, and Beggs (101), from 1962 to 1966, compared student teachers to regular classroom teachers as to the formers' impact on attitude and achievement of high school students. Questioning skills was one of the variables examined. The investigators found that students taught by student teachers exclusively reached a higher level of achievement than did students taught by regular teachers.

Clegg (20, 21), in 1967, conducted an investigation at the elementary level similar to Davis and Tinsley's studies
with secondary students (26, 112). His conclusions were that a complete range of cognitive levels in the questions asked by the student teachers existed. Out of this range, 27 per cent of the questions were classified as memory types. Also in 1967, Pankratz (89) studied the differences in verbal behavior patterns of thirty twelfth grade physics teachers. The teachers were divided into two groups according to ratings of "high" or "low." The results revealed a general trend for teachers in the "low" group to ask more narrow, short-answer questions and for teachers in the "high" group to ask more broad questions encouraging student-emitted responses. Teachers in the low group also tended to use the cyclic pattern of question-answer-question-answer in their classrooms (89, p. 201). Students' questions and teachers' answers to students' questions were more frequent in the high group classes and there was a marked difference in the length of answers to students' questions between the two groups (89, p. 208).

The frequency of questions asked and the types of questions used by art teachers at the first grade, seventh grade, and college levels were studied by Clements (22) and reported in 1967. He found that teachers seldom paused to give students a chance to think about their answers. The total amount of questions asked during a fifty minute period ranged from ten to 340. First grade teachers asked beginning and intent questions followed by experience questions.
Seventh grade and college teachers tended to precede and follow judgment questions with their own opinions. Over half of the questions asked by the art teachers at all three levels received answers of a second duration or less. Clements concluded that art teachers seem to attach little value to their pupils' opinions (22, p. 22).

During the school year of 1967-68 Hunter (60) investigated the results of training on first grade science teachers. She examined the verbal behavior of eleven of these teachers and compared them to eleven teachers who had not received the training. Among other things, she hypothesized that the teachers who had been trained would use a greater range of verbal behavior and would ask more questions with a greater percentage of these questions being of the divergent type. She found, however, no significant difference in the amount of questioning behavior between the trained and the untrained teachers. "About 95% of all questions asked were of the cognitive memory type; . . ." (60, p. 7). Hunter concluded that this finding was of particular interest to publishers of educational materials who pose broad questions in their teacher guides. Teachers will generally begin with a broad question and narrow it down if it is not immediately answered so that they often take the divergent, convergent or evaluative question and make it cognitive memory. "Since most teachers have little training in question asking, they tend to use cognitive memory
questions almost exclusively" (60, p. 8). Hunter decided that the questions designed by the publishers were not measuring the objectives as specified in the materials.

Huenecke in 1970 (56) inquired into the possible differences in teaching behavior of teachers who use curriculum guides and those who do not in terms of what the author called "pre-active" behavior—written instructional objectives and "interactive" behavior—teachers' oral classroom questions (56, p. 379). She used Bloom's Taxonomy to analyze both objectives and questions. Huenecke found that the teachers who used curriculum guides wrote objectives and asked classroom questions at cognitive levels that did not significantly differ from the methods used by teachers who did not use the guides. Although not significant, Huenecke reported that there was a tendency for "Non-users" to write more objectives at the four higher levels than "Users" and that "Users" tended to ask more questions at the four higher levels than "Non-users" (56, p. 382).

The area of student involvement as related to questions asked by teachers was investigated by Dunbar (32) in 1970. Content used by Dunbar included memory, application and evaluation levels of questions for all groups. Students were measured in terms of "time at task," "interest in task," and "performance at task." He found that question types did have a significant effect on the "interest at task" variable. The higher level questions created more
student interest in each group of content area materials. The other two variables of task involvement showed no significant differences.

Blosser in 1970 assessed the effectiveness of a procedure designed to develop skill in questioning, as a teaching technique, by prospective secondary school science teachers and sought to determine if the questioning skill developed during the instructional sequence would transfer to the student teaching experience (15, p. 96). After analyzing the results, Blosser concluded that questioning is a skill which can be developed through instruction and practice (15, p. 108) but this ability did not transfer from the pre-student teaching micro-teaching methods training sessions to the student teaching experience (15, p. 89).

Ladd and Andersen believe that the reason so many teachers continue to ask low-level, cognitive-memory questions is that they are not aware of the level of inquiry which questions elicit in students (67, p. 396). They, therefore, carried out a study, reported in 1970, to investigate the effectiveness of a modified Meux and Smith (79) classification scheme for measuring the level of inquiry present in a class discussion and the effect of inquiry on student achievement. Ladd and Andersen found "... that teachers' questioning behavior strongly influences student achievement" (67, p. 398). They concluded that the students
of teachers who used high inquiry questions performed significantly better on all tests given them.

Yost (119) used a programmed text in Newtonian Mechanics science with interspersed questions at different levels of complexity in a study carried out in 1970. Students made written responses to the questions. Yost found that as the complexity of the questions to which students responded increased, student achievement on relevant and incidental subject content increased; the amount of time taken by students to complete the instructional materials increased; the number of errors made by students in responding to the interspersed questions increased; and as the number of errors made by the students increased, relevant and incidental subject content achievement increased (119, pp. 72-73).

Hernandez (53) in 1970 developed an observation system to analyze cognitive content of teacher discourse in mathematics. The cognitive dimension had as its subdivisions memory, convergent and divergent production. She found from the codings of the class sessions that the predominant cognitive process was memory. Each of the four teachers used in the study had very little discourse judged as convergent and almost none showed divergent production.

John and Ann Belland and Thomas Price (11) trained intern teachers in questioning techniques and in analyzing questions. They used the trainees to compare two systems which have been designed to examine questioning strategies
of teachers. They compared the Price-Belland Question Analysis System (11, p. 2) to the Hough-Duncan Observational System for Instructional Analysis (11, p. 3; 55, p. 130). Their 1970 report disclosed that intern teachers trained to recognize memory questions when placed in practice situations were able to conduct a question-asking session with a lower percentage of memory questions than they did before the training and practice (11, p. 7). They also found upon comparing the two observational forms that there were significant differences in the way the behaviors were recorded on the systems and concluded that the decision-making process on the part of the observers using the two systems influenced the category outcomes (11, p. 11). Their recommendation was that a more reliable questioning-behavior analysis system needs to be developed.

Derhammer (28) in 1971 sought to cause a significant increase in correlation between higher order classroom and test questions, significant increases in higher order classroom and test questions, and significant differences in higher order questioning behavior through training and feedback. He found that there was a significant correlation between classroom and test questioning practices of teachers and training of these teachers. Instruction in higher order questioning techniques was also found to be effective in increasing the incidence of higher order written test questions but not in improving higher order oral classroom
questioning behavior. Derhammer concluded that feedback in itself was not superior to training alone in increasing the incidence of higher order classroom or test questions asked by teachers.

Douce (29) conducted a similar study in 1971, using experienced teachers. The experimental group received instruction in question-asking along with learning packets on Bloom's Taxonomy, including sample questions and activities for the teachers to perform. Douce found no significant difference in the treatment effect between the instructed and the non-instructed groups. There also was no significant interaction between the instruction for the experimental teachers and levels of questions. Douce did find a significant difference between levels of questions.

Using a monograph describing methods of probing as a questioning technique which he called "symbols modeling," Goodwin (46), in 1971, gave this treatment to one group of eight secondary student teachers. To another group he gave the monograph plus an observation form on which they reacted to demonstrations of probing situations, which Goodwin called "symbolic-live" modeling. His third group received no treatment. He found that symbolic modeling did improve the probing questioning behaviors of the intern teachers, but that the symbolic-live modeling treatment did not produce any significant difference when compared to the group who received only the monograph or symbolic modeling alone.
In 1971 Dwyer reported on two studies he conducted on the effect of questions on visualized instruction (33, 34). One project involved the placement of written questions after visual material to determine if the visual illustrations were affected by the question placement (33). The visuals were designed to complement the textual material and the questions were designed to direct the students' attention to relevant learning cues. Dwyer found that "on the identification, terminology, comprehension, and total critical tests, S's receiving the presentation without visuals achieved as well as did those receiving the visualized treatments" (33, p. 182). He attributed these results to the fact that the questions effectively reinforced relevant learning cues.

Dwyer's second study (34) replicated the first (33), except that he preceded the illustrated page by a page containing a question designed to focus the subjects' attention on the relevant learning cues. He wanted to find out this time if written questions function as advanced organizers in complementing visualized instruction. He concluded that "... results indicate that the use of questions as advanced organizers to focus S attention on relevant learning cues is not an important instructional technique for increasing S achievement of all types of learning objectives" (34, p. 263).
Five third grade reading teachers, with at least one year of experience, were the subjects of a study conducted by Trosky (114) in 1971. He sought to relate the modifications in teachers' questioning behavior, as a result of supervisory conferences, to six dimensions of reading comprehension (114, p. 2). Trosky found that four of the five teachers made modifications of their behavior, decreasing the number of recognition questions. The fifth teacher made no changes, but indicated an understanding of how to do so after the self-analysis conference. Trosky stated that the self-evaluation conference facilitated significant changes in both the frequency and proportion of questions evoking the dimension of recognition. This type of question decreased (114, p. 19).

Wadsworth and Flagg (115), in 1971, used written interspersed questions with sixth grade social studies and science classes, and college students. These questions were administered shortly after students had read text material. The investigators wanted to determine the effects of such questions on incidental (recall) learning. The effects on sixth graders and college students of the same text passages, written for the appropriate level of the students, were also examined. They found that "... there was no general facilitative effect of interspersed questions (after relevant text materials) on incidental learning" (115, p. 17).
Zimmerman and Bergan (120) reported their study in 1971 of the use of four school districts throughout the nation which were utilizing the Tuscon Early Education Model (TEEM), a program emphasizing the intellectual processes in elementary school children. They studied teacher questioning behaviors as a means for initiating intellectual operations in students. The TEEM teachers were trained in a process curriculum approach that encourages divergent and higher levels of mental processes. They were compared to a similar group of teachers who had not received the training. Findings supported the hypotheses that experimental teachers trained in process curriculum did significantly differ from the control teachers in one or more of the question categories. They asked a significantly greater percentage of perceptual questions and a significantly smaller percentage of cognition and memory questions (120, p. 23). The experimental teachers displayed twice the amount of perception questions evinced by the control teachers (120, p. 24). The former group placed a greater emphasis on a process approach to teaching critical thinking.

The foregoing broad review of the literature related to research in teacher questioning strategies reveals an upsurge in such studies in the last few years and contradictory findings. However, the pattern is much the same as Soar reported in 1967 concerning the repeated findings of classroom studies on the typical cycle of teacher-pupil interaction.
He said that it "... is one of a very low cognitive level in which the teacher asks a factual question, the pupil responds and the teacher evaluates whether the pupil has produced the answer she wanted in the form she wanted it" (109, p. 10). Research on the use of questions in teaching which was considered relevant to this study is reported in the next section.

A Summary of Research Relevant to This Study on the Use of Questions in Teaching

Following the preceding general summary of research on questioning, special attention will be made, in this section, to investigations of the use of questions in teaching, which were used as prototypes in designing this study. Research which represents the approach recommended by Rosenshine (95, p. 7) will be reviewed in more detail. His criteria and the ones used as objectives in this investigation include (1) training teachers to teach a class in a certain manner, (2) obtaining observational measures to verify that the teachers behaved as they were trained to do, and (3) securing end-of-experiment pupil measures to determine the influences of teachers on their students.

While researchers and educators expound on the importance of good questioning strategy in the classroom by teachers, few researchers have explored the relationship between teachers' questions and student outcomes. Gall (42) says that the most important work in the area to date is
the research by Hunkins (57-59). His purpose was to determine whether the variable of question type bears any relationship to student achievement. Hunkins used written sets of questions for sixth grade social studies students. The questions were keyed to a social studies text. In one group the questions stressed knowledge and in the other, analysis and evaluation processes. Bloom’s Taxonomy (14) was used to order the question types. The analysis-evaluation group earned a significantly higher score on a specially constructed multiple-choice post-test. On six levels of the Taxonomy there were no significant differences between groups, but on application and evaluation questions of the subtests, the analysis-evaluation group scored significantly higher. Gall questions Hunkins’ use of a written multiple-choice type of test to measure evaluation-level thinking processes since such questions may not really have one “correct” answer (42, pp. 714-715). Still Gall feels that Hunkins’ studies suggest that if a group of students is exposed to certain types of questions and if their responses are monitored to improve the quality of these responses rather than correctness, “... then they will be able to answer similar types of questions better than a group of students who have not had the exposure” (42, p. 715).

Schreiber (98) in 1967 classified actual questions asked by fifth grade social studies teachers. She found that
the most prevalent type of question asked was the factual recall. The kinds of questions that were used very little were those that called for the sequential ordering of information, defining and clarifying information, and drawing conclusions. Schreiber then conducted four, one-hour meetings with the teachers in which the purposes of asking questions, and guidelines to use in asking them, were discussed. Different types of questions, and their functions, were identified and questions were constructed by the teachers over selected social studies materials. Schreiber followed these training sessions with a second series of classifications of the teachers' questioning practices. She found a significant decrease in recall questions and an increase in the higher-level questions asked. She concluded that teachers can be trained to ask higher-level questions.

Using the Questioning Strategies Observation System he developed, in 1968, based on the Bloom-Sanders' Taxonomy of questions (97) with a format similar to the OSCAR 5V (27, p. 717), Morse (81) studied the effectiveness of laboratory and non-laboratory instructional techniques in producing desired behaviors in questioning practices. He gave half of his subjects laboratory-centered instruction in questioning tactics, including simulated teaching experience, followed by evaluation and feedback. Both groups were assigned the same readings and instructed to achieve the same objectives. The control group received no laboratory instruction or no
teaching experience. Both groups performed final simulated teaching assignments in which their questioning behaviors were coded by Morse. He found that the laboratory-trained teacher candidates asked significantly more cognitive questions and reacted more positively to student responses to questions than did non-laboratory subjects.

Rogers (93), in the same year, conducted an investigation to find out if questioning strategy of student teachers could be altered to include higher percentages of questions at various levels of a taxonomy of questions. She also studied the effect of questioning on the achievement of students. The experimental student teachers participated in seminars on the purposes and use of varying cognitive levels of questions prior to planning and teaching a four-day unit. Rogers coded both groups with her Teacher Oral Question Observation Schedule, which yields seven cognitive and four non-cognitive categories of questions. She found that student teachers in the experimental group asked significantly higher percentages of high-level questions both in oral and written form. Control group student teachers asked no oral questions beyond the application level and no questions beyond the interpretation level on final examinations, which they constructed. They also asked more procedural questions than did the experimental student teachers. Pupils taught by student teachers who emphasized higher cognitive levels of oral questions did not achieve significantly
different from pupils taught by student teachers who emphasized memory and other lower levels of questions. "Rogers also concluded that a special preparation component on varying cognitive emphases of questions does influence teacher candidates to alter significantly their questioning practices" (27, p. 715).

Even experienced teachers tend to ask a large per cent of low-level questions. Cornell (23) in 1969 used fourth grade teachers with seven consecutive years of teaching experience. The experimental group heard audio tapes of high-level evaluation questions based on Bloom's Taxonomy before they taught micro-lessons. Cornell found that 66.2 per cent of all questions asked by this group fell into the two lowest categories of the Taxonomy. He concluded that there needs to be a more effective method for improving the questioning behavior of in-service teachers.

Olmo (88) in 1970 worked with thirty teaching-methods students. They were assigned to four classes in a laboratory high school and given a theme to develop into a micro-unit in social studies. The students also developed a chart based on Bloom's Taxonomy to analyze questions and used this to identify questions from lessons they observed and designed for their own unit. In a methods class the questions were evaluated in terms of the thinking required. The students then prepared their own questions for the micro-teaching sessions which they each taught to four high
school students in the micro-class. Later each methods student prepared and taught a full class period. They recorded and analyzed questions in the class they were to teach. After their teaching, students were critiqued in their methods course (88, pp. 504-505).

Olmo concluded that the methods students became more perceptive in identifying questions as well as in asking them and that they encouraged their students to ask their own questions, thus giving the prospective teachers practice in considering problems in an analytical manner. She believes this micro-teaching followed by class analysis helped the methods students evaluate their own questions in terms of how effectively they dealt with the situations (88, pp. 507-508).

Tinsley, Watson, and Marshall (113) studied five classes of students and teachers in 1970 who viewed themselves as content-oriented; that is, the program in these classes tended to emphasize acquisition of knowledge. They also observed five classes which considered themselves to be process-oriented; that is, the program in these classes tended to use discovery procedures in student-centered learning situations. The researchers found that there was no significant difference between the content-oriented and process-oriented programs in mean number of questions asked. Overall, memory, interpretation, and procedural questions were asked most frequently by both teachers and students.
The mean number of questions asked per class was slightly higher for the content-oriented program. More memory questions, 28.68 per cent, were asked than any other category. Interpretation types of questions ranked second and procedural questions were third. The investigators concluded that even though process-oriented teachers and students perceived their curriculum as focusing on problem solving, only the lowest form of mental processes was incorporated as questioning behavior (113, p. 6). The content-oriented teachers did tend to ask more memory questions while their students asked more procedural questions. Students in the process-oriented program tended to ask more analysis and evaluation questions.

To determine what effect cognitive instruction had on the classroom cognition level of secondary English and social studies student teachers, Williams (117, 118), in the fall semester of 1969-70, used sixty subjects. The week prior to the beginning of their internship Williams treated the experimental group to four separate training sessions. In the first session he lectured, questioned, discussed and gave handouts of Bloom's Taxonomy; the second session was devoted to stating behavioral objectives with each student teacher having to state at least five objectives at each cognitive level; the third session was practice in stating classroom questions and formulating test questions at the various levels of the cognitive domain with handouts being
given on the Gallagher-Aschner and Bloom-Sanders classifications; and the final session required each student to simulate a teaching lesson using peers as subjects and using behaviorally stated objectives at each cognitive level to teach from (117, p. 75). Williams then coded the student teachers during their internship. He found significant differences between the cognitive behavior of the experimental and control groups of student teachers. He also found a significant difference between the cognitive behavior of pupils of the experimental and control groups. There was no significant difference between groups at the four lower cognitive levels for either teachers or their pupils, but statistical significance was consistently found between the experimental and control groups of teachers and their students at the five higher cognitive levels (117, pp. 79-81). Williams concluded that cognitive instruction with pre-service student teachers can increase their cognitive behavior in classroom instruction and that as teachers increase the cognitive structure of their instruction, the cognitive behavior of their pupils will similarly increase (117, p. 83).

Konetski (65) in 1970 utilized two instructional strategies with pre-service science teachers. Both procedures included the same programmed instructional booklet which both groups of students read. Following the reading of the program, the training strategies differed. The experimental
pre-service teachers received six hours of instruction during which they got a handout describing the categories of questions which Konetski had arranged as cognitive-memory or convergent and divergent and evaluative. The experimental group also categorized and designed their own questions. They conferred with the instructor of the science teaching methods course they were enrolled in during their micro-teaching, and they discussed their questioning behavior. The control group did not receive formal instruction on questioning, but they did meet in conferences with the instructor during their micro-teaching and they received the handout on categories of questions used during the study.

Konetski compared the two instructional strategies with the number of divergent and evaluative questions asked, the proportion of divergent and evaluative questions asked, and total number of questions asked. His conclusions were that instruction in classifying questions asked and instruction in designing questions significantly and positively affected both the number and proportion of divergent and evaluative questions asked, instruction in the use of divergent and evaluative questions significantly and negatively affected the total number of questions asked and that conferences between the instructor and the pre-service science teachers were most effective in affecting the divergent and evaluative questions asked when used in conjunction with the formal instruction presented during his investigation.
Yet, despite these findings, Konetski interpreted the data as an indication that although teachers think they are asking thought-provoking questions, they are, for the most part, asking questions which require little more than recall (65, p. 2). He suggested that many students do not normally answer questions because they fear getting the wrong answer. If the teacher would ask divergent questions for which there are no prescribed answers, the fear of giving wrong replies might be partially eliminated (65, p. 15). This conclusion supports Gallagher's findings (43) that a slight increase in the number of divergent and evaluative questions asked by teachers produces a much larger number of divergent and evaluative responses from students (65, p. 13).

Using some of the relevant and functional aspects of the preceding studies as guidelines, the design for this study was developed and carried out.

A Review of the Analysis of Questioning Tactics As Reported in Writings Other Than Research Studies

In addition to the various studies cited in previous sections, the literature contains articles written by educators and researchers extolling the value of questioning as a viable classroom strategy.

Burkhart (19) said that teachers must personalize education by learning of students' values. "Questioning patterns are the instruments of informality which provide us with means for establishing interactive freedom" (19, p. 85).
To Burkhart, asking empathy questions evokes student values and lets the teacher get deeper into the feelings of her students. He indicated that the process of questioning has struggled to gain recognition in the past decade, but it has been rejected because teachers are expected to make sure the students always respond with correct answers (19, p. 86). Burkhart believes that the youth of today are abandoning traditions of the past. Teachers must change their methods of instruction, particularly in questioning techniques.

Dauterman (25) prescribes four steps for effective(150,723),(856,814) question. The teacher should call on the less demonstrative students, being careful to correct misinformation so that all in the class, experience some degree of success; sequence and pace questions carefully; avoid using questions which deal with insignificant content or which require fact-recall; and adapt questions to the particular subject areas under discussion (25, pp. 29-30). He feels that the verbal behavior of the teacher is the most important classroom variable, which affects student learning. Through effective questioning, the teacher challenges students to perform a variety of mental operations and to struggle with subject matter (25, p. 32).

Duke (31) and Fraenkel (39) advise that teachers must put careful thought and planning into the questions they ask, and learn to build on the responses of the students, to carry questioning to the highest level of critical thinking.
Duke emphasizes the correct use of the open-ended question for helping the student think aloud and become involved in the thought process, without having to second-guess the teacher as to what "correct" answer she wants. Said Duke:

"There has been too long an emphasis in the classroom on didactic and deductive approaches to instruction; more time needs to be spent on the hypothetical and the inductive, for it is in this area that the child begins to apply and evaluate the concepts with which he must work most of his life (31, p. 471)."

Fraenkel proposed a taxonomy of questions which teachers should use in planning their units of study and daily lessons (39, pp. 200-201). Questions should be developed at all levels. Teachers should not push their own ideas on the students, but encourage the students to present and defend their own reasons for thinking as they do (39, p. 201). Too few teachers, said Fraenkel, promote the highest level--heuristic learning, which he classified as creative thought and divergent thinking. "It is with this last level that teachers, in effect, enter the stratosphere. Teachers who encourage students to think heuristically breathe rarefied air" (39, p. 201).

Many students look upon the language of the printed text as an alien domain and thus are disadvantaged learners in the traditional school classroom, according to Gantt (45). He presents a discussion which illustrated that pupils emit verbal discourse cues, which may be identified by the teacher to help the students with their thinking. Gantt believes...
that pre-service and in-service teachers can be given guidelines for helping the disadvantaged learner develop his thought processes in the classroom (45, p. 16).

If the handling of question-answer class situations in high school teaching is going to change, the teachers and the administrators are going to have to be the first persons to make the change, says Lindman (68). Students have been conditioned too long by the classroom environment and especially teacher attitudes to believe that total ignorance will be more easily forgiven than a wrong answer. "Students consider question-answer sessions as evaluative situations (and quite often they are) and therefore cloak themselves in the safety of silence" (68, p. 211). Students are also greatly influenced by their peers and would rather remain silent than risk rejection by their group (68, p. 212). Lindman challenged teachers to seek ways of reinforcing desirable behavior in students and be aware of factors that penalize it. He cautioned teachers to be good listeners and not to accept vague or erroneous statements from students. Especially should the teacher not rush student oral response or become preoccupied with other matters. "One of the most effective methods to discourage student oral response is to appear impatient or hurried" (68, p. 214). Lindman believes that it is the teacher who must take the initiative to sharpen and evaluate oral response in the classroom.
Piercey (90) suggests that not only should teachers be trained to ask questions effectively but that they should also act as models from whom students can be trained to ask questions. She believes that if the teacher identifies the art of questioning as a learning technique and guides the students in its use, the school can be changed from a place where students are taught to answer to a place where they learn to question and solve problems on their own initiative (90, p. 2).

Like Piercey, Thompson (111) says that students do not know how to ask questions. "... it is through questioning, not through knowing, that knowledge acquires its vitality" (111, p. 468). Thompson attributes the shortcomings of America's educational system to too much emphasis upon the didactic and deductive approaches used in teaching and not enough emphasis upon the hypothetical and inductive methods. To Thompson, teachers spend too much attention to answers and too little attention to questions (111, pp. 467-468).

Saadeh (96) reviewed studies dealing with developing critical thinking abilities in people and concluded that such abilities can be developed through the processes of logical operations which include high level inquiry methods. He ascertained that subject matter, knowledge, and intelligence are not enough by themselves to develop in people the ability to think critically. Any meaningful approach used
by teachers that challenges students and involves them in thought processes, underlying understandings, is a good procedure to develop critical thinking (96, p. 90). Critical thinking ability is not obtained by persons simply through a growing-up process, as studies cited by Saadeh (96, p. 88) show. Persons must be taught to think critically and proper questioning tactics by the teacher is one method to achieve this goal.

Thus, general writings on classroom questioning, as well as research studies dealing with the technique, indicate that educators and researchers are concerned with how teachers use questions and what effect teacher questions will have on student behavior. The next section reviews some current teaching methods textbooks, in their treatment of questioning in teaching, and discusses training programs designed to change teachers' questioning behavior.

A Review of Some Current Teaching Methods Textbooks Related to Questioning in Teaching and Specialized Training Programs Designed to Change Teachers' Questioning Behavior

The concept of questioning as a teaching device has been advocated by educators for many years. Blosser (15) reported that textbooks of general methods of teaching written in the late nineteenth century and early decades of this century contain descriptions of the use of questions in teaching. "In 1924 Odell . . . wrote 'The Use of the Question in the Classroom,'" a publication filled with
descriptive advice designed to help teachers improve their questioning practices" (15, p. 14). Blosser noted a decrease in emphasis on questioning in methods textbooks published in the 1940's and 50's, although some authors did devote portions of a chapter to questioning.

An examination of several relatively current texts, which are designed to be used in methods-of-teaching courses or as supplements to such course texts, reveals that all but two devote only a small portion of their coverage to questioning strategies.


Amidon and Hunter (6) in their 1966 paperback, Improving Teaching, described instruction as an interactive process and presented classroom situations in which the interaction was categorized by the Verbal Interaction Category System (VICS). Teacher-initiated talk coded on VICS includes questioning (6, p. 211) and several dialogues in the text are examples of how a teacher uses questions. These situations are accompanied by an analysis of teacher-pupil talk and a skill session in which the reader can practice the strategies used in the dialogue.
Flanders' 1970 book, *Analyzing Teacher Behavior*, covers the methodology of questions in twelve pages (36, pp. 292-303). He includes reference not only to questions that stimulate pupil initiative but also to the formulation, arranging, and sequencing of questions about feelings. One chapter contains a table, which presents categories for classifying questions about affect (36, p. 295), and a section on asking questions based on Hilda Taba's strategies (36, p. 301).

Grambs, Carr, and Fitch in their third edition of *Modern Methods in Secondary Education, 1970*, (47) expended only four pages, out of 457, to discuss the inquiry technique in teaching (47, pp. 180-183). They did devote a chapter to testing and evaluating procedures in which examples of questions are shown, but the emphasis is on the mechanics of test construction rather than question analysis.

Hough and Duncan (55), from their extensive research on classroom interaction using the *Observational System for Instructional Analysis*, have written a text, *Teaching: Description and Analysis, 1970*, (55). A detailed description is given to the process of analyzing teacher behaviors, using the OSIA. In Chapter Six, the authors included a section on "Solicitation of Student Behavior" (55, pp. 188-192) in which questioning is discussed as a teacher-initiated strategy. The section describes questions coded on the OSIA, as being either open or closed, and stresses the importance of
open ended questions to promote high-order, divergent thinking processes (55, pp. 191-192).

Hyman edited a text, which came out in 1968, entitled *Teaching: Vantage Points for Study* (61). In it a section on the "Cognitive Process" contains articles of researchers on verbal behavior and how questioning affects student response (61, pp. 79-146). This section concludes with a report on the research on classroom questioning by Davis and Tinsley (26), which was discussed earlier in this chapter.

Hyman wrote a text, published in 1970, entitled *Ways of Teaching* (62) in which he included an entire chapter on "The Art of Questioning," Chapter 9 (62, pp. 217-255). He drew heavily on the work of Gallagher and Aschner (43, 44), Bloom (14), and Sanders (97). Examples of actual classroom dialogue, with a discussion and analysis, are included for each level of the cognitive domain. Results of studies on questioning are given to exemplify how students' behavior may be affected by the questioning procedures of teachers. This treatment of questioning in teaching is one of the most comprehensive of any in recent texts reviewed and it was used in training the experimental student teachers in this study.

Jewett wrote a chapter on "The Problems Approach and the Senior High School" (63) as part of a 1971 NEA National Council for the Social Studies Curriculum Bulletin in which he makes a case for increased use of the inquiry method in teaching. By use of dialogue excerpts from classroom
situations, Jewett demonstrates how a teacher can destroy an intellectually permissive atmosphere in the class by stifling divergent responses from students (63, p. 64). He offers suggestions to improve the inquiry technique all the way from the question-asking and answering stage to full-fledged problem-solving situations.

Although the main purpose of their 1970 book, Classroom Management, Theory and Skill Training, is directed toward teachers' work with the affective domain of student behavior, Johnson and Bany (64) did write a section on "Guiding Class Discussion" (64, pp. 202-206) in which the effective use of questioning is discussed and an "Incident-Simulation" exercise is included. Students using this exercise are instructed in how to gain practice and skill in conducting class discussion and inquiry sessions.

In their paperback text published in 1970 on Systematic Instruction, Popham and Baker wrote three pages on questioning as one type of classroom transaction (91, pp. 102-104). The approach is very general and it centers on the mechanics of asking questions rather than on their content. The authors do refer to Sanders' Taxonomy and suggest that asking higher-level questions is one characteristic of an effective teacher.

The two texts which deal entirely with questioning are Groisser's How to Use the Fine Art of Questioning, 1964, (48) and Sanders' Classroom Questions, What Kinds? 1966, (97).
Groisser made no reference to studies of questioning techniques or of any scientific basis for his procedures, yet they follow the pattern suggested by others who have researched the issue. He uses sample dialogue in six subject areas to illustrate the sequencing of questions (48, pp. 15-20). Groisser's last chapter is a check-list for the teacher to guide his use of questions and it contains an exercise for the teacher to use in evaluating a hypothetical teacher's ability to question students (48, pp. 61-64). The criteria for effective questioning established by Groisser can still be utilized by teachers because they compare favorably with more recent research findings which show that proper use of questions does influence student achievement.

Sanders' Classroom Questions (97) has been widely used in recent years for research studies on questioning. It has also been used in seminars and teacher workshops and as a supplementary text in teaching methods courses. Although Sanders' examples are drawn from the social studies, one familiar with its contents can extrapolate to other subject areas. The author used Bloom's Taxonomy for his question categories and made modifications which resulted from five years of field development and evaluation in the public schools of Manitowoc, Wisconsin. Sanders devotes a separate chapter to each of the categories he designed. They are the basic six levels from Bloom's Cognitive Domain plus an extension of the lowest level which Bloom refers to as
knowledge. Sanders makes two levels here calling the lowest one "memory" and the second lowest, "translation." Each chapter of the text is devoted to one of the seven levels and it is written in a programmed fashion so that at the end of each chapter the reader answers questions on the chapter just read and can get immediate feedback as to the correctness of his responses by looking at the answer pages. In addition to the objective-type items used in the chapter questions, Sanders includes essay questions on the level of the mental process with which the chapter is concerned. The last two chapters deal with evaluation questions, suggestions for composing questions, and planning for questions. This text was the primary source used in the training of the experimental student teachers in this study. Its use will be explained in Chapter Three.

With few exceptions teaching-methods texts, even to the present (1972), still devote only a small portion of their space to any discussion of research on questioning or give procedures for improvement of questioning strategies by teachers.

A few specific programs have been implemented to train teachers to improve their questioning practices. Gall reported that in 1938 Houston developed an in-service program for the purpose of changing teachers' questioning methods (42, pp. 716-717). Houston used group conferences, stenographic reports of each teacher's lessons, supervisory
evaluation and self-analysis. Statistical analysis of eleven teachers trained by Houston indicated that most of them were able to bring about substantial changes in specific aspects of their questioning behavior.\(^{(42, \text{p. 717})}\).

Oliver and Shaver\(^{(87)}\), in 1964, trained teachers in the use of questioning skills appropriate to the discussion of controversial issues in the social studies.

Gall described the work of Taba and her associates, in 1964 and 1966, with their development of a system of teacher training, centered around questioning strategies, viewed as techniques which teachers could use to develop their students' abilities in forming concepts, explaining cause-and-effect relationships, and exploring implications.\(^{(42, \text{p. 718})}\). He also tells of the 1958 work of Suchman, who identified inquiry skills for science classes and trained teachers in their use. Suchman found the students of the trained teachers showed a significant increase in the number of questions they asked.\(^{(42, \text{p. 718})}\).

The most recent (1970) training program was developed, at the Far West Laboratory for Educational Research and Development, by Borg, Kelley, Langer, and Gall.\(^{(17)}\). It is called a minicourse and it is self-contained for an inservice training package requiring about fifteen hours to complete. The minicourse incorporates the use of video tapes and 16mm motion picture film, and relies on modeling, self-feedback, and microteaching, to effect behavioral change.
in teachers. Gall (42, p. 717) and Borg (16) reported that in a field test with forty-eight elementary school teachers, Minicourse 1 produced significant increases in the frequency of redirection questions, thought questions, and probing questions asked by the teachers. A refresher course was administered to one-third of the main field test sample two months after they had completed Minicourse 1 and they were compared to the other teachers who had no refresher course. Those who had this retraining did not perform significantly better on post-course tapes (16, p. 77). The data "... indicated not only that the Minicourse instructional model brings about substantial change in teacher performance but also suggest that these improvements are for the most part incorporated into the teachers' permanent repertorie of skills" (16, p. 79).

Chapter Eight, Minicourse 3, "Effective Questioning in A Classroom Discussion—Secondary Level" (17, pp. 128-138), puts considerable emphasis on framing higher cognitive questions. The authors feel that "High school pupils are capable of attaining much higher conceptual levels in discussions than are lower grade pupils" (17, p. 128). The main field test of Minicourse 3 was conducted from January through March of 1969, using seventy-four junior and senior high school teachers. Results showed that there was a significant increase in high-level cognitive questions asked by the
teachers who had been trained and that the proportion of teachers using high-level questions increased (17, p. 132).

The Far West Laboratory is now developing twenty additional minicourses dealing with tutoring, role-playing, lecturing, and the inquiry method, which includes training in questioning skills appropriate to particular teaching-learning contexts (42, p. 717).

A Statement of the Relationship of the Reported Research to This Study

A review of the literature as presented in this chapter may allow one to conclude that systematic observation can be validly and reliably used in research applications to analyze teacher effectiveness. It can help pre-service teachers, in training, change their classroom verbal behaviors. Medley has said of such procedures

If the project is dependent upon producing changes in teacher behavior—whether by better training or supervision, by introducing new curricula or materials, or by reorganizing the administrative structure to free teachers to teach better, systematic measurements of teacher behavior can make an important contribution to its success. Such measurements can tell you whether or not teachers do in fact teach differently after the change; they may tell you why some teachers improve and others get worse after the innovation; if they are used for feedback they can be instrumental in expediting the changes teachers need to make in their behavior if the new program is to achieve its goal (71, p. 50).

The survey of the literature also focused attention on efforts to identify teachers' questioning behaviors as they
relate to students' behavior and achievement and to ascertain if training in certain strategies could change behaviors of both teachers and students. Conclusions indicated changes could occur. Gall says that "...the value of focusing on teachers' questions is that they are the basic unit underlying most methods of classroom teaching. If this is true, then their continued study deserves the strong support of researchers" (42, p. 719).

Rosenshine (95) spoke of the importance of observational studies of "natural" teaching as being particularly useful for teacher education because they can identify certain behaviors or patterns of behavior which discriminate between high-achieving and low-achieving teachers. He feels that this type of research is necessary today.

But those educators involved in the pre-service and in-service education of teachers have a further question. They want to know whether training teachers to increase their use of certain behaviors and decrease their use of others results in changes in student product measures. To answer their question, experimental studies must be conducted and reviewed. But only certain types of experimental studies are relevant to their question, those in which (a) a number of teachers were trained to teach a class of students in a certain manner, (b) observational measures were obtained to verify that the teachers behaved as intended, and (c) end-of-experiment measures were obtained (such as achievement scores) (95, p. 7).

Rosenshine in reviewing the literature up to May 1970 found only four studies which met these criteria (95, p. 7).
The studies were discussed earlier in this chapter as were investigations which have followed in the two years since Rosenshine reported.

The research contained in this study was an attempt to meet the three criteria stipulated by Rosenshine in the above quotation. Since teacher questioning has been a major method of classroom discourse from the beginning of recorded educational history and since it is likely to remain a staple in classroom interaction, further study can be justified. Gallagher and Aschner (44) give reason enough for continuing research in inquiry techniques when they found that teachers generally ask cognitive-memory and convergent questions and that there is a relationship of the thought processes expressed by students to the types of teacher questions posed. They concluded that the kind of thinking young people engage in depends upon the kinds of questions teachers ask (44, pp. 191-192). Such being the case, the more researchers can find out about how and why teachers ask questions and can provide effective training procedures to alter less effective inquiry methods, the better will be America's schools.


a conference sponsored by Padogogisches Zentrum, Berlin, Germany, November 12-15, 1968 (ERIC, ED 040 113).


37. Flint, Shirley Helene, "The Relationship Between the Classroom Verbal Behavior of Student Teachers and the


46. Goodwin, Jr., William Earl, "The Relative Effects of Symbolic and Symbolic-Live Modeling on the Probing Questioning Behaviors of Selected Teacher Interns,"


50. Harris, B. M. and K. E. McIntyre, A Manual for Observing and Analyzing Classroom Instruction, Austin, Texas, The University of Texas at Austin, Extension and Field Service Bureau, 1964.


60. Hunter, Elizabeth, "The Effects of Training in the Use of New Science Programs Upon the Classroom Verbal Behavior of First Grade Teachers as They Teach Science," *Classroom Interaction Newsletter*, 4 (May, 1969), 5-11.


65. Konetski, Louis C., "Instruction on Questioning," paper read before the National Association for Research in Science Teaching Annual Meeting, Minneapolis, Minnesota, March 5-8, 1970 (ERIC, ED 040 040).

66. Kysilka, Marcella Louise, "The Verbal Teaching Behaviors of Mathematics and Social Studies Teachers in Eighth and Eleventh Grades," unpublished doctoral
dissertation, College of Education, University of Texas at Austin, Austin, Texas, 1969.


90. Piercey, Dorothy, "Questioning Strategies in Reading or Is the Right Person Being Trained to Ask the Questions?" paper read before a meeting of the Western College Reading Association, Los Angeles, California, April 1-3, 1971 (ERIC, ED 050 896).


101. Seagren, Alan T. and others, "The Impact of Student Teachers Upon the Attitude and Achievement of High School Students," Kansas City, Missouri, Mid-Continent Regional Education Laboratory, Incorporated, Lincoln, Nebraska, Teachers College, University of Nebraska, December 31, 1967 (ERIC, ED 026 346).


114. Trosky, Odarka S., "Modifications in Teachers' Questioning Behavior in the Development of Reading
Comprehension and A Series of Supervisory Conferences," paper read before the International Reading Association Meeting, Atlantic City, New Jersey, April 19-23, 1971 (ERIC, ED 051 982).


CHAPTER III

METHODS AND PROCEDURES

The purpose of this chapter is to describe the samples used in the study and explain the procedures for their selection, to relate the research methods employed to implement the study, and to delineate the procedures for treating the data. Five major areas are considered: (1) selection and description of the groups, (2) description of the instruments, (3) description of the study, (4) description of training in questioning techniques, and (5) procedures for treating the data.

Selection and Description of the Groups

The objective of this study, as previously indicated in Chapter I, was to investigate the effects training secondary teacher candidates in questioning strategies had on the achievement and critical thinking ability of their students. In order to measure change in these variables, standardized test instruments for specific subject areas had to be used. This fact, and requirements for statistical significance, dictated that the academic area, which had the most candidates for student teaching during the semester in which the study was carried out, had to be selected. Student teachers in the social studies provided the most candidates. All
secondary education teacher candidates in the social studies who were enrolled in the student teaching course during the spring semester of 1971, from a North Texas area private university School of Education, were used. Their schedules allowed them to remain in the public school on a half-day basis for the semester. A total number of sixteen student teachers met these conditions. These sixteen candidates had completed a minimum of eighteen semester hours in required social studies courses, nine semester hours in professional education courses and all had grade point averages of 2.5 or higher on a 4 point scale. They were all due to complete requirements for graduation and teacher certification at the end of their student teaching internship.

The general design and purpose of the study were explained to the candidates without revealing the treatment procedures. The student teachers were asked if they would participate in the investigation. All agreed to do so.

Authorization to conduct the study in a large North Texas metropolitan public school system was obtained from the Assistant Superintendent for Instruction, the Director of Research, the Director of Teacher Education, and the Consultant for Secondary Social Studies.

The sixteen student teachers were arbitrarily assigned by the Director of Teacher Education to eight senior high schools throughout the city. The schools represented varied socio-economic communities. One of them was a technical
high school located in the inner city. In five of the high schools, two student teachers each were assigned to the same supervising teacher because of a state law requiring a balance in supervising teacher-student teacher ratio. This resulted in a total of eleven supervising teachers for the sixteen student teachers.

Since the public school assignments of the student teachers could not be controlled for this study, after they were made, the student teachers were randomly selected, by use of a table of random numbers (10, pp. 381-385), for placement of eight in the experimental group and eight in the control group. Six of the eight experimental student teachers taught American History, one instructed World History, and one taught United States Government. Three of the control group student teachers were teaching American History, three were teaching World History, and two were teaching United States Government classes.

One supervising teacher who had a student teacher in the control group and one in the experimental group, each teaching a different area of social studies, requested that each student teacher include two classes in the study, rather than just one class, as was the case of the other student teachers. The request was honored and thus, the control group student teacher had two classes of World History and the experimental group student teacher had two classes of American History.
Table I shows the distribution of student teachers to supervising teachers by experimental and control groups and by subject area taught.

**TABLE I**

**DISTRIBUTION OF STUDENT TEACHERS TO SUPERVISING TEACHERS BY GROUP AND SUBJECT AREA**

<table>
<thead>
<tr>
<th>Supervising Teacher</th>
<th>Number of Student Teachers Assigned</th>
<th>Experimental Group</th>
<th>Subject Area</th>
<th>Control Group</th>
<th>Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1</td>
<td>1</td>
<td>W.H.***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>2</td>
<td>1</td>
<td>A.H.*</td>
<td>1</td>
<td>W.H.***</td>
</tr>
<tr>
<td>3.</td>
<td>2</td>
<td>1</td>
<td>A.H.*</td>
<td>1</td>
<td>Gov't.**</td>
</tr>
<tr>
<td>4.</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>W.H.***</td>
</tr>
<tr>
<td>5.</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>W.H.***</td>
</tr>
<tr>
<td>6.</td>
<td>2</td>
<td>2</td>
<td>A.H.*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>1</td>
<td>1</td>
<td>A.H.*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
<td>A.H.*</td>
</tr>
<tr>
<td>9.</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>A.H.*</td>
</tr>
<tr>
<td>10.</td>
<td>1</td>
<td>1</td>
<td>A.H.*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>2</td>
<td>1</td>
<td>Gov't.**</td>
<td>1</td>
<td>Gov't.**</td>
</tr>
</tbody>
</table>

*American History

**United States Government

***World History
Examination of Table I reveals that three of the eleven supervising teachers were assigned two student teachers each, one of whom was in the control group and one of whom was in the experimental group. Of the remaining two supervising teachers who had two student teachers each, one teacher had two student teachers in the experimental group and the other had two student teachers in the control group. In none of these cases were the student teachers present at the same time nor did they teach the same class groups. One was present in the morning and the other during the afternoon.

The assignment of student teachers and random selection of them for the control and experimental groups determined the number of public school students from which data would be gathered and analyzed for each group. Initially the control group had 275 students and the experimental group had 263 for a total of 538. Because of student dropouts and uncompleted or unusable tests, the final number used in the data analysis was 263 for the control group and 258 for the experimental group with the total equal to 521. Of these subjects in the control group, three were in the ninth grade, ninety-one were tenth graders, 102 were eleventh graders and sixty-seven were in the twelfth grade. Ninety-six of these control group students were taking American History, 116 were taking World History, and fifty-one were enrolled in United States Government classes.
The experimental group had a distribution of one ninth grader, thirty-four tenth graders, 15 in the eleventh grade and sixty-nine in the twelfth grade. Of these, 193 were taking American History, thirty-three were in World History classes, and thirty-two were enrolled in United States Government classes.

Not all students in either group took or completed all three tests which were administered them as part of the study. A complete description and illustrations of the distributions by tests will be treated in detail in the next chapter.

Description of Instruments

To establish a measure of the control variable and to measure the independent and dependent variables as well as to obtain the student teachers' reaction to their supervision, the following instruments were utilized in the study:

1. The Iowa Tests of Educational Development, Test 5, Ability to Interpret Reading Materials in the Social Studies. To measure the students' initial knowledge of social studies for the control variable, Form Y-4 was used as a pre-test. Form X-4 was administered as a post-test to establish a criterion variable of achievement.

The instrument was chosen because the eighty multiple-choice items are structured to test higher mental processes
beyond recall of facts. The ITED manual for teachers and counselors lists the skills measured by Test 5 as the ability to: (a) comprehend what is stated in a selection, (b) interpret what is implied in a selection, and (c) analyze and evaluate a selection critically (6, pp. 20-21). Students cannot rely on recall of previously learned facts or simply retrieve information contained in each of eight passages in order to answer questions correctly. The eight reading passages each are followed by as few as eight and as many as ten multiple-choice items from which the student picks the most correct answer. The manual describing the objectives of Test 5 of the battery states that

The eight passages in the test concern topics drawn from the areas of geography, sociology, contemporary social problems, economics, political science, and history. In general, the selections tend to deal with ideas rather than things, with matters of opinion rather than matters of fact. . . . The most important thoughts are often implied and the reader is frequently called upon to make qualitative, subjective evaluations (6, p. 22).

The Iowa Test battery is widely known, used, and respected as a valid measure of the general educational development of high school students. Since a current objective in education is to teach people to think critically rather than just be able to recall facts, the Iowa Test 5 was chosen because it does measure higher cognitive processes than memory only.

2. The Sequential Tests of Educational Progress--Social Studies. To measure other dimensions of the students'
initial knowledge of social studies for the control variable, Form 2B was used as a pre-test. Form 2A served as a post-test to establish a criterion variable of achievement.

As with the ITED, the STEP battery was developed to measure broad outcomes of general education rather than narrow results of a subject-matter area. The manual for interpreting STEP scores says of the battery, "STEP focuses on skill in solving new problems on the basis of information learned, rather than on ability to handle only lesson material" (14, p. 5).

The social studies examination of STEP is a seventy-item multiple-choice test which uses maps, charts, graphs, cartoons, pictures, diagrams, and the printed word to test competencies. STEP items are written so that students must

1. Identify generalizations, main points, and central issues.
2. Identify, compare, and contrast underlying values, attitudes, assumptions, biases, and motives.
3. Distinguish fact from opinion and recognize propaganda.
4. Assess the adequacy of data with respect to its relevancy, sufficiency, verifiability, and consistency.
5. Compare and contrast data.
6. Apply appropriate outside information and criteria.
7. Draw valid generalizations and conclusions (14, p. 7).

This instrument also measures understandings of the ways social change affects man's ways of living. Content is drawn from all of the social sciences: history, geography, economics, government, and sociology.
The manual summarizes the STEP social studies test by saying that it, "Measures social studies understandings, abilities to read and interpret social studies material . . ., skills in seeing relationships among basic facts, trends, and concepts, and ability to analyze such material critically" (14, p. 5).

Since measurement of cognitive processes on a higher level than factual recall was important to this study, the STEP social studies examination was chosen to compliment the ITED in measuring additional aspects of students' skills, understandings, and abilities.

3. Watson-Glaser Critical Thinking Appraisal. As a measure of students' initial critical thinking ability, Form ZM was given as a pre-test to measure the control variable. Form YM was administered as a post-test to establish a criterion variable of critical thinking ability.

The instrument consists of 100 objective items divided into five subtests which measure inference, recognition of assumptions, deduction, interpretation and evaluation of arguments. These are considered to be important abilities involved in critical thinking.

Watson and Glaser view critical thinking as a composite of attitudes, knowledge, and skills.

This composite includes: (1) attitudes of inquiry that involve an ability to recognize the existence of problems and an acceptance of the general need for evidence in support of what is
asserted to be true; (2) knowledge of the nature of valid inferences, abstractions, and generalizations in which the weight or accuracy of different kinds of evidence are logically determined; and (3) skills in employing and applying the above attitudes and knowledge (16, p. 10).

Various studies are cited by the authors (16, pp. 15-16) in which the Critical Thinking Appraisal has been used in research applications.

As the ability to think critically is an objective of schooling and is considered more relevant today than ever before, the Critical Thinking Appraisal test was chosen to measure student teachers' effects on this skill in their students.

4. The Observation Schedule and Record, Form 5, Verbal (OSCAR 5V). This systematic observation record was chosen to retrieve information concerning the student teachers' levels of questioning, primarily, along with other verbal behaviors displayed by them in the classroom. OSCAR 5V is a sign system which can record sixty-eight different events (9, p. 10). It is the latest revision in the OSCAR series and is designed to record only verbal behaviors. The use of simple, objective criteria for discriminating categories from each other in recording teacher behaviors makes the OSCAR 5V especially suitable for this study as only one observer was used. Medley and his associates (9) designed OSCAR 5V to be used by one observer and "... the observer's task is to put himself in the pupil's shoes and attend
to only those aspects of teacher behavior a pupil would see" (9, p. 3). The accuracy of the records obtained by using OSCAR 5V is further enhanced by the fact that the observer need attend only to the teacher. "His primary concern is with teacher verbalizations only; pupil utterances need concern him only when the teacher is silent and is listening to them" (9, p. 6).

As to reliability of codings by one observer only, Medley designed OSCAR 5V so "... that perfect agreement between different observers may not be an altogether desirable thing" (9, p. 4). He says that if two observers coding the same teacher behavior at the same time disagree as to the behavior it is quite possible that half of the students might also disagree on that behavior. He contends "... that the results obtained when the observers disagree would describe the teacher's behavior better than the results obtained if they agreed" (9, p. 4). He declares "... that disagreements will arise from ambiguities in the behavior of the teachers rather than from misunderstandings of the category system on the part of the observers" (9, p. 4). Thus, OSCAR 5V was chosen for this study in which a sole observer coded student teacher verbal behavior.

The coder's job with OSCAR 5V is to classify and record teacher utterances as "entries" and "exits." "The first teacher utterance ... is called an entry; the last one, his reaction to what the pupil has said, is called an exit."
The whole series of these behaviors is called an event" (9, p. 7). The Observation Schedule provides for two kinds of events which are recorded in different ways. One is called a "statement," an event containing only one utterance, which is recorded as an entry; the other is an "interchange" which is an event containing both an entry and an exit. The entry sets up a pupil utterance to be evaluated by the teacher and this evaluation constitutes the exit (9, p. 7).

Space is provided on the OScAR 5V form for recording each event in terms of what the authors call the "first word" and the "second word" each of which contains nine symbols representing nine categories describing pupil and teacher behaviors. "The first word . . . is used primarily for recording entries to interchanges. The second word is used for recording either entries to statements or exits from interchanges" (9, p. 8).

Both types of events, statements or interchanges, are recorded by using tally marks as a single event. Statements are classified in a one-way scheme according to the entry and only one tally mark is recorded for each one. Interchanges are classified in a two-way scheme, according to entry and exit, and two tally marks are made to record each interchange. The latter procedure results in a set of tallies which are counted as one for purposes of identifying and enumerating behaviors (9, p. 8).
Permission was obtained to use the OScAR 5V form for this study and to reproduce it for use in coding the student teachers.

5. Student Teacher Response on Supervision. (See Appendix.) This questionnaire, designed for the study, contains a total of thirty-two items in two parts. Part A, "The Supervising Teacher's Assessment of My Student Teaching," has twenty Yes-No items. Within these questions, items eight through twelve were written to find out if the supervising teacher had any influence on the student teacher concerning questioning techniques since such strategies were an important aspect of the study.

Part B, "The College Supervisor's Methods of Supervising My Student Teaching," contains twelve Yes-No items designed to yield information on how the student teacher felt he or she was guided by the college supervisor. Responses to this part of the questionnaire were incidental to the study.

Description of the Study

The sixteen student teachers entered their assigned schools at the beginning of the university spring semester in 1971. They and their supervising teachers had been briefed on the general design and procedures of the study with the exception of the treatment which would be given one group. They were told that they had been randomly assigned
to a group. The eight student teachers in the control group were referred to as "Group One" and the eight in the experimental group were "Group Two." Neither the student teachers nor the supervising teachers knew which group was control and which was experimental. The student teachers were told not to discuss seminar sessions with members of the other group.

Agreement had been reached, in preparing for the study, that all student teachers would spend the first two weeks after entering the schools in observing and choosing the classes that would be used in the investigation. During this two week period all sixteen student teachers met three times on the university campus with the college supervisor who also conducted the research. They met as one group for these three sessions and were given handouts which included directions for administering the three pre-tests: the Iowa Tests of Educational Development, Test S, Form Y-4 (5), the Sequential Tests of Educational Progress--Social Studies, Form 2B (13), and the Watson-Glaser Critical Thinking Appraisal, Form ZM (16). Purposes for the tests, procedures for administering them, and the importance of the examinations were emphasized. In the last meeting prior to testing, the student teachers were given the pre-test booklets, answer sheets, pencils and other supplies and a schedule was approved for administering the tests.
During the third week the student teachers were in the schools they gave the tests. Since three pre-tests had to be given and because two of them—the Iowa Test 5 and the STEP test—had to be given in two parts on two separate days, the entire school week was needed for pre-testing. One supervising teacher who had two control group student teachers would not allow a week to be devoted to pre-testing and another week for post-testing, so the STEP examination was arbitrarily omitted in those two classes as it was decided that the Iowa Test 5 was more necessary for measuring the skills and understandings deemed significant for the dependent variables in the study.

The student teachers were instructed in administering certain tests to the effect that when they gave the first part of a two-part examination on one day, they must give the second part on the following day. Other than this provision no order of test administration was specified. The entire class period was needed to give the examinations each day. During the pre-test week, the college supervisor visited the student teachers to direct the testing. At the conclusion of the pre-testing period, all materials were returned to the college supervisor for hand scoring.

Starting with the fourth week the student teachers were in the schools (the week following the pre-testing period), they assumed their teaching responsibilities which would continue for the remainder of the semester. Fourteen
student teachers taught one class each and two student teachers, one in the experimental group and one in the control group, taught two classes each. During this same week the college supervisor started observing each student teacher once a week coding interaction on the OScAR 5V. A separate form was used for each student teacher each week which yielded eight codings for each student teacher.

Although previous studies using the Observation Schedule and other observation systems have limited the coding period to around fifteen, twenty, or thirty minutes per session, since frequency of questioning was an integral aspect of this study, students were coded all period at each visit. Questioning activities in most cases continued throughout the period, so a more accurate measure of the student teachers' questioning behaviors was obtained by observing and coding the entire period each time. The class periods for all student teachers were of the same length, fifty minutes in duration.

Any and all of the sixty-eight events which can be recorded on OScAR 5V were coded if they occurred, but special attention was paid to Convergent (CVG) and Divergent (DVG) questions asked by student teachers. Medley defined the convergent question as one which calls for one right answer, while the divergent question allows for more than one answer which may be acceptable or correct (9, p. 11). The emphasis on recording questions of these two types was for the purpose
of measuring for and later testing Hypothesis Three of the study.

Beginning with the fourth week, also, separate seminars were held for the control and experimental groups by the college supervisor. These meetings lasted from an hour-and-a-half to two hours. During these sessions student teachers in the control group received standard feedback from their peers and the college supervisor concerning teaching methods, planning, and management of classroom routine. No reference was made to the control student teachers' questioning techniques. They were not shown their OSCAR 5V codings. The experimental student teachers did receive coding feedback.

Training of the experimental group of student teachers in questioning techniques began the fifth week they were in the schools—their second week of actual classroom teaching. Details of the training are described in the next section of this chapter.

Each group continued to meet separately and on a weekly basis for an hour and a half to two hours with the college supervisor. The control group continued to receive standard feedback and the seminars consisted of the control student teachers relating experiences they had during the week with comments from their peers and the college supervisor. No specific feedback concerning behaviors recorded on the OSCAR 5V form was discussed with the control group.
student teachers. The forms on which their verbal behaviors were coded were not shown them.

During the weeks the experimental group student teachers received training in inquiry techniques, feedback from the college supervisor was given at the weekly seminar meetings and in individual conferences with the student teachers. They were shown their codings on each OSCAR 5V form with interpretations made by the college supervisor. Special emphasis was drawn to their patterns of questioning and they were encouraged to increase the number of divergent questions asked. If there was progress in this direction, it was reported to the experimental student teachers at succeeding seminar sessions.

Weekly visitations, codings and seminar meetings including training for the experimental student teachers continued until approximately one-and-a-half weeks before the student teachers were to conclude their internship in the public schools. A total of eight weeks was available for coding between pre- and post-tests. When the student teachers completed their teaching, they were given instructions and materials for administering the post-tests. Form X-4 of the Iowa Test 5, Form 2A of the STEP examination, and Form YM of the Critical Thinking Appraisal were used for the post-tests. One week was allowed for post-testing and the same procedures used for administering the pre-tests were followed in giving post-tests.
After all student teachers had completed the post-testing of their students they were given a copy of the Student Teacher Response on Supervision questionnaire to answer privately and return by mail to the college supervisor. Student teachers were instructed not to sign their names or identify themselves on the forms. Each questionnaire was coded by the college supervisor so that it could be identified as being from either a control or an experimental student teacher. All sixteen student teachers returned the completed questionnaire.

When all post-tests and student teacher questionnaires were received by the college supervisor and after all student teachers had completed their internship, they were informed of their role in the study and the control student teachers received information from their OScAR 5V codings. However, the results of the pre- and post-test scores and responses on the student teacher questionnaire were never made known to the student teachers.

Description of Training in Questioning Techniques

Student teachers in the experimental group were given twelve hours of training in questioning methods through eight one-and-a-half-hour seminars conducted by the college supervisor during the eight weeks the student teachers were actually teaching.
During the first two training seminars the student teachers were given handouts showing examples of questions written at each of the six levels in Bloom's Taxonomy of the Cognitive Domain (1). The Bloom schema was compared to the hierarchy developed at the University of Utah (2). A handout from the Utah publication, Frontiers of Thinking, illustrating questions in constitutional government at the various levels was given the student teachers. The Utah hierarchy contains four levels instead of six but it includes the same cognitive processes as Bloom's Taxonomy. Level I, the lowest cognitive classification, is retrieving, identifying, discriminating and perceiving; Level II is inferring, comparing and contrasting; Level III is defining, judging and evaluating; and Level IV, the highest cognitive classification, is creating and generalizing (2, fold-out sheet).

By the third and fourth seminars, the experimental student teachers were instructed to have read Sanders' Classroom Questions, What Kinds? (12). These seminars emphasized Chapter Nine, "Planning for Questioning" (12, pp. 155-173). Student teachers were asked to include questions on all seven levels of the Sanders' schema in their lesson plans which were checked by the college supervisor on his visits to the classrooms.

The fifth and sixth meetings focused on the Gallagher-Aschner classification of cognitive-memory, convergent and
divergent questions (3, pp. 186-187). Chapter Nine, "The Art of Questioning," which appears in Hyman's book, Ways of Teaching (4, pp. 217-255) also draws heavily on Gallagher and Aschner. Examples of convergent and divergent questions from Hyman (4, pp. 227-229) were put on overhead transparencies and shown to the experimental student teachers. A discussion followed in which the student teachers were asked to formulate examples in their own areas and with units they were teaching. Stress was put on asking divergent types of questions. Lesson plans and materials formulated and used by the student teachers were checked by the college supervisor for the presence of high-level, divergent questions.

The eighth and final training seminar included a summation of previous sessions and a reemphasis of the importance of asking high-level, divergent questions. Progress reports on changing patterns of questions by the experimental student teachers were given at each seminar. In the last meeting, the college supervisor showed each student teacher his cumulative profile of verbal behaviors as recorded on OScAR 5V. Attention was called to the frequency of convergent and divergent questions asked over the preceding weeks.

The student teachers had approximately five more days, a school week, to teach following the last training session. They were each coded during this week on OScAR 5V forms.
Procedures for Treating the Data

After the pre- and post-tests were hand scored, the raw scores on the Iowa Tests were converted to "standard scores" using conversion tables (8, pp. 1-2) and percentile bands using tables in the interpretive supplement (7, pp. 9-15) in accordance with the publisher's recommendations (6, pp. 33-34). The Sequential Test raw scores were changed to "converted scores" by using the appropriate table from the 1962 SCAT-STEP Supplement (15, p. 46) and percentile bands using the tables from the STEP Manual for Interpreting Scores (14, pp. 22-23). This procedure was in keeping with the publisher's recommendations (13, p. 9; 14, pp. 9-14).

Raw scores on the Watson-Glaser Critical Thinking Appraisal were also recorded as percential ranks and stanines using appropriate tables for each form (16, p. 5) as recommended by the authors.

Answer sheets of tests which students failed to complete, of which they did not take one part (of a two part test), or which were obviously marked by blind guessing, were eliminated. The answer sheets of students who did not take a pre- or post-test were also not used. Of the total number of students taking the Iowa Test of Educational Development, Test 5, 220 in the control group and 212 in the experimental group were used. From the Sequential Tests of Educational Progress--Social Studies, 179 were usable.
in the control group and 227 in the experimental group. Two hundred and forty six Critical Thinking Appraisals could be analyzed from the control group and 218 from the experimental group.

No control could be maintained over where the student teachers were assigned and thus no balancing could be experimentally made to equate public school students in the control and experimental groups. Intact student samples were used. To make the groups statistically equal an analysis of covariance as described by Popham (10, pp. 221-256) and Roscoe (11, pp. 254-263) was employed.

Standard scores on the ITED, converted scores on the STEP and raw scores on the Critical Thinking Appraisal were submitted to the North Texas State University Computer facility for treatment of data using the analysis of covariance program (revised, May, 1971) with the pre-tests being the covariants. The result of using this statistical procedure is equivalent to matching the control and experimental groups with respect to the variables being controlled (pre-tests). Tables and analyses of the statistics on the three tests used in this study are included in Chapter IV.

Codings of the frequency of convergent and divergent questions recorded on OSCAR 5V for all student teachers in the study were analyzed by using the chi-square tests of independence. Such tests "... are used to determine whether an observed frequency distribution departs
significantly from a hypothesized frequency distribution" (11, p. 190). Popham (10, pp. 277-279; 291-296) and Roscoe (11, pp. 196-201) describe the rationale, function, computation, and interpretation of the procedure. A table and analysis of the chi-square tests are included in the next chapter.

Student teachers' reactions recorded on the Student Teacher Response on Supervision questionnaire were tallied separately for control and experimental groups. Items eight through twelve of Part A dealing with the supervising teachers' effect on student teachers' questioning behavior are shown in a table and analyzed in Chapter IV.

Procedures for administering and scoring tests used in this study and for coding and tabulating verbal behavior frequencies on OScAR SY were followed precisely as recommended by the publishers and authors of the instruments.
CHAPTER BIBLIOGRAPHY


2. Frontiers of Thinking (author not given), A Faculty Publication of the William M. Stewart School, University of Utah, Salt Lake City, Utah (no copyright date).


CHAPTER IV

STATISTICAL TREATMENT AND ANALYSIS OF THE DATA

Statistical Treatment

The basic purpose of this study, as was reported in Chapters I and III, was to ascertain whether training a group of student teachers in questioning strategies would result in increased achievement and critical thinking abilities of their students and a change in the questioning patterns of the trained student teachers.

Pre-tests of initial knowledge in social studies and critical thinking ability of the public school students served as the control or relevant variables. Post-test scores were the criterion or dependent variables. Frequency tallies of convergent and divergent questions asked by the student teachers were a measure of the independent variable.

The statistical computations necessary to test hypotheses one and two dealing with pupil achievement and critical thinking ability were performed by the North Texas State University Computing Center and the .05 level of confidence was selected as a minimum for significance to accept the three research hypotheses.
Null Hypothesis One stated that for students in the experimental group, there would not be significantly higher adjusted group means on two post-tests of achievement than those for students in the control group.

To test this hypothesis, group means from pre- and post-tests on equivalent forms of two standardized social studies examinations were gathered. Since the control and experimental groups of public school students could not be matched, the original criterion means for the groups were adjusted to compensate for initial differences between the groups on the control variables (2, p. 227). The pre- and post-test group means and the adjusted means for the experimental and control groups on one of the social studies tests, the Iowa Tests of Educational Development, Test 5, are shown in Table II.

### TABLE II

<table>
<thead>
<tr>
<th>Source</th>
<th>N</th>
<th>Criterion Post-Achievement</th>
<th>Control Pre-Test Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Unadjusted</td>
<td>Adjusted</td>
</tr>
<tr>
<td>Experimental</td>
<td>212</td>
<td>15.29</td>
<td>15.46</td>
</tr>
<tr>
<td>group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>220</td>
<td>14.47</td>
<td>14.31</td>
</tr>
<tr>
<td>group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>432</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

SOCIAL STUDIES STUDENTS' CRITERION AND CONTROL VARIABLE MEANS FOR THE IOWA TESTS OF EDUCATIONAL DEVELOPMENT, TEST 5
The experimental group pre-test mean on the ITED was 16.14 and the unadjusted post-test mean decreased to 15.46. The control group pre-test ITED mean was 15.70 and the unadjusted post-test mean decreased to 14.31. The adjusted means on the criterion variable were obtained by analysis of covariance for the experimental and control groups. The adjusted criterion means were 15.29 for the experimental group and 14.47 for the control group. It was these adjusted means upon which the final analysis was based and from which statistical inferences were drawn.

A summary of the statistical analysis of the variable of pupils' social studies achievement on the Iowa Tests of Educational Development, Test 5, is revealed in Table III.

TABLE III
ANALYSIS OF COVARIANCE OF EXPERIMENTAL AND CONTROL SOCIAL STUDIES STUDENTS' ACHIEVEMENT PERFORMANCE ON THE IOWA TESTS OF EDUCATIONAL DEVELOPMENT, TEST 5

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>1</td>
<td>72,441</td>
<td>72,441</td>
<td>3,086*</td>
</tr>
<tr>
<td>Within</td>
<td>429</td>
<td>10,067.867</td>
<td>23,468</td>
<td>...</td>
</tr>
<tr>
<td>Total</td>
<td>430</td>
<td>10,140,308</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

* No significant difference.
The between sum of squares on the ITED scores was 72.441 with 1 degree of freedom, which gave a mean square of 72.441. The within sum of squares was 10,067.867 with 429 degrees of freedom, which yielded a mean square of 23.468. In comparing mean differences on the adjusted criterion variable between the experimental and control groups, an F ratio of 3.86 was required for the .05 level of significance, using 1 and 429 degrees of freedom.

The null hypothesis of no difference between groups on social studies achievement, as measured by the ITED, after adjusting with the covariate, was retained. The .05 level of confidence, which was established as the criterion for rejection, was not reached. A .079 level of confidence was reached, but this was greater than the .05 necessary for rejection.

A second measure of pupils' social studies achievement for testing Hypothesis One was obtained from administering the Sequential Tests of Educational Progress—Social Studies. The pre- and post-test group means and the adjusted means for the experimental and control groups on the STEP examinations are shown in Table IV.

The experimental group STEP pre-test mean was 276.33 and the unadjusted post-test mean decreased to 274.55. The control group STEP pre-test mean was 276.25, while the unadjusted post-test mean decreased to 275.89. The adjusted means on the criterion variable were obtained as a part of
the analysis of covariance for the experimental and control groups. The adjusted criterion means were 274.53 for the experimental group and 275.92 for the control group. These adjusted group means were used for the final analysis and were the basis for the statistical inferences which were drawn concerning the significance of the first hypothesis.

TABLE IV

SOCIAL STUDIES STUDENTS: CRITERION AND CONTROL VARIABLE MEANS FOR THE SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS--SOCIAL STUDIES

<table>
<thead>
<tr>
<th>Source</th>
<th>N</th>
<th>Criterion Post-Achievement</th>
<th>Control Pre-Test Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Un-adjusted</td>
</tr>
<tr>
<td>Experimental group</td>
<td>227</td>
<td>274.53</td>
<td>274.55</td>
</tr>
<tr>
<td>Control group</td>
<td>179</td>
<td>275.92</td>
<td>275.89</td>
</tr>
<tr>
<td>Total</td>
<td>406</td>
<td>. .</td>
<td>. .</td>
</tr>
</tbody>
</table>

The smaller number of students in the control group who took the STEP tests resulted, as was explained in Chapter III, because one of the supervising teachers who had two student teachers in the control group would not allow all three tests used in this study to be administered. The STEP test was considered to be the most expendable and was thus not given to students in these two classes.
A summary of the data from the analysis of covariance on the variable of pupils' social studies achievement, on the Sequential Tests of Educational Progress--Social Studies, is shown in Table V.

**TABLE V**

ANALYSIS OF COVARIANCE OF EXPERIMENTAL AND CONTROL SOCIAL STUDIES STUDENTS' ACHIEVEMENT PERFORMANCE ON THE SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS--SOCIAL STUDIES

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>1</td>
<td>194.089</td>
<td>194.089</td>
<td>1.812*</td>
</tr>
<tr>
<td>Within</td>
<td>403</td>
<td>43,145.429</td>
<td>107.060</td>
<td>...</td>
</tr>
<tr>
<td>Total</td>
<td>404</td>
<td>43,339.518</td>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>

*No significant difference.

The between sum of squares for the STEP was 194.089 with 1 degree of freedom, which gave a mean square of 194.089. The within sum of squares was 43,145.429 with 403 degrees of freedom, which yielded a mean square of 107.060. Upon comparing mean differences on the adjusted criterion variable between the experimental and control groups, an F ratio of 3.86 was required for significance using 1 and 403 degrees of freedom. Since the F ratio value was not equal to or greater than that needed for rejection at the .05 level of confidence, the null hypothesis of no difference between
groups on social studies achievement, as measured by the STEP tests, after adjusting with the covariate, was retained.

As the criteria for rejecting Null Hypothesis One on the ITED and STEP test scores were not reached, the null hypothesis was not rejected.

Null Hypothesis Two stated that for students in the experimental group, there would not be a significantly higher adjusted group mean on a post-test of critical thinking ability than that for students in the control group. Data from pre- and post-tests on equivalent forms of the Watson-Glaser Critical Thinking Appraisal were gathered and subjected to an analysis of covariance. The adjusted and unadjusted group means on the Critical Thinking Appraisal are shown in Table VI.

**TABLE VI**

SOCIAL STUDIES STUDENTS: CRITERION AND CONTROL VARIABLE MEANS FOR THE WATSON-GLASER CRITICAL THINKING APPRAISAL

<table>
<thead>
<tr>
<th>Source</th>
<th>N</th>
<th>Criterion</th>
<th></th>
<th>Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Post-Achievement</td>
<td></td>
<td>Pre-Test</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Un-adjusted</td>
<td>Achievement</td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>218</td>
<td>59.73</td>
<td>60.07</td>
<td>56.42</td>
<td></td>
</tr>
<tr>
<td>Control group</td>
<td>246</td>
<td>61.53</td>
<td>61.23</td>
<td>55.36</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>464</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The experimental group pre-test mean on the Critical Thinking Appraisal was 56.42 and the unadjusted post-test mean increased to 60.07. The control group pre-test Critical Thinking mean was 55.36 and the unadjusted post-test mean increased to 61.23. The adjusted means on the criterion variable were obtained for the experimental and control groups. The adjusted criterion means were 59.73 for the experimental group and 61.53 for the control group. These adjusted group means were used for the final analysis and were the basis for the statistical inferences which were drawn.

A summary of the data from the analysis of the covariance on the variable of pupils' critical thinking abilities as measured by the Watson-Glaser Critical Thinking Appraisal is shown in Table VII.

TABLE VII

ANALYSIS OF COVARIANCE OF EXPERIMENTAL AND CONTROL SOCIAL STUDIES STUDENTS' ACHIEVEMENT PERFORMANCE ON THE WATSON-GLASER CRITICAL THINKING APPRAISAL

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>1</td>
<td>373.863</td>
<td>373.863</td>
<td>5.171*</td>
</tr>
<tr>
<td>Within</td>
<td>461</td>
<td>33,326.625</td>
<td>72.292</td>
<td>...</td>
</tr>
<tr>
<td>Total</td>
<td>462</td>
<td>33,700.488</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant beyond the 0.05 level.
The between sum of squares was 373.863 with 1 degree of freedom, which gave a mean square of 373.863. The within sum of squares was 33,326.625 with 461 degrees of freedom, which yielded a mean square of 72.292. In comparing mean differences on the adjusted criterion variable between the control and experimental groups, an F ratio of 3.86 was required for significance, using 1 and 461 degrees of freedom. The obtained F ratio value was greater than that needed for rejection of the null hypothesis at the .05 level of confidence. Therefore, the hypothesis of no difference between groups on critical thinking ability, after adjusting with the covariate, was rejected. A significant difference in the adjusted criterion means was determined to exist. The control group was superior to the experimental group in critical thinking ability.

Null Hypothesis Three stated that for student teachers in the experimental group, there would not be a significantly greater frequency of high-level, divergent questions asked during the eight weeks between pre-testing and post-testing than there would be for student teachers in the control group for the same period.

To test this hypothesis, frequency tallies of divergent and convergent questions asked by all student teachers were obtained on the Observation Schedule and Record SV. Medley and his associates recommended what they call a Level I
interpretation on counts of how often events coded on OSCAR SV occur during a given period of observation and stated that such a procedure is most likely to be useful when quantitative measures of behavior are needed in a research project attempting to relate teacher behaviors to pupil learning (1, pp. 42-44).

The frequency of divergent and convergent questions asked by the experimental and control groups of student teachers are shown in Table VIII which is a bivariate or 2 by 2 contingency table to which chi-square tests of independence were applied (2, pp. 282, 293; 3, pp. 198-199).

<table>
<thead>
<tr>
<th>Source</th>
<th>Divergent Questions</th>
<th>Convergent Questions</th>
<th>Row Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>(221.7)</td>
<td>(293.2)</td>
<td>515</td>
</tr>
<tr>
<td></td>
<td>243</td>
<td>272</td>
<td></td>
</tr>
<tr>
<td>Control group</td>
<td>(193.2)</td>
<td>(255.7)</td>
<td>449</td>
</tr>
<tr>
<td></td>
<td>172</td>
<td>277</td>
<td></td>
</tr>
<tr>
<td>Column Subtotal</td>
<td>415</td>
<td>549</td>
<td>964 Total</td>
</tr>
</tbody>
</table>

*Expected frequencies in parentheses.

Degrees of Freedom = 1.

Chi-square = 7.31**.

**Significant beyond the 0.01 level.
The experimental group of student teachers asked 243 divergent questions where the expected frequency was 221.7; they asked 272 convergent questions with an expected frequency of 293.2. Total questions asked by the experimental student teachers were 515. The control group of student teachers asked 172 divergent questions where the expected frequency was 193.2; they asked 277 convergent questions with an expected frequency of 255.7. Total questions asked by the control student teachers were 449. Total of all divergent-convergent questions asked was 964. The obtained chi-square was 7.31. With 1 degree of freedom the chi-square value needed at the .05 level of confidence for rejection of the null hypothesis was 3.84. Since the obtained chi-square value was greater than this— in fact, greater than the .01 level of 6.63— the null hypothesis of independence between groups and types of questions asked was, therefore, rejected. A significant relationship between groups and types of questions asked was determined to exist.

At the conclusion of the study all student teachers were given a copy of the Student Teacher Response on Supervision questionnaire. (See Appendix.) They were asked to answer anonymously and return it to the college supervisor. The form contains twenty Yes-No questions in Part A dealing with supervising teachers' guidance of the student teachers. Five items asked about the supervising teachers' influence
on student teachers' questioning behavior. Table IX shows how the student teachers responded to these items.

### TABLE IX

RESPONSES OF STUDENT TEACHERS TO QUESTIONS* CONCERNING THE SUPERVISING TEACHERS' INFLUENCE ON THEIR CLASSROOM QUESTIONING BEHAVIOR

<table>
<thead>
<tr>
<th>Questions from Part A</th>
<th>Student Teacher Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control Group (N=8)</td>
</tr>
<tr>
<td>8. Did your supervising teacher make any specific reference to the types of questions you asked the students?</td>
<td>2 6 8</td>
</tr>
<tr>
<td>9. Did the supervising teacher refer to the different &quot;levels of questioning&quot;?</td>
<td>2 6 8</td>
</tr>
<tr>
<td>10. Did the supervising teacher suggest that you ask more questions calling for mental processes above the recall level of your students?</td>
<td>2 6 8</td>
</tr>
<tr>
<td>11. Did the supervising teacher give specific examples of &quot;higher level&quot; types of questions you should ask?</td>
<td>2 6 8</td>
</tr>
<tr>
<td>12. Do you think you changed your pattern of asking questions as a result of directions or suggestions from your supervising teacher?</td>
<td>2 6 8</td>
</tr>
</tbody>
</table>

*For the complete questionnaire form see Appendix.

**Y = Yes; N = No; T = Total.

***One abstention.
Two of the eight student teachers in the control group answered "Yes" to all five items indicating that their supervising teacher had talked to them specifically about questioning and had influenced them in the way they asked questions to the point of changing their patterns of questioning. The remaining six student teachers in the control group responded that no mention or suggested changes in questioning strategies were made to them by their supervising teacher and that the supervising teacher did not influence or change their questioning methods.

With the experimental students, two of the eight said that their supervising teacher made specific reference to the types of questions they asked their students. One of these two felt that the supervising teacher had changed the student teacher's pattern of asking questions. A third experimental student teacher responded that his or her supervising teacher gave specific examples of higher level types of questions the student teacher should ask the students. Yet, this same person indicated that the supervising teacher did not refer to different levels of questioning or suggest that the student teacher ask questions above the recall level. The remaining five experimental student teachers answered "No" to the five items indicating that their supervising teacher did not refer to questioning strategies, identify levels of questions, or try to get the student teacher to change his or her question-types asked.
the students. One experimental student teacher refused to answer questions ten through twelve because he or she felt that they implied "... that student teachers ask only low level questions and are incapable of asking higher level questions."

Part B has twelve Yes-No items concerned with how the student teachers were guided by the college supervisor. Responses to this section of the questionnaire were incidental to the study. They are, therefore, not included in Table IX or discussed.

No statistical analysis was made of the responses to the questionnaire. Implications will be discussed in Chapter V.

Analysis of the Data

The scores obtained from the Iowa Tests of Educational Development, Test J and the Sequential Tests of Educational Progress--Social Studies were measures of pupils' social studies achievement. These data, after statistical analysis, indicate that the experimental group did not differ from the control group to any significant degree on the variable of pupils' achievement in social studies knowledge as was predicted by Null Hypothesis One. Examination of Tables II and IV reveals that both the experimental and control groups had higher unadjusted mean scores on pre-tests of the ITED and the STEP than they did on the unadjusted post-test
means. Adjusted post-test means yielded a higher mean for the experimental group on the ITED, but it was not a significant difference; while the adjusted post-test STEP mean was higher for the control group, but not significantly. Gain on post-test means for both groups was expected. Factors that could have contributed to the negative effect of lower post-test scores can be suggested.

Both the Iowa Tests and the Sequential Tests had to be administered in two parts each on two successive days. Coupled with the Critical Thinking Appraisal administration, this meant that students spent five school days taking pre-tests and five more days taking post-tests. The students knew that scores from these examinations were not to be used in evaluating their achievement in school and would not appear on their school records. Since there was neither intrinsic nor extrinsic motivation for them to perform well, many made no real effort to try their best, especially on the post-tests. On some answer sheets a "blind-guessing" pattern was obvious. These tests were eliminated from data analysis. However, other students could have been more subtle in their guessing and thus their apathy toward taking the tests could not be detected by the way they answered.

A week or two prior to taking the post-tests for this study, the same students had to take another battery of standardized tests for records being kept by the school system in which the study was conducted. The importance of the
school-administered tests was emphasized over the post-tests for this study. At least two of the supervising teachers, although they allowed administration of the tests for this investigation, did not encourage their students to perform at their best. As was explained in Chapter III and again referred to earlier in this chapter, one supervising teacher who had two student teachers in the control group would not allow five days for pre-testing and five for post-testing, so one test had to be eliminated. Finally, the post-tests were given within three weeks of the close of the public school year and students were probably weary of school in general, and examinations in particular. It is known that several students stayed away from school after they took one post-test, until the testing was completed.

Scores on the Watson-Glaser Critical Thinking Appraisal provided a measure of students' critical thinking ability and data for testing Hypothesis Two. Examination of Table VI shows that there was an increase in unadjusted post-test means for both the experimental and the control groups, but the control group had a higher mean. When the post-test means were adjusted, a significant difference between groups was found to exist in favor of the control group.

The Critical Thinking Appraisal was administered in one period each for the pre- and post-tests. This factor could account for the increased mean post-test scores for both groups. Such gains were expected. The students were
probably more enthusiastic about a test of one period dura-
tion as compared to the other two given in two parts, each.

Frequency tallies of divergent and convergent questions
served as measures of student teachers' questioning patterns
and were used to test Hypothesis Three. After statistical
analysis, the data indicate that there was a significant
relationship between student teacher groups and types of
questions asked. An examination of Table VIII shows that
the experimental student teachers asked significantly more
divergent questions than did the control student teachers.
Both groups asked more convergent than divergent questions.
However, the experimental group asked fewer convergent
questions and more total questions than did the control
group of student teachers.

Data from the analysis of covariance on the Watson-
Glaser Critical Thinking Appraisal (Tables VI and VII) and
the chi-square tests on divergent and convergent questions
(Table VIII) indicate that an increase in divergent ques-
tions asked by the experimental student teachers produced
a decrement in critical thinking of their students. Perhaps
the emphasis in training experimental student teachers to
ask more divergent questions caused a shift in their atten-
tion to the structure of the questions rather than their
cognitive content. They may have been more concerned with
asking questions which allowed for more than one answer than
in listening to the quality of students' answers.
Examination of Table IX reveals that two of the eight experimental student teachers said that their supervising teacher did make specific references to the types of questions they asked their students. One of these two also said that he or she changed the pattern of questioning as a result of the supervising teacher's influence. Another experimental student teacher said that the supervising teacher gave specific examples of higher level types of questions he or she should ask but that this did not influence the student teacher to alter his or her method of asking questions.

The same two student teachers in the control group said that their supervising teacher did speak to them of the types of questions they asked their students, referred to levels of questioning above recall, and gave examples of higher-level questions the student teachers should ask their students. These student teachers also said they changed their questioning pattern as a result of the supervising teacher's influence. The other six student teachers in the control group indicated that their supervising teachers did not refer to the student teacher's questioning methods, speak of levels of questioning above recall, or give them specific examples of higher-level questions they should ask. None of the six felt that anything the supervising teacher said concerning questioning strategies influenced their own methods of asking questions.
Student teachers were told not to affix their names to the questionnaire so no comparison can be made among student teachers in either group, their responses on the questionnaire, and their actual classroom patterns of asking questions. It is possible that the two control student teachers who said they were told of different question types and methods by their supervising teacher and who think they changed their questioning patterns, did not interpret the terms "levels of questioning" or "higher-level" questions in the same manner as the terms were defined for this study. Thus, it is possible that the perceived changes of the two control student teachers were not the same as those recorded and tested for Hypothesis Three.

The fact that only two of eight student teachers in the control group and one of eight in the experimental group responded that their supervising teacher influenced their questioning style would tend to negate each other. Therefore, the influence of the supervising teachers in both groups on their student teacher's questioning behavior was minimal or non-existent.
CHAPTER BIBLIOGRAPHY


CHAPTER V

SUMMARY, FINDINGS, CONCLUSIONS, FACTORS BEARING UPON THE FINDINGS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

This chapter is a summary of the investigation, a discussion of the findings, the conclusions, factors that had a bearing on the findings, and recommendations for further research.

Summary, Findings, and Conclusions

The problem of this study was the effect of levels of questioning used by secondary social studies student teachers on pupil achievement and critical thinking ability. A sub-problem of the study was the effect of feedback from the university supervisor to student teachers on the latters' use of convergent and divergent levels of questioning. Purposes of the study were (1) to determine the effect of levels of questioning used on secondary public school students in social studies, as measured by their achievement scores and their critical thinking ability, (2) to determine the effect of feedback to student teachers on their patterns of asking convergent and divergent questions, and (3) to draw conclusions from the findings and develop implications concerning levels of questioning used by teachers, and the use of feedback from college supervisors to student teachers.
Subjects of this study were an experimental group of student teachers and their students and a control group of student teachers and their students. There were eight student teachers and 258 pupils in the experimental group and eight student teachers and 263 students in the control group. The student teachers were randomly divided into two equal classes and randomly assigned to the control and experimental groups. All of them taught a subject in social studies. They all were assigned to secondary schools in a metropolitan area.

Three research hypotheses were formulated. The .05 level of confidence was selected as a minimum criterion of significance to accept the hypotheses.

Hypothesis One

For students in the experimental group, the adjusted group means on two post-tests of achievement would be significantly higher than those for students in the control group.

Procedures.—Two different pre-tests in social studies were administered to establish the pupils' initial knowledge of social studies. Student teachers gave the pre-tests during the third week they were in the schools, and before they started teaching. For the next eight weeks, student teachers taught their assigned classes and their verbal behavior was coded on the Observation Schedule and Record 5V. Student
teachers in the experimental group received twelve hours of training in questioning techniques during this time for the purpose of developing high-order reasoning and critical thinking abilities in their students. The experimental student teachers also received feedback from the college supervisor on their codings and were informed of their patterns of asking questions. They were encouraged by the college supervisor to ask more high-level, divergent questions.

Two post-tests in social studies, using the equivalent forms of the examinations utilized for pre-testing, were administered by all student teachers to their students during the twelfth week of their internship—the week immediately following the conclusion of their actual teaching. These post-tests established the pupils' levels of achievement in social studies at the end of the semester. Data from the pre- and post-tests were analyzed by an analysis of covariance to determine if there was a significant difference on adjusted criteria (post-test) means between the experimental and the control groups.

Findings.—An F ratio of 3.86 was needed to reach the .05 level of significance. This level was not reached on the two social studies adjusted post-test means. Therefore, the findings indicated that the two groups did not differ
significantly on the two post-tests of achievement. Research Hypothesis One was rejected.

Conclusions.--In light of the findings of this phase of the study it cannot be expected that students will show significant gain in performance on social studies achievement, as measured by the Iowa Tests of Educational Development, Test 5 and the Sequential Tests of Educational Progress--Social Studies when taught by student teachers who received training in questioning techniques. These trained student teachers did not significantly increase the level of their students' social studies achievement, when compared to pupils of student teachers who received no training.

Hypothesis Two

For students in the experimental group, the adjusted group mean on a post-test of critical thinking ability would be significantly higher than that for students in the control group.

Procedures.--The Watson-Glaser Critical Thinking Appraisal test was chosen to measure pupils' critical thinking ability. Procedures for testing this variable, for training and coding student teachers' verbal behavior, for giving feedback to student teachers, and for analyzing the data were the same as described above for Hypothesis One.
Findings.--An F ratio of 3.86 was needed to reach the .05 level of significance. There was a significant difference beyond the .05 level of confidence on adjusted criteria group means between the experimental and control groups on the Watson-Glaser Critical Thinking Appraisal. The difference, however, was such that the control group had a higher adjusted post-test mean than did the experimental group of students. The findings indicated that the two groups did differ significantly on the post-test of critical thinking ability. Research Hypothesis Two was rejected, however, because the significant difference was in favor of the control rather than the experimental group as had been predicted. The control group proved to be superior to the experimental group in critical thinking ability.

Conclusions.--In light of the findings of this phase of the study it cannot be expected that students will show significant gain in performance on critical thinking ability, as measured by the Watson-Glaser Critical Thinking Appraisal, when taught by student teachers who received training in questioning techniques. These trained student teachers did not significantly increase the level of their students' critical thinking skills when compared to the pupils of student teachers who received no training. In fact, the data indicate that the training had a negative effect on experimental pupils' critical thinking.
Hypothesis Three

For student teachers in the experimental group, there would be a significantly greater frequency of high-level, divergent questions asked during the eight weeks between pre-testing and post-testing than there would be for student teachers in the control group for the same period.

Procedure.—The experimental student teachers were trained for twelve hours in techniques of questioning and were given feedback during the training period, as to their progress in asking high-level, divergent questions. All student teachers' classroom verbal behavior and interaction were coded once weekly on the Observation Schedule and Record 5V. The feedback to the experimental student teachers came from the OScAR 5V codings of frequency of convergent and divergent questions they asked. Interpretations were made of the codings and they were encouraged to increase their high-level, divergent questions. Eight codings were obtained for each student teacher in both the experimental and control groups.

After the student teachers completed their teaching, chi-square tests of independence were performed on the frequency of divergent and convergent questions asked by both the experimental and control student teachers. Percentages of convergent, divergent, and total questions asked were completed for both groups.
Findings.—A chi-square value of 3.84 was needed to accept the research hypothesis at the .05 level of confidence. Since the obtained value was greater, the research hypothesis concerning frequency of high-level, divergent questions asked by the experimental student teachers was retained. The findings indicated that a significant relationship between experimental and control groups and types of questions asked did exist.

Of the total 964 divergent-convergent questions asked by student teachers in both groups, the experimental student teachers asked 54 per cent of these. Of the total 549 convergent questions asked by student teachers in both groups, the experimental student teachers asked a smaller percentage, 49 per cent, compared to 51 per cent asked by the control student teachers. Concerning divergent questions, of the total 415 asked, the experimental student teachers asked more, 59 per cent, compared to 41 per cent divergent questions asked by the control student teachers. The experimental student teachers asked 47 per cent divergent and 53 per cent convergent questions while the control student teachers asked 38 per cent divergent and 62 per cent convergent questions. Even though more convergent than divergent questions were asked on the whole, the experimental student teachers asked fewer convergent questions and significantly more divergent questions than did the control student teachers.
Conclusions.—In light of the findings of this phase of the study it can be expected that, through training, student teachers can increase the number of high-level, divergent questions they ask. The total number of questions they ask can increase. As a result, student involvement and classroom verbal interaction may also intensify.

A prediction can also be made that as a result of feedback from the college supervisor to student teachers on the latters' use of various levels of questions, the student teachers can change the frequency and kinds of questions they ask in the classroom according to the kind of training they receive. However, as indicated by the findings, it cannot be predicted that student teachers who increase the frequency of divergent, high-level questions they ask their students would find a significant change in the level of achievement and critical thinking skills of the students. In fact, the data point to the conclusion that an increase in divergent questions, asked by the experimental student teachers, produced a decrease in critical thinking of their students. These student teachers may have been more concerned with the levels of questions than the response of the students. Perhaps, they did not evaluate the students' answers and give the proper follow up to improve discussion and increase cognitive skills as had been hypothesized. The control student teachers, on the other hand, may not have been especially aware of their question structure. They may
have paid more attention to students' answers and caused the students to raise their levels of cognition through guided discussion and inquiry.

An assumption of this study was that any feedback from supervising teachers to student teachers concerning the latter's levels of questioning would be infrequent and would not significantly affect the student teachers' questioning behavior. To obtain an indication of the validity of this assumption, at the conclusion of their internship, each student teacher was asked to answer questions concerning their supervision by the public school teacher with whom they interned and the college supervisor. This was accomplished by a questionnaire on which the student teacher made Yes-No responses. The student teachers were not identified by name. They were identified as being from either the control or the experimental group. Hidden among the questions were items for which the questionnaire was primarily designed. These five items dealt with what influence the supervising teacher had on the student teacher's questioning behavior. Two of eight student teachers in the control group and one of eight in the experimental group responded that the supervising teacher influenced his or her questioning behavior to the point of having them change their pattern of asking questions. The supervising teachers' effect on student teachers' questioning techniques was, therefore, considered to be minimal.
Factors Which Bear Upon the Findings and Recommendations for Further Research

As a result of this study, factors which could have affected the results and recommendations for controlling or eliminating them from replications of the study are discussed below.

1. The necessity of having to administer three pre- and post-tests, two of which had to be given in two parts on successive days, and which resulted in five days of pre-testing and five more for post-testing, had an influence on the outcomes.

A replication of the study should be conducted using only one test of achievement. Science Research Associates, publishers of the Iowa Tests of Educational Development, now have a shorter one-class-period version which would mean using only one period for the pre-test and one period for the post-test. Since the data from the study indicated a trend toward a significant difference in favor of the experimental group on the ITED, use of the shorter form of this examination is recommended.

2. The student teachers and supervising teachers were asked to explain the purpose of the study to their students and encourage them to perform at their best, so perhaps the control student teachers and their supervising teachers were a more positive motivating force on the students because the pattern on two of the tests was in the direction of higher
adjusted post-test means for students in the control group. They performed better on the Watson-Glaser Critical Thinking Appraisal and the Sequential Tests of Educational Progress.

It is recommended that a method of using both intrinsic and extrinsic motivation on the public school students, to get them to perform at their best on the tests, be incorporated by the student teachers and the supervising teachers. If the pupils could see the results of pre- and post-tests and have scores explained to them, perhaps they would take the tests more seriously. It is also suggested that if the students could be given some credit on their school record perhaps they would be more motivated to take the examinations in earnest.

3. The fact that students in both groups had to take other standardized tests required by the school system just before they took the post-tests for this study could have influenced their attitude toward taking the latter examinations.

A recommendation for anyone who replicates this study is to get public school officials to allow the researcher to use only those students who do not have to take other test batteries or to excuse the students in the investigation from having to take standardized test batteries other than the ones used in the study. Thus students would not be so apathetic to taking tests needed in the investigation. They could better appreciate the tests used in the study.
4. Perhaps the control variables of pre-test scores alone were not sufficient measures to establish matched groups even when the analysis of covariance was performed on test scores.

If intact groups have to be used and, thus, data treated by the analysis of covariance, it is recommended that other control variables in addition to pre-test scores be used as covariates. Traits such as students' intellectual aptitude (I.Q.) or previous scholastic performance (grade point average), which are unlikely to be influenced by experimental manipulation, could be used.

5. The subject-matter actually taught by the student teachers between pre- and post-tests may not have been closely enough related to the content of the social studies tests used to measure levels-of-achievement in students. In other words, the student teachers did not teach for the tests.

Further investigations of this nature should use instruments which will measure outcomes of objectives actually taught by the teachers in the classroom. Perhaps the researchers could develop and validate tests especially for the study.

6. Training of the experimental student teachers in techniques of questioning, although successful in increasing the frequency of divergent questions they asked, may not have been sufficient enough to help these student teachers
heed the answers given by the pupils. The student teachers in the experimental group may have been more concerned with trying to ask high-level questions and not aware or responsive to students' answers. They may not have known how to follow-up the answers to bring about high-order cognitive processes in the students through discussions.

Also, the twelve hours of training over an eight-week period may not have been long enough or effective enough for the experimental student teachers to apply their training in such a manner that it would affect the outcome of the post-tests. By the time the training reached the main emphasis on high-level, divergent questions, the student teachers had only ten to fifteen school days left in which to bring about change in their students.

It is recommended that training of student teachers in questioning methods include exercises wherein the student teachers actually write lesson plans and assignments using various levels of questions. These should be evaluated by their peers and the college supervisor in seminars and then actually carried out in the classroom. More emphasis should be put on how the student teachers respond to and follow up students' answers to all questions, especially high-level ones.

The period between pre- and post-testing should be increased to fifteen or sixteen weeks, if possible, so that student teachers who received training in question-asking
would have more time to put their training into practice and so that their students would have more time to internalize such teaching. This might possibly increase the public school students' levels of achievement and critical thinking ability.

7. The Watson-Glaser Critical Thinking Appraisal may not be a valid instrument for measuring convergent and divergent questions as defined in this study. It is possible for students to correctly answer divergent questions with little more than recall of facts and conversely, to answer convergent questions correctly by arriving at the answer through such processes as deduction and drawing conclusions, which are measures on the Critical Thinking Appraisal.

If the Watson-Glaser instrument is used in further investigations of this nature, perhaps a redefinition of types of questions needs to be formulated or a better distinction made between identifying divergent and convergent questions. It is suggested that divergent and convergent questions be classified as high-level types and that cognitive-memory types of questions be identified as low-level.

8. How much previous knowledge of and training in the methods of asking questions student teachers had received prior to their internship was not determined or controlled in this study. These factors may have been part of the training student teachers received in courses prior to their internship and could have influenced the results.
It is recommended that the researcher who attempts a study of this nature again, screen student teacher candidates to look for these factors and perhaps give them a questionnaire prior to being included in the study in order to establish prior training in and knowledge of questioning strategies.

9. Student teachers were asked to indicate on a questionnaire whether or not their supervising teachers had any influence on them in getting them to change their patterns of asking questions. Although only two in the control group and one in the experimental group said such influence had occurred, there was no precise way of finding out exactly how much influence the supervising teachers did have on the student teachers' question-asking methods.

It is recommended that more control be kept on how the supervising teachers influence what the student teachers teach and on how they affect the student teachers' methods of asking questions. For instance, in the experimental group, units and lesson plans could be agreed upon which would control the levels of questions asked and prescribe procedures for guiding class discussions. Supervising teachers could also be instructed to avoid suggesting how student teachers should question.

The researcher in further replications should work with public school authorities to make sure that no two student teachers are assigned to the same supervising teacher. This
is especially important if one student teacher happens to be assigned to the experimental group and the other in the control group as was the case in this study.

10. The instrument used for coding verbal interaction, especially the types of questions student teachers asked, might not have been definitive enough in identifying the exact nature of divergent and convergent questions. It may not have actually differentiated between high-level and low-level questions asked by the student teachers. The question types were noted by a tally mark, but unless the coder wrote down the exact wording of the questions—which was not the usual practice—feedback to the student teachers was probably too general. The pupils' responses to student teachers' questions were recorded on the form as being either substantive or nonsubstantive and this information was given to the student teachers in the experimental group. Such information, alone, was probably not adequate enough to help the student teachers guide their students in discussions which might have helped develop the latters' critical thinking skills.

It is recommended for further studies that an instrument other than the one employed in this investigation be used for coding student teachers' questions. Instruments such as the Teacher Question Inventory, the Florida Taxonomy of Cognitive Behavior, the Questioning Strategies Observation System, the Teacher Oral Question Observation
Schedule, the Question Analysis System, or the Observational System for Instructional Analysis, all of which have been referred to earlier in this paper, focus on questioning rather than on all possible verbal behaviors which transpire in the classroom. Such a form would provide for more specific and accurate identification of different levels in student teachers' questions and thus could enhance the effectiveness of the feedback from college supervisor to the student teacher, improve training sessions, and more validly measure question types.

It is further recommended that the coding of verbal interaction on a printed form be coupled with the recording of such interaction on audio tapes. In this way student teachers could actually hear themselves asking questions and could hear the students' responses and their own follow-up to these responses. Such a procedure would enrich the feedback from the coder and be much more effective in specifying levels of questions than trying to learn from a piece of paper with tally marks on it. Also, both the coder and student teacher could review certain tapes as much as necessary to identify more accurately the verbal interaction deemed important for the study.

11. Even with adequate training of student teachers, perhaps eight, or even fifteen or sixteen weeks is not long enough to teach students to think critically. Maybe such training can only be accomplished through all of the years
of a person's schooling with each teacher contributing to the ongoing process—a process which must be continued by individuals throughout their life.
APPENDIX

STUDENT TEACHER RESPONSE ON SUPERVISION

Below are questions relating to two aspects of your student teaching internships: (A) the nature and amount of assessment of your teaching given you by your supervising teacher and (B) the nature and amount of supervision given you by your college supervisor.

Please answer the questions honestly. Your supervising teacher or the principal will never see this instrument or know how you answered the questions. The college supervisor will not know to which supervising teacher you are responding nor will he know whose answers are on these sheets since neither your supervising teacher’s name nor your name is to appear on these pages.

Answer by drawing a circle around the "YES" or "NO" beside each question, depending upon how you feel about the item.

If you wish to make any comments not covered within the questions, write on the bottom and back of the last page. DO NOT MENTION NAMES, SIGN YOUR NAME, OR IDENTIFY YOURSELF IN ANY WAY.

A. The Supervising Teacher’s Assessment of My Student Teaching

1. Did your supervising teacher talk to you about how you were doing? ..... YES NO

2. Did your supervising teacher critique you at the beginning of your internship? ..... YES NO

3. Did he/she critique you regularly throughout your internship? ..... YES NO

4. Did he/she primarily critique you after you had concluded your student teaching? ..... YES NO

5. Were suggestions from your supervising teacher specific so that you could apply
them to your teaching before your internship was concluded? YES NO

6. Were the comments primarily negative criticism? YES NO

7. Did the supervising teacher point out methods, techniques, and/or presentation of subject matter that you used which he/she thought were particularly effective? YES NO

8. Did your supervising teacher make any specific reference to the types of questions you asked the students? YES NO

9. Did the supervising teacher refer to the different "levels of questioning"? YES NO

10. Did the supervising teacher suggest that you ask more questions calling for mental processes above the recall level of your students? YES NO

11. Did the supervising teacher give specific examples of "higher level" types of questions you should ask? YES NO

12. Do you think you changed your pattern of asking questions as a result of directions or suggestions from the supervising teacher? YES NO

13. Did the supervising teacher give you help when you needed it? YES NO

14. Do you feel the supervising teacher gave you enough "academic freedom" to plan and present material in the manner which best suited your training and philosophy of education? YES NO

15. Did the supervising teacher leave the room while you were teaching? YES NO

16. Was he/she regularly absent from the room while you were teaching? YES NO

17. Do you feel that the supervising teacher let you teach enough during the internship? YES NO
18. Did you get to teach 50 per cent or more of the time you were interning? ... YES NO

19. Do you think that the supervising teacher indoctrinated you into all phases of teachers' responsibilities so that you are prepared to meet these responsibilities as a beginning teacher yourself? ... YES NO

20. When compared with your college supervisor do you believe your supervising teacher was more successful and necessary for helping you and making your internship a success? ... YES NO

B. The College Supervisor's Methods of Supervising My Student Teaching

1. Was your college supervisor available or easily reached when you needed advice or help? ... YES NO

2. Do you feel the college supervisor visited your classroom enough while you were actually teaching? ... YES NO

3. Did you feel "threatened" by the presence of the college supervisor in your classroom? ... YES NO

4. Did your conferences with the college supervisor help you in solving your problems and improving your teaching? ... YES NO

5. Did you feel free to discuss any and all matters pertaining to your student teaching with your college supervisor even though it might include negative criticism of your supervising teacher? ... YES NO

6. Was the college supervisor consistent in the methods or techniques he used in working with you throughout your internship? ... YES NO

IF ANSWER TO NUMBER 6 IS "NO," ANSWER ITEMS 7 AND 8.
IF THE ANSWER TO NUMBER 6 IS "YES," SKIP TO ITEM 9.
7. If there was a change in the way the college supervisor worked with you, did it occur after the first few weeks of your internship? ...... YES NO

8. If there was a change in the way the college supervisor worked with you, was the latter change in technique more beneficial to you in helping you improve your teaching? ...... YES NO

9. Was the advice given you by your college supervisor too general to apply to your individual situation? ...... YES NO

10. Was the college supervisor helpful in "tying in" concepts and knowledge from education courses so that they would apply to actual practice in your classroom? ...... YES NO

11. Do you think a college supervisor who is a specialist in the subject area in which you taught would have been more helpful than one from the School of Education? ...... YES NO

12. After the initial visit to the school do you think your internship would have been just as successful without the college supervisor visiting your classroom and having conferences with him? ...... YES NO
BIBLIOGRAPHY

Books


Articles


Guszak, Frank J., "Teacher Questioning and Reading," The Reading Teacher, 21 (December, 1967), 227-234.


Hunter, Elizabeth, "The Effects of Training in the Use of New Science Programs Upon the Classroom Verbal Behavior of First Grade Teachers as They Teach Science," Classroom Interaction Newsletter, 4 (May, 1969), 5-11.


Medley, Donald M., "Experiences with the OSsAR Technique," *Journal of Teacher Education*, 14 (September, 1963), 267-273.


Medley, Donald M., "Some Behavioral Correlates of Teacher Effectiveness," *Journal of Educational Psychology*, 50 (December, 1959), 239-246.


Educational Resources Information Center (ERIC) Reports


Dauterman, Philip, "Are There Any Questions?". Alberta English '70, 10 (Summer, 1970), 29-32 (ERIC, ED 045 672).


Konetski, Louis C., "Instruction on Questioning," paper read before the National Association for Research in Science Teaching Annual Meeting, Minneapolis, Minnesota, March 5-8, 1970 (ERIC, ED 040 040).

Medley, Donald M., "Studying Teacher Behavior with the OScAR Technique," summary in Proceedings Conference on Instructional Methods and Teacher Behavior, Berkeley, California, Far West Laboratory for Educational Research and


Piercey, Dorothy, "Questioning Strategies in Reading or Is the Right Person Being Trained to Ask the Questions?" paper read before a meeting of the Western College Reading Association, Los Angeles, California, April 1-3, 1971 (ERIC, ED 050 896).

Seagren, Alan T. and others, "The Impact of Student Teachers Upon the Attitude and Achievement of High School Students," Kansas City, Missouri, Mid-Continent Regional Education Laboratory, Incorporated, Lincoln, Nebraska, Teachers College, University of Nebraska, December 31, 1967 (ERIC, ED 026 346).


Trosky, Odarka S., "Modifications in Teachers' Questioning Behavior in the Development of Reading Comprehension and A Series of Supervisory Conferences," paper read before the International Reading Association Meeting, Atlantic City, New Jersey, April 19-23, 1971 (ERIC, ED 051 982).

Yost, Jr., Michael, "The Effect on Learning of Post Instruc-
tional Verbal Responses to Questions of Different
Degrees of Complexity, Final Report," Project Number
9-D-062, Grant Number OEG-4-70-0005(057), Washington,
D. C., U. S. Department of Health, Education, and Wel-
fare, Office of Education, April 15, 1970 (ERIC, ED 045
391).

Monograph

Flanders, Ned A., Teacher Influence, Pupil Attitudes, and
Achievement, OE-25040 Cooperative Research Monograph
Number 12, Washington, D. C., U. S. Department of
Health, Education, and Welfare, Office of Education,
1965.

Publications of Learned Organizations

Bebb, Alson M., Arlene F. Low, and Floyd T. Waterman, "Some
Ways of Studying Teaching," Supervisory Conference on
Individualized Teaching, Washington, D. C., Association

Bowers, Norman D. and Robert S. Scog, "Studies of Human
Relations in the Teaching-Learning Process, V. Final
Report: Evaluation of Laboratory Human Relations
Training for Classroom Teachers," Cooperative Research
Project Number 469, U. S. Office of Education, Depart-
ment of Health, Education, and Welfare, Chapel Hill,
North Carolina, University of North Carolina, Columbia,
South Carolina, University of South Carolina, 1961
(mimeographed).

Clegg, Jr., Ambrose A., "Analyzing the Cognitive Level of
Classroom Questioning: A Preliminary Report," paper
read before the fourth Conference on Instruction spon-
ored by the Massachusetts Teachers Association,
Amherst, Massachusetts, University of Massachusetts,
May 8, 1967 (mimeographed).

Frontiers of Thinking (author not given), A Faculty Pub-
lication of the William M. Stewart School, University
of Utah, Salt Lake City, Utah (no copyright date).

Gallagher, James J., "Productive Thinking in Gifted Chil-
dren," Cooperative Research Project, Number 965, U. S.
Department of Health, Education, and Welfare, Office of
Education, Urbana, Illinois, University of Illinois,
Institute on Exceptional Children, 1965 (mimeographed).


Unpublished Materials


Rogers, Virginia M., "Varying the Cognitive Levels of Classroom Questions in Elementary Social Studies: An Analysis of the Use of Questions by Student Teachers,"


Test Materials and Observation Manuals


Harris, B. M. and K. E. McIntyre, A Manual for Observing and Analyzing Classroom Instruction, Austin, Texas, The University of Texas at Austin, Extension and Field Service Bureau, 1964.


Sequential Tests of Educational Progress, Directions for Administering and Scoring, Princeton, New Jersey, Cooperative Test Division, Educational Testing Service, 1957.


