INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
A Xerox Education Company
BROADWATER, Ernest Hoff, 1944–
AN EXPLORATORY STUDY OF THE REPRESENTATIVE VALIDITY OF THE OBSERVATIONAL SYSTEM FOR INSTRUCTIONAL ANALYSIS.

The Ohio State University, Ph.D., 1972
Education, teacher training

University Microfilms, A XEROX Company, Ann Arbor, Michigan

© 1972

ERNEST HOFF BROADWATER

ALL RIGHTS RESERVED
AN EXPLORATORY STUDY OF THE REPRESENTATIVE
VALIDITY OF THE OBSERVATIONAL SYSTEM
FOR INSTRUCTIONAL ANALYSIS

DISSertation

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By
Ernest H. Broadwater, B.Ed., Ed.M.

The Ohio State University
1972

Approved by

Advisor
Curriculum and Foundations
PLEASE NOTE:

Some pages may have
indistinct print.
Filmed as received.

University Microfilms, A Xerox Education Company
ACKNOWLEDGMENTS

The completion of this study would not have been possible without the encouragement and support given the author by many persons. The author was fortunate to have worked within the supportive climate which existed in the Department of Curriculum and Foundations at The Ohio State University.

The assistance given by Dr. John Kennedy in the analysis of the data was most beneficial.

The guidance and help received from Dr. John Belland was sincerely appreciated.

Special consideration must be given to Dr. John Hough for his continuing counsel, understanding and patience in furthering my career.

To my wife, Linda, I owe my deepest appreciation. Without her continued warm support and understanding on those many working weekends this dissertation never would have been completed.
VITA

September 6, 1944 . . . . . . Born - Philadelphia, Pennsylvania

1967 . . . . . . . . . . . . . B.Ed., Keene State College
Keene, New Hampshire

1968 . . . . . . . . . . . . . Ed.M., Temple University
Philadelphia, Pennsylvania

1968-1969 . . . . . . . . . . Peace Corps Volunteer
Sokoto, Nigeria

1969-1970 . . . . . . . . . . Chairman, Basic Education Department
Community Action Program
Manchester, New Hampshire

1970-1971 . . . . . . . . . . Teaching Associate, Department
of Curriculum and Foundations
The Ohio State University
Columbus, Ohio

1971- . . . . . . . . . . . . . Assistant Professor, Department
of Educational Foundations
Marshall University
Huntington, West Virginia
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>11</td>
</tr>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. STATEMENT OF THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Background of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>7</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>8</td>
</tr>
<tr>
<td>Assumptions, Delimitations and Limitations</td>
<td>10</td>
</tr>
<tr>
<td>Summary</td>
<td>11</td>
</tr>
<tr>
<td>II. REVIEW OF RELATED LITERATURE</td>
<td>13</td>
</tr>
<tr>
<td>Introduction</td>
<td>13</td>
</tr>
<tr>
<td>Coefficients of Reliability</td>
<td>14</td>
</tr>
<tr>
<td>Observational Research</td>
<td>16</td>
</tr>
<tr>
<td>Perceptual Variables</td>
<td>26</td>
</tr>
<tr>
<td>Investigation of Behavior Classification</td>
<td>32</td>
</tr>
<tr>
<td>The Observational System</td>
<td>34</td>
</tr>
<tr>
<td>Peer-Microteaching</td>
<td>35</td>
</tr>
<tr>
<td>Summary</td>
<td>36</td>
</tr>
</tbody>
</table>
### Chapter

<table>
<thead>
<tr>
<th>III. METHODS AND PROCEDURES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>38</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>39</td>
</tr>
<tr>
<td>Training</td>
<td>41</td>
</tr>
<tr>
<td>Selection of Subjects</td>
<td>49</td>
</tr>
<tr>
<td>The Peer-Microteaching Lessons</td>
<td>51</td>
</tr>
<tr>
<td>Data Collection</td>
<td>52</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>54</td>
</tr>
<tr>
<td>Summary</td>
<td>58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. PRESENTATION AND ANALYSIS OF RESULTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V. SUMMARY AND CONCLUSIONS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>77</td>
</tr>
<tr>
<td>Discussion of the Findings</td>
<td>80</td>
</tr>
<tr>
<td>Implications for Further Study</td>
<td>88</td>
</tr>
<tr>
<td>Summary</td>
<td>93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BIBLIOGRAPHY</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>94</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                                    Page

1. Scott Coefficient for Sixteen Subjects          50
2. Mean Per Cent of Behaviors Recorded in Each Category for Each of the Sixteen Peer-Microteaching Lessons           60
3. Cohen Coefficients for Four Peer-Microteaching Lessons for the Four Groups          62
5. Mann-Whitney U Values for Pairs of Subjects by Peer-Microteaching Roles for the Sixteen Peer-Microteaching Lessons          66
6. Mann-Whitney U Test for Teacher-Students and Observer-Students for Sixteen Microteaching Lessons          67
7. Kruskal-Wallis One-Way Analysis of Variance for Each of the Four Microteaching Groups          69 Table 7—Continued          70
8. Chi-Square Test for Each Category on the Ten Selected Behaviors in Each Microteaching Lesson          72
9. Frequency of Categories 6, 7 and 8 Recorded Within the Ten Selected Behaviors for Each Lesson          74
CHAPTER I

STATEMENT OF THE PROBLEM

Background of the Problem

Persons who work in positions in which their instructional behavior influences the learning of others have reason to be concerned with both the qualitative and quantitative nature of their behavior. Information about their instructional behavior and its influence on the learning and behavior of others may be gained through the use of observational systems. Systems for recording the instructional behavior of teachers have increased in number and frequency of use as evidenced by the sixteen volume anthology of observational systems compiled by Simon and Boyer. Many of these systems supply teachers, researchers and supervisory personnel with a tool for gathering information about the type and frequency of their verbal and nonverbal behavior. The term behavior as used in this study will refer to observable acts by the teacher or students which are significant enough to record using an observational system. Providing teachers with this information about their instruction is often referred to in the field of education as giving feedback. This feedback can provide teachers with information about their daily instructional behavior, and thus may serve as a valuable source of
information which they may use to analyze, and thus become more aware of their classroom teaching.

Most frequently the gathering of information about instructional behavior is done by a person who comes into the classroom and records the behaviors of the teacher and students whom he observes. Sometimes, however, the teacher may classify his and his students' classroom behavior from an audio- or videotaped recording of the teaching. In both instances, persons classify the behavior of others. Data collected through the use of observational systems is then usually summarized and analyzed. The data may then be used for such purposes as the following:

a. increasing teacher awareness
b. recording changes in teacher behavior
c. evaluating training
d. evaluating teacher performance
e. researching classroom behavior

All of these purposes for gathering data imply the expenditure of considerable human resources, and are usually motivated by a desire on the part of professionals to maximize the classroom environment for student learning. It is therefore crucial that these data be valid. If the feedback were not a valid representation of the classroom behaviors that actually occurred, the feedback should be rejected. It would be undesirable for teachers to use invalid information as a basis for change. There is then an obvious need for valid data in observational reports of classroom behavior.
Previous development and use of observational instruments have been concerned with the question of validity. Much of this earlier work, however, has dealt indirectly with the issue of representative validity, and has focused on construct validity. In some cases only concurrent validity has been established.

The previous work in the area of observational systems has assumed that the observer, trained in the use of an observational system, has, in fact, provided valid data on the type and frequency of classroom instructional behaviors. H. H. Anderson\(^{(30)}\) is considered by many to have been influential in fostering the extensive contemporary development and use of observational systems. Anderson's research used observers to record teacher and pupil behaviors in classrooms. Anderson reported high inter-observer agreement between observers trained in the use of his observational system. From the observational data of teacher and pupil behaviors, Anderson identified types of teacher classroom behavior consistent with his dominative and integrative constructs. He concluded that teachers using either dominative or integrative behaviors influenced pupils to behave differently under each condition. Thus Anderson's trained observers recorded teacher behaviors which he proposed influenced student classroom behaviors.

Withall,\(^{(51)}\) following the work of Anderson, developed a category system with a primary focus on teacher behaviors in the classroom. His observers also agreed with each other on the classification of behaviors as defined by the system. Withall's work postulated a
relationship between the frequency and type of teacher behavior and a resultant quality of classroom social-emotional climate.

Flanders,\(^{(10)}\) in his classic Minnesota and New Zealand studies, expanded Withall's categories to include student behaviors. There, behavior was classified by a trained observer. Scott's coefficient was used to determine the extent to which observers agreed with each other and with a standard (an authority in the system and its use) on the classification of behaviors. Flanders found a relationship between some classroom behaviors and student achievement and attitudes towards school. The use and increase in the number of observational systems following Flanders' work has been phenomenal.

Research on the training of teachers in the use of observational systems followed. In a study conducted at The Ohio State University in 1966 by Lohman,\(^{(56)}\) preservice teachers were trained in Interaction Analysis. Later, during their student teaching, these students had their classroom behavior classified by trained observers. By comparing the data of these student teachers with that of a control group, a measure of the training effect was obtained. The research observers in this study were considered ready to observe when they reached a point in their training where they demonstrated a high level of inter-observer agreement and agreement with a standard.

From the brief overview of four classic studies that have used observational systems, it is evident that the studies have certain factors in common. First, persons were trained to recognize instructional behaviors and then match those behaviors to categories
which constituted an observational system. Second, observer training continued until the individuals reached a high degree of agreement with each other and with a standard on the classification of classroom behaviors. In this way researchers have attempted to establish the construct validity of observational data. Third, observational data, recorded by trained observers in the classroom, has been related to selected dependent variables from which indirect measures of representative validity has been inferred. These factors, however, do not include direct measures of the representative validity of observational data.

A measure of representative validity of observational data involves analyzing behaviors recorded for a given lesson by the teacher and students as well as the classroom observer. The agreement of all of these persons on the classification of the instructional behaviors provides a measure of the representative validity of the observational data. In this study, observational data was collected using the Observational System for Instructional Analysis (OSIA) to record classroom instructional behaviors. The categories for recording instructional behaviors are: clarification; response to solicitation; initiation of information; solicitation of response; corrective feedback; confirmation; acceptance; positive personal judgment; negative personal judgment; silence; confusion. A thorough description of the Observational System for Instructional Analysis is presented later.
The following examples are intended to clarify for the reader the nature of the research question by providing specific examples of the representative validity problem as it may exist in the classroom. A classroom observer hears a teacher saying, "All right," as a response to a pupil. Hearing this, the observer records the category of confirmation for the teacher's statement. The teacher, however, classifies his behavior in the acceptance category. The student perceives the teacher's behavior as communicating in the positive personal judgment category. In this example the one behavior was perceived in three different ways. However, it is the datum recorded by the observer which typically would be reported. This being the situation, what is one to say of the representative validity of that reported datum. Certainly in this example the representative validity of that datum would be in serious question. Other combinations of behavior recorded by an observer, which might not agree with both teacher and pupil, could be the following: teacher says to a student, "You are not as dumb as I thought you were." The observer promptly records a negative personal judgment for this teacher's sarcasm. The teacher intended his comment to be a negative personal judgment. The student, however, perceives the teacher's comment as a positive personal judgment. While this example differs from the other, in that two of the three persons involved agree, there would still be a problem of the representative validity of the datum reported by the observer.
Thus, the issue of observational data validity has typically taken two forms: one has been to establish the construct validity of the data; and the second has been to establish representative validity through indirect means. For example, observational data has been related to dependent variables such as hostility, social-emotional climate and achievement, and these variables have been theoretically assumed to be dependently related to the use of selected instructional behaviors. Using variables that are assumed to be related to recorded classroom behaviors as a way of approximating the representative validity of observational data is clearly an indirect attempt to establish the representative validity of observational data. The representative validity of classroom observational systems remains largely unmeasured. This failure to establish the representative validity of classroom observational instruments represents a significant omission in research in the field of observational systems. Just because trained observers agree on the classification of teacher and student instructional behavior (and agree with a standard) does not necessarily mean that the participants in the instructional setting would agree with the observers' classification of their behavior.

Statement of the Problem

This study was designed to directly investigate the representative validity of one classroom observational system, the Observational System for Instructional Analysis. To establish an estimate of
representative validity necessitated obtaining observational data not only from the observer, but also the teacher and students involved in a given lesson. Therefore, answering questions concerning the representative validity of the Observational System for Instructional Analysis was dependent on concurrent recording of observational data by subjects in typical roles of observer, teacher, and student.

Specifically the study was designed to investigate the extent to which subjects trained in the Observational System for Instructional Analysis who were placed in the role of teacher, student, and observer for peer-microteaching lessons, agreed on the classification of behaviors recorded during each lesson. These subjects classified these peer-microteaching behaviors in which they themselves assumed one of the above roles. The representative validity of the classified behaviors using the Observational System for Instructional Analysis was investigated by analyzing inter-observer agreement of behaviors recorded by observers, teachers and students.

Definition of Terms

Representative Validity of Observational Data (hereafter referred to as representative validity) - the degree to which an observer and participants in an instructional situation agree on the classification of behaviors in which the observer and participants were involved.

Construct Validity - evidence gathered to show a correlation between a construct generated by a theory and some measurable criteria thought to represent the construct.
Observational System for Instructional Analysis (OSIA) - an instrument used to classify classroom behaviors of both teacher and students. In this research study only the first nine categories of behavior of the OSIA were used (clarification, response, initiation of information, solicitation, corrective feedback, confirmation, acceptance, positive personal judgment, and negative personal judgment). The usual distinctions between substantive and managerial behaviors were not made because of the nature of the lessons that were taught. Distinctions between student and teacher talk were also not made. It was not considered necessary to make such distinctions in this study because it was obvious to the subjects who the teacher was and who the students were. A tenth category of silence (replacing the OSIA categories of covert and overt silence) and an eleventh category, confusion (replacing the OSIA category of nonfunctional behavior) constituted the categories of the observational system used in this study.

Behavior - an observable expression of the teacher or student which is significant and clear enough to be recorded using the revised form of the Observational System for Instructional Analysis.

Participants - preservice teachers who played the role of students and teacher in peer-microteaching lessons, and whose behaviors were classified by using the revised form of the Observational System for Instructional Analysis.

Roles - a. Teacher - a subject who was responsible for instructing for a ten minute peer-microteaching lesson.
b. Student - a subject who conformed to the student role by following the instructional directions of the teacher. This subject was not responsible after the lesson for the content presented during the lesson.

c. Observer - a subject who did not participate in the peer-microteaching lesson. He viewed the behaviors of the teacher and students and later (along with the teacher and students who participated in the lesson) recorded the behaviors from a tape of the lesson.

Peer-Microteaching - as used in this study refers to the teaching of a lesson by an undergraduate education major. The lesson was taught to his peers, i.e. other undergraduate education majors. The exercise was a scaled down teaching experience: the number of students was two; a time limit of ten minutes was used; an obvious limit existed on the content presented. This microteaching to peers did not include the usual evaluation of teacher performance for later improvement, nor was the teacher trying to follow a specified teaching model. The subject in the role of teacher was responsible for the planning and presentation of the lesson. The instruction was directed toward the subjects who were playing the role of students. In this peer-microteaching a subject in the role of observer watched the lesson.

Assumptions, Delimitations, and Limitations

Assumptions - In any research study certain basic assumptions must be made. In this study, it was assumed that preservice teachers who
learned to tally at an inter-observer agreement coefficient of .80, as calculated by the Scott (\textsuperscript{2}) coefficient, would be able to record behaviors, used in a peer-microteaching lesson, in a way that would be highly consistent with the constructs of the Observational System for Instructional Analysis as they perceived such behaviors as participants in, or observers of, microteaching lessons. It was further assumed that subjects would behave in ways that would be typical of the roles of teacher or student they represented in the peer-microteaching lessons. It was also assumed that in listening to a taped recording of the lesson in which they participated, the teacher and students could cognitively and emotionally reassume the roles they played during the peer-microteaching lesson.

**Delimitation** - The present study included only preservice teachers. No inservice teachers, supervisors, administrators or researchers were included. The teacher and student behaviors recorded were limited to those which occurred during a peer-microteaching lesson as defined above.

**Limitations** - The subjects (preservice education students enrolled in an educational psychology course) were not controlled on such variables as sex, GPA, age, IQ, and cultural background. No personality variables were used in the selection of subjects, nor were any existing interpersonal attitudes between students measured.

**Summary**

For observational systems to effectively serve the purpose for which they are intended i.e., improvement of teaching, training,
supervision, and research, it is evident that the data must be valid. This investigation sought to study the representative validity of the Observational System for Instructional Analysis. Preservice teachers trained in the use of a modified version of the Observational System for Instructional Analysis tallied verbal behaviors from a tape recording of lessons in which they were involved in one of three roles: observer, teacher, or student.

Information gathered on the representative validity of observational data recorded from peer-microteaching lessons may have transfer value to other uses of observational systems. Thus, representative validity becomes meaningful for the teacher who is using observational data to increase awareness and to provide an information base for change. Supervisors, administrators, and researchers also want to be more sure of the representative validity of the data from which they base their evaluations, recommendations, instructional generalizations, and research conclusions. The availability of more comprehensive information on the representative validity of observational systems will be useful to all such people who work in various phases of the field of education.
CHAPTER II

REVIEW OF RELATED LITERATURE

Introduction

The purpose of this chapter is to present a review of the literature relevant to the question of representative validity of observational system data. The first section contains a review of the statistical techniques traditionally used to establish what has often been referred to as observer reliability. In the second section interaction analysis research will be presented in chronological order. Particular emphasis in that review will be on describing the method of observer training used by various researchers. Also reviewed will be prior attempts to establish the validity of observational data. The third section of this chapter reviews literature relating to the influences of personality and perceived variables as they effect the objective recording of classroom behaviors. A challenge to present observational practices and specific construct restrictions in the systems themselves will be investigated. In addition, information about reinforcement and emotional variables interacting in the recording of observational data will be presented. The fourth section is a brief statement of the relationship between the observational system used in this study and the more widely used
A fifth section that comments on the use of peer microteaching in teacher education will conclude the review.

**Coefficients of Reliability**

The numerical index referred to as the Scott coefficient has been used extensively by researchers to support the validity and reliability of observational data. The Scott coefficient, for nominal scales, is described by Scott as giving a measure of inter-observer agreement. This tool has been subsequently used by researchers to describe what they have referred to as the reliability of the observers. More accurately, reliability in these cases has been used to mean a mixed estimate of validity and reliability; where validity usually means construct validity (though sometimes concurrent validity) and reliability usually means interjudge reliability, i.e. the extent to which two or more observers record the same event with consistency. The formula for the Scott coefficients is

\[ \bar{W} = \frac{P_0 - P_e}{1 - P_e} \]

where \( \bar{W} \) = index of inter-coder agreement.

- \( P_0 \) = observed per cent agreement.
- \( P_e \) = per cent agreement expected by chance.

While \( P_0 \) may be readily obtained from the data, \( P_e \) must be computed. The \( P_e \) is derived by squaring the per cent of tallies for one of the observers for each category of the system that was used in recording the observed lesson, and summing the squares. This then is the formula most frequently followed by researchers reporting on the agreement between observers.
In an extensive inservice training project conducted under the direction of Flanders at the University of Michigan in 1963, a detailed description of the training practices are given. The project engaged in a two stage training program to teach observers to become reliable in the use of interaction analysis. The Scott coefficient was used as the index for observer reliability both in training and actual classroom observation.

Cohen developed a formula for a similar purpose which he described in an article titled "A Coefficient of Agreement for Nominal Scales." The formula for the Cohen coefficient is: 

\[ K = \frac{P_o - P_e}{1 - P_e} \]

where \( K \) = coefficient of interjudge agreement.

\( P_o \) = proportion of units in which the judges agreed.

\( P_e \) = proportion of units for which agreement is expected by chance.

Cohen states that Scott assumes "that the distribution of proportions over the categories for the population is known and is taken to be equal for the judges." Cohen feels this assumption questionable; thus, \( P_e \) for Cohen is computed by finding the product of the proportion for each of the categories for each observer and then summing the products.

Note that both Scott's \( \gamma \) and Cohen's \( K \) are indices of inter-observer agreement. It is typical for one of the observers to be considered as an expert in the use of the observational system, and thus this observer's data becomes the standard against which agreement is measured. The expert is assumed to know the system,
the theoretical construct from which categories are derived and the
definition of the categories. In addition the expert is assumed to
be highly competent in using the system, i.e. matching behaviors with
categories. Thus high agreement of a trained observer with the expert
becomes an estimate of the construct validity of the trained observer's
data. In those cases where neither observer is an expert in the
system, the use of either Scott's and Cohen's coefficients yield an
estimate of concurrent validity.

Observational Research

The following review of research with observational systems will
emphasize the training methods and statistical methods used in train-
ing observers. A second central point will be presenting some of the
indices, external to the specific observed behaviors, which have been
related to the recorded observational data. Studies representative
of research in the field of observational systems have been selected
for review. These studies are typical of the ways in which observers
have been trained. The uses made of the recorded observational data
are also similar to work done in other studies. Thus, while volumi-
nous research has been conducted using observational systems, those
other studies do not differ significantly from the selected studies
in the way observers have been trained nor in the relating of observ-
vational data to some other measures external to the observed
classroom behaviors.

In many of the studies reviewed in this section of the chapter
the researchers derived category systems from theoretical constructs,
and thus when they report coefficients of inter-observer agreement (often referred to as reliability) they are in fact reporting coefficients of construct validity. Most of these research reports, however, do not contain convincing evidence that observers consciously made use of a theoretical construct either during training or during observation. Thus, the authors' reports of coefficients of inter-observer agreement may, in part, be reports of estimates of concurrent validity.

Research in the use of observational systems was conducted by Anderson\(^{29}\) as early as 1936. His work is generally recognized as providing a valuable referent for the subsequent development of observational instruments. Anderson constructed a twenty category check list to record teacher and pupil behaviors to measure the occurrence of integrative and dominative behaviors, and observers were trained to use the system until they reached high agreement among themselves. Then using Anderson's check list as an observational system, the observers recorded the number and type of teacher and student contact behaviors in a preschool setting. From observer data Anderson gained support for his construct of dominative and integrative classroom behaviors. Anderson's observers, trained in the use of a check list, could record classroom behaviors from which information relevant to dominative and integrative contacts could be obtained. Thus, because Anderson's observers learned to use his check list with substantial inter-observer agreement, and it is assumed that they understood Anderson's integrative-dominative construct, their agreement could be considered as a measure of construct validity.
Withall's research on classroom behavior followed closely the work of Anderson. Withall modified Anderson's construct of dominative and integrative behavior to create a construct involving the influence of teacher and learner centered behavior on classroom social emotional climate. In Withall's research, only the behavior of the teacher was observed and recorded because of his theory regarding teacher influence on classroom social emotional climate. In his research, he trained observers in the use of a seven category observational system. The "objectivity" (Withall's term) of the system was established, in part, by having four trained judges apply his system independently to four transcriptions of teaching situations. The percentage of agreement of each judge with the researcher was then computed. Mean percentages of agreement ranged from 64% to 75%. One of Withall's conclusions from his classroom observational data was that "a valid measure of social-emotional climate of groups is obtainable by categorizing teacher statements." The validity of this conclusion was, in part, established by an experimental class of ten students who during the class pushed a plus or minus button to represent their feelings at a given point in the lesson. Withall observed that students felt slightly more positive in what he defined as the "learner centered" part of the class and much more negative during the "teacher centered" part of the class.

Flanders built upon Withall's construct of social-emotional climate. However, he constructed an observational instrument, interaction analysis, to record teacher direct and indirect behavior
and student behavior. He, too, depended upon classroom observers to record behaviors. These observers' tallies provided the basic data which he used in his later analysis. The Scott coefficient was used during training to provide an estimate of inter-observer reliability. After establishing inter-observer reliability in training, Flanders had observers collect data from the classroom. In addition he measured student achievement and attitudes toward school in these classes. Flanders concluded that gains in student achievement were related to teacher and student classroom behavior patterns as measured by his observational instrument. He proposed from these findings that teacher direct and indirect behaviors have differential effects on the classroom social-emotional climate which in turn influences student attitude and achievement. Thus, behaviors reported by persons trained to be reliable observers, as measured by the Scott coefficient, identified classroom behavior patterns which correlated with student achievement and attitude. By correlating observational data with measures of achievement and attitude, Flanders was testing an hypothesis derived from his theory of social-emotional climate, and in so doing was in part, indirectly establishing the representative validity of his observational system. This approach is closely related to a type of predictive validity.

Lohman studied the classroom instructional behavior of student teachers. He selected two groups of student teachers from a larger population, one group of which had received formal training in the use of Flanders' System of Interaction Analysis as a means of
analyzing teaching. The other group had learned to analyze teaching without the aid of a formal category system. This training was conducted in a general methods course prior to their student teaching. Lohman and another researcher observed both groups during their student teaching experience. The observers were trained in a revised form of the Flanders' System of Interaction Analysis by an assumed expert in the system, and their 'reliability' (Lohman's term) was calculated by the Scott coefficient. Observer reliability was checked six times during the study, and a mean coefficient of .76 was reported for the six reliability checks. The observers tallied the behaviors used in the classes in which the student teachers were instructing. The resulting interaction analysis data were used to measure differences in the behavior of student teachers that had learned interaction analysis and the group that had not learned a formal category system. Lohman reported that data gathered by trained, reliable observers clearly indicated significant differences in the frequency of use of the teacher behaviors of clarification, initiating information, and direction giving and the student behavior of student initiation. Because Lohman's observers were trained by an assumed expert in the system, this observational data may be considered to have construct validity.

Hill (54) conducted a study of inservice teachers to observe possible behavior change as a result of several patterns of training in interaction analysis. His reliability and the assumed construct validity of his data were established by having Amidon serve as the
standard, using the Scott coefficient as the index of validity and reliability. Teachers in the study were observed prior to training in interaction analysis and were observed again after training. At the conclusion of training, behaviors recorded by the observer were used to measure changes in teacher instructional behaviors as a result of the several different training designs they had received. Hill concluded that teachers did change their teaching behaviors in two of the Flanders' categories, but the changes were not related to the different training designs or modes of teacher feedback.

Kennedy, Haefele and Ruff (16) also trained in-service teachers in interaction analysis. Their research studied the effect of two different training procedures in teaching interaction analysis to in-service teachers. Effectiveness of the two training methods was measured using the interaction analysis data recorded for each of the two groups. Samples of classroom behaviors were recorded on audio tapes for the teachers from each group. A trained staff tallied from the audio tapes. Differences in the tallied behaviors were used in evaluating the two training programs.

Each of the last three studies used interaction analysis in a training program. Lohman studied the student teaching behaviors of preservice teachers trained in the use of a formal category system and others not trained in the use of a formal system; Hill trained in-service teachers in interaction analysis and then studied changes in their behavior as a function of in-service training and different types of feedback; Kennedy, Haefele and Ruff used two different
training methods in teaching interaction analysis and tallied the teaching behaviors of both groups. All tallies of classroom behaviors were made by trained observers either in the classroom (Lohman and Hill) or from audio tapes (Kennedy, Haefele and Ruff). All three studies used interaction analysis data provided by the observers for the purpose of evaluating the effects of different means of training.

Another study of the effect of training in interaction analysis training was conducted by Carline. (53) His study was a further investigation of the relationship between teacher instructional behavior and student learning. His study differed from Flanders' classic study in several ways. While Flanders studied teachers who were naturally direct or indirect, Carline trained teachers in one school in Interaction Analysis to intentionally increase their indirect behavior. Teachers in a second school (control) were not trained in Interaction Analysis, but were given a placebo treatment to minimize a possible Hawthorne effect. He used trained observers who had attained agreement with a standard of .80 or better by the Scott coefficient. Carline showed that teachers trained in Interaction Analysis did become more indirect in their teaching than teachers in the control group, but increased indirect teaching by those teachers did not result in greater student achievement than the achievement of students in the control group. Thus, while trained observers gathered data on classroom instructional behaviors in the established patterns of earlier studies, the lack of higher student achievement in classes where teachers became more indirect challenges an earlier assumption of interaction analysis.
Samph\( ^{(58)} \) conducted a study to determine the effect of the observer on classroom behavior. Without the knowledge of the teacher or students, he was able to monitor the verbal behavior in the classroom when an observer was not present. From those audio tapes, classroom behaviors were classified using Interaction Analysis. This data was then compared with the observer's data and Samph concluded that the teacher's verbal behavior did change when an observer was present. This finding increased the concern about the ability of the classroom observer to collect representative samples of classroom behavior. A later study of the same subject by Masling and Stern\(^{(13)} \) yielded the opposite results. However, the observer was always present in the latter study.

In addition to problems of observer training and establishing reliability and validity, some persons suggest that the process of recording teacher behavior as being directed toward a group, especially when many teacher-student contacts are dyadic in nature, may be inappropriate. Good and Brophy\(^{(12)} \) state there are two general assumptions concerning interaction analysis and proceed to challenge those assumptions. The two are:

1. That classroom interaction that takes place is between the teacher and the class as a group rather than the teacher and individual class members.

2. The differences which exist within the class are less important than the differences between classes.

They, too, used observers to gather their classroom data, but their
system recorded behaviors between the teacher and individual students. They concluded that teacher appraisal behaviors vary with individual students. Good and Brophy support their findings by citing several studies which show teachers criticize boys more often than girls. Thus, teacher appraisal behaviors are differentially applied for both individuals and subgroups within the class.

Amidon and Flanders, both having done extensive research with interaction analysis, studied the effects of two kinds of teaching behaviors on dependent-prone students. Teaching styles were identified using Flanders' System of Interaction Analysis. The authors concluded that dependent-prone students respond differently than independent-prone students to different instructional patterns.

This section has reviewed a number of representative studies from research done with observational systems. Many of these studies have used interaction analysis. A common approach throughout these studies has been the practice of training observers to agree with an expert, i.e., a standard. This approach to establishing construct validity has often paralleled efforts to indirectly establish the representative validity of observational data by relating observational data to theoretically related dependent measures, e.g., student achievement and attitudes. Thus, validity of observational data typically has been established by the following means:

1. Observers have obtained agreement with an expert and then between themselves; sometimes this agreement has been checked by having two observers record in the same classroom.
2. The observational data has been related to some dependent variable which is believed to be theoretically related to classroom behaviors, e.g. student achievement and attitudes.

In reviewing these methods of establishing validity, the need for a more direct measure of representative validity becomes obvious. Observer records of classroom behaviors should correspond with the perception of teachers and students whose behavior is being recorded by an observer.

A review of observer training, shows that little attention has been paid to observer reliability, (defined as the consistency with which an observer uses the categories of an observational system). For example, there have been few reports of behaviors recorded by an observer on a given tape being compared with behaviors recorded by the same observer at a later time using the same tape.

In training persons in interaction analysis, an implicit assumption has been made. That assumption is that indirect teacher behavior increases student achievement. The findings of Carline have not confirmed this assumption. This failure to confirm one of the theoretical constructs used for validating observational systems suggests the need for further research on observational data. Research by Samph showing that the teacher can change his behavior when knowingly being observed places an added validity problem upon observational data gathered by the classroom observer. The earlier emphasis on construct validity and the findings of Carline and Samph
require a more thorough investigation into establishing the validity of observational data.

The statement by Flanders that, "... the trained observer must remember that he alone is in the best position to judge whether the teacher is, in general, restricting or expanding the freedom of action of the students;..." needs clarification. Findings by Amidon and Flanders and Good and Brophy have shown students differ in their response to instruction. This suggests that it may be extremely difficult for an observer to record the differing student perceptions of teacher behaviors. The following review of the influence of numerous perceptual variables in perceiving social stimuli further points to the need for measures of observational data which include the perceptions of participants as well as the observer.

Perceptual Variables

The literature in social psychology suggests that persons vary, and therefore, one could assume that various persons would not be expected to agree on the classification of a given sequence of events. Secord and Backman conclude the section in their book, Social Psychology, on 'The Perceiver in Person Perception' by stating, "This section has emphasized individual differences among perceivers in their perception of other persons." Good and Brophy assert this point of view when they state that reinforcement of one student may not mean reinforcement for the entire class. Beyond not expecting various persons to agree perceptually on the classification of a given
event, there is reason to believe that different roles would also influence perceptions. Certainly many factors influence perception and various combinations of those factors are received by individuals. Peak adds that, "The perception of complex social stimuli will as a rule involve a series of perceptual sets."(21:350) It is to the factors influencing the perception of complex stimuli to which we now turn.

In reviewing influences on one's perception, Taft(50) included: traits, age, sex, family background, intelligence, training in psychology, aesthetic ability and sensitivity, emotional stability and character integration, self-insight, social skill and popularity, attitude toward social relations. The following variables were listed by Taft as having a high correlation with ability to judge: intelligence and academic ability, physical science specialization, aesthetic and dramatic interest, insight into status with respect to peers, emotional adjustment, social skills. Bruner and Tagiuri(7) stated that similarity between the judge and judged had an advantage in accurate perception. Many other sources include characteristics of stimulus and perceiver as well as the nature and types of interaction as factors influencing perception accuracy (Shrauger and Altrocchi(49), Bieri(6), Tagiuri and Petrullo(27)). The numerous variables studied and the many conclusions reached by Norman(45) in his article "The Interrelationships Among Acceptance-Rejection, Self-Other Identity, Insight Into Self and Realistic Perception of Others" further adds to the complexity of individual factors which have meaning with respect to the perception of others.
Motivation as an important element in all human behavior is widely accepted. It is obvious then that motivation would be a significant variable in a person's perception of the behavior of others. Speaking directly to this point, Henle suggests that the strength of motivation influences a person's ability. Taft lists the following as the most important factors in the ability to judge:

1. possessing appropriate judgmental norms
2. judging ability
3. motivation

Of the three factors, Taft states motivation could be most important. He also adds that when the judge is similar to the persons being judged that the judge is more likely to use the appropriate judgmental norm.

The role of experience has been shown to be another factor affecting a person's perception. Gage and Cronbach refer to social perception as "...a process dominated far more by what the judge brings to it than by what he takes in during it." While the specific content of the research by Segull, Campbell and Herskovits was not on social perception, but on visual perception, one of their statements seems relevant in the present context. They reported that "...each individual's experiences combine in a complex fashion to determine his reaction to a given stimulus situation." A further addition is that perception is just one of many components in human development and as such it is affected by individual experience. Asch suggests that if you want to know why a person
perceives a situation the way he does, you review his prior experiences relevant to these conditions rather than review the situation itself. This referent to past experiences as a determiner of perception has numerous implications for the present study. Hyman's statement is appropriate here: "When we observe - visually and/or auditorily - we do so from a particular vantage point."(H1:313)

Before reviewing several research projects completed on the related perceptual question, the inclusion of several other variables in social perception seems necessary. Roles in interpersonal relations influence a person's perception of others; this generally accepted fact was demonstrated by Price and Iverson.(H7) Role influences play a conspicuous part in classroom behavior. A related point was made by Bruner and Tagiuri(7) who reported that persons in various roles actually cued on different aspects of a stimulus.

It is significant to note that in the earlier discussion of observational systems there was an emphasis on the affective or social-emotional climate of the classroom. Bruner and Tagiuri(7) reviewed several studies relating to the accuracy of judging; a composite result showed the accurate judge to be relaxed while the more emotional judge was not as accurate. Asch(32) states that one's presence in a social field makes it impossible for that person to be free of the effects of the field. Thus, for Asch, it would be difficult to conceive of an observer not having his accuracy influenced by the emotion of the classroom.
Heider\(^{13}\) is concerned with the accuracy of perception and reports, "...not only is it true that perception leads to evaluation but evaluation can lead to perception." Sherif\(^{24}\) reports that peoples' perception of written material varied with the name of the author identified with that material. Research by Anderson and Barrios,\(^{31}\) and Luchins\(^{17,18,42}\) has demonstrated material presented first in a sequence has a greater effect on a person's perception than material coming later. Thus Heider's statement along with the other research could mean that observers recording classroom behaviors may be influenced by characteristics of the teacher or students in addition to the order in which the behaviors in that class occur.

Conn, et. al.,\(^{35}\) showed that elementary school students differed in their accuracy in perceiving vocal expressions of emotion and also differed in responding to positive teacher expectations. One significant variable in the perception of emotion was the sex of the student and the sex of the teacher. In concluding their study they stated that further investigation will be needed to understand the complex factors involved in individual student perception of communication. Finally in this consideration of the influence of evaluation and emotion on perception, Jones and Thibaut succinctly state that, "We may dislike a person when he thwarts us but feel quite neutral and objective when he thwarts another."\(^{15,153}\)

Bieri\(^{6}\) identified two different populations for a perceptual experiment. One group was social workers, predominantly female, and a second group was army reservists, predominantly male. He presented
each group with behaviors which were not consistent along a dominance-
submissive continuum. From the results he concluded that the two
groups did differ in their perceptions of the same behaviors.

In concluding this section on perceptual variables which influence the individual's view of his world, the multiplicity and complexity of those variables has become evident. Certainly, this review suggests that people do perceive as individuals but may have similar perceptions with others by having certain factors in common, i.e. sex, roles, experience, abilities. These findings present a particular problem for the classroom observer who is recording classroom behaviors from his own perceptual base.

A second point included in the perceptual literature relates to that which is being experienced within the observer himself. The above mentioned variables apply to him as well, but in addition to those are his level of motivation and emotional involvement. These latter two factors may significantly affect the representative validity of the data reported by the observer. The effect of these numerous perceptual variables involving not only teacher and students but also the observer may receive some clarification by obtaining observational data recorded from each of those persons.

Noting this situation further accentuates the need for investigating the representative validity of data derived from the use of observational systems. The literature on perception raises a serious question regarding the assumption of many researchers that an observer can record behaviors as if he were the person (teacher or
student) who is emitting a behavior or the person to whom the behavior is directed in a classroom setting.

Investigation of Behavior Classification

A final aspect in the recording of classroom instructional behaviors in addition to the observation research and perceptual literature is a brief inspection of the category systems themselves. Within the perception and classification of complex stimuli exists the problem of category system itself in establishing the units of behavior to be differentiated.

Bruner and Tagiuri (7) note that the nature and degree of the distinction required of the judge can present a problem. It would seem that the extent of the problem would depend upon the categories of a particular observational instrument. Biddle speaks directly to the present task by concluding; "Thus, all observer rating techniques suffer both from limitation of content and from unreliability of the encoding process." Belland, Belland and Price (52) studied the classification of teacher questions using two different observation instruments and concluded that the decision-making process required of the observer will influence his classification of behaviors. Thus, Bruner and Tagiuri, Biddle, and Belland, Belland and Price all express concern for the nature of the observational system limiting the reliability and validity of recorded observation data.

Another example of the category system itself influencing the resulting data is presented by Good and Brophy. (12) They have
suggested that the individual student should be the unit of interaction analysis data gathering. They reported that individuals and subgroups within the class are evaluated differently by the teachers. Compounding the complexity of differential teacher treatment of pupils are the findings of Premac.\(^{(46)}\) His research on students' differences in perception of reinforcers within a first grade class lend further evidence to suggest that one experience is not seen by all students as having the same reinforcing value. Thus, contrary to establishing reinforcers for all class members, Premac summed his work by stating that reinforcement is relative to the individual as opposed to a universal phenomenon for all individuals. From the findings reported by Good and Brophy, and Premac, one may tentatively conclude that teachers evaluate individual pupils differently and that what is reinforcing differs for individual pupils.

Thus the category system itself presents a further complex problem for the observer who is attempting to record data which is both reliable and valid. The distinctions and decisions required of the observer as well as the breadth and focus (individual or group) of the category system all serve to influence the data reported. Because the social-emotional oriented category systems (e.g. Flanders' System of Interaction Analysis) and other systems such as the Observational System for Instructional Analysis both contain categories directly related to potentially reinforcing behaviors, the fact that a reinforcer may not be the same for all individuals adds a final variable for the classroom observer in his effort to report valid data. By
having observational data reported by the teacher, student and observer, an index can be obtained to estimate the extent to which the observer records data consistent with the teacher and student, i.e. a measure of representative validity.

The Observational System

The observational system used in this study was the Observational System for Instructional Analysis. Subjects were trained to record instructional behaviors with this system, and later used this system to record teacher and student behaviors used in microteaching lessons. A thorough presentation of the system may be found in *Teaching: Description and Analysis* by Hough and Duncan. (14)

The Observational System for Instructional Analysis is a category system for the classification of classroom instructional behaviors. Only the first nine categories in the system were used in this study. These categories identify classroom instructional behaviors which include substantive and appraisal behavior categories. The system was developed to describe classroom instructional behaviors and specifically to describe patterns of behavior that define the instructional strategies used by teachers and students in the classroom.

The Observational System for Instructional Analysis was developed to describe classroom instructional behaviors and instructional strategies, and that development is different from the development of the social-emotional systems. The Flanders' System of Interaction Analysis, for example, followed a theoretical position about the
existence and effect of classroom climate on learning. The system was used to test those theoretical assumptions. The Observational System for Instructional Analysis was not developed to test any theoretical assumptions about the classroom but rather to describe instructional behavior and strategies.

The Observational System for Instructional Analysis and the Flanders' System of Interaction Analysis do not differ significantly in the way in which recordings of classroom behaviors are made. In both, time sampling systems have been used together with a change in category classification, should the behavior change within the set time interval. Both systems have used trained observers and preservice and in-service teachers have also been trained to record using these systems.

**Peer-Microteaching**

The use of microteaching in the training of preservice and in-service teachers has experienced considerable growth over the last decade. Scaled down teaching experiences had been conducted earlier under other rubrics, but microteaching as a specific type of scaled down teaching experience was developed concurrently with technological advances which provided relatively inexpensive audio- and videotape feedback after the teaching exercise. With the new technology, microteaching has established itself as a valuable and integral part of teacher education in many teacher training institutions. Allen and Ryan\(^{(1)}\) note in their book *Microteaching* that it is also an
excellent technique to use in conducting research. This input into the teacher education curriculum was undoubtedly accelerated by the work at Stanford University with microteaching. There, microteaching was viewed as a real teaching experience, i.e. teachers taught short lessons to a few students who were expected to learn the content.

There is evidence that the behaviors of preservice teachers during peer-microteaching is indicative of their behavior during student teaching. Ober(57) studied the relationships between the instructional behaviors used by preservice teachers in peer-microteaching, and the behaviors that they used later in student teaching. In the Ober study preservice teachers taught to their peers, while the Stanford microteaching model used school children. Ober reported that the preservice student teacher behavior under peer-microteaching conditions was an effective predictor of the way the preservice teacher would behave in the classroom during student teaching. Behavior during peer-microteaching then is behavior one could expect to be emitted during actual classroom teaching.

Summary

This chapter has reviewed several aspects of observational systems with specific emphasis on related research relevant to the present investigation of the representative validity of data gathered by using the Observational System for Instructional Analysis.

The means by which observers have been certified as reliable have consistently involved the training of persons to reach a certain level of agreement with another individual who has tallied the exercise and
is considered an expert. Indices of observer reliability in tallying the same tape two or more times have not been often reported. Studies showing correlational relationships between observational data and dependent variables such as student achievement and attitudes were most common. Observer training practices, the indirect means used to establish the representative validity of observational systems, and research findings that students differ in their response to a given teacher's behavior emphasize the need for a more direct investigation of the representative validity of observational data.

Numerous variables have been shown which influence an individual's perception. It has also been shown that individuals may perceive a given behavior differently. These findings present a potential problem of infinite complexity for the classroom observer who is recording classroom behaviors. Research is needed to learn the extent to which these classroom behaviors recorded by an observer, agree with recordings of the same behaviors by the teacher and student.

The development of and the categories in the Observational System for Instructional Analysis were described. The characteristics of peer-microteaching were presented as was research that gives reasonable credence to the assumption that teaching behaviors used by preservice teachers in peer-microteaching are similar to the behaviors used in student teaching. It appears, therefore, that peer-microteaching represents an appropriate context in which to explore the representative validity of data generated by the use of the Observational System for Instructional Analysis.
CHAPTER III

METHODS AND PROCEDURES

The purpose of this chapter is to present the research questions and hypotheses, to describe the training of the subjects and the experimental setting, and finally to present the methods of data collection and analysis.

Context

In the questions and hypotheses that follow, inter-observer agreement refers to the agreement between two people with respect to the classification of instructional behaviors used in peer-microteaching. All inter-observer agreement measures were obtained by using the Cohen coefficient. The research questions are asked within a context in which sixteen subjects were assigned to four groups of four subjects each. Within each of the groups, each subject taught one peer-microteaching lesson, served as an observer once and as a student twice. The first three questions and hypotheses relate to the sixteen peer-microteaching lessons as a whole. The fourth question and hypothesis refers to the four peer-microteaching lessons taught by each of the four groups of subjects. In questions one through four, inter-observer agreement refers to the total tallies for each lesson. The fifth question and hypothesis refers to the
agreement between the experimenter and subjects with respect to ten behaviors selected from each of the sixteen lessons.

Hypotheses

This exploratory study of the representative validity of data generated by use of the Observational System for Instructional Analysis was guided by the following questions and hypotheses:

**Question 1:** Does the role of the subject (teacher, student, or observer) cause that person to classify instructional behaviors differently than persons playing other roles? For example, do students classify the same instructional behaviors differently than the teacher? More specifically, the question may be stated as follows: given six possible classifications of pairs of subjects, will the inter-observer agreement of any one classification of pairs of subjects differ from the inter-observer agreement of any other classification of pairs of subjects? The following six classifications of pairs are possible when four persons are matched:

- $T - S_1$
- $T - S_2$
- $T - O$
- $O - S_1$
- $O - S_2$
- $S_1 - S_2$

where: $T =$ teacher; $S_1 =$ student one; $S_2 =$ student two; $O =$ observer.

**Hypothesis 1:** There will be no difference in the inter-observer agreement of any of the six possible classifications of pairs of subjects.
Question 2: Will the inter-observer agreement of any two pairs of subjects differ in the classification of peer-microteaching behaviors for the sixteen lessons?

The fifteen two pair combinations are listed below:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>S₁</td>
<td>T</td>
<td>S₂</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>S₁</td>
<td>T</td>
<td>0</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>S₁</td>
<td>O</td>
<td>S₁</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>S₁</td>
<td>O</td>
<td>S₂</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>S₁</td>
<td>S₁</td>
<td>S₂</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>O</td>
<td>S₁</td>
<td>O</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>O</td>
<td>S₂</td>
<td>O</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>S₁</td>
<td>S₂</td>
<td>O</td>
</tr>
</tbody>
</table>

Hypothesis 2: There will be no difference in the inter-observer agreement of any two pairs of subjects.

Question 3: Will the inter-observer agreement of the teacher and students differ from the inter-observer agreement of the observer and students?

Hypothesis 3: There will be no difference in the inter-observer agreement of teacher and students when compared with the inter-observer agreement of the observer and students.

Question 4: The previous three questions asked if there were differences in inter-observer agreement for the total sixteen
peer-microteaching lessons. The fourth question asks if there were differences in inter-observer agreement between subjects for each of the four lessons within each of the four groups.

**Hypothesis 4:** There will be no differences in inter-observer agreement between subjects for each of the four peer-microteaching lessons within each of the four peer-microteaching groups.

**Question 5:** In answering this question, a fifth person (the experimenter acting as a standard) was involved in the classification of behaviors. The question may be stated as follows: do subjects in the role of teacher, student, or observer and a standard (the experimenter) differ in the classification of specific behaviors which occur in the lessons in which those students participated? This question involved the following five persons:

\[ E \quad T \quad S_1 \quad S_2 \quad O \]

where: \( E \) = experimenter.

**Hypothesis 5:** Subjects in the role of teacher, student or observer and the researcher will not differ in their classification of selected behaviors used in the peer-microteaching lessons.

**Training**

The subjects for this study were undergraduate students at a private midwestern university. All were education majors in their
junior year enrolled in a required educational psychology course. As there was nothing atypical about this section of the course, there is no reason to believe that it was not a representative sample of the university's student body. None of the students had training in behavioral analysis in any previous methods course nor had any of the students learned a formal observation system for recording behavior. Therefore, the learning of the Observational System for Instructional Analysis and doing peer-microteaching was a new experience for all class members. Training for this class in the Observational System for Instructional Analysis was conducted by this experimenter. The members of the class who completed the training and participated in the peer-microteaching received credit toward their final grade. The assigned course instructor continued with other class activities during the weeks while the observational training took place.

The means by which the subjects were trained will be presented in chronological order for the six weeks of training. During this six week training period students received an average of two hours of training per week.

At the initial training session each subject received a copy of the categories of the revised version of the Observational System for Instructional Analysis which was used for the training. In this study no subscripts were used to distinguish between teacher and student behaviors. Subjects were directed to note the code symbol which corresponds to each category in the system. Each of the categories were then defined for the subjects; behavioral examples of the
categories were given, and student questions were answered. Briefly
the categories are defined as follow:

Category 1: **Clarification** - is the seeking of clarification in
response to a previous statement which for some reason needs clarify-
ing. The seeking of clarification is such that it follows directly
from the preceding comment of another person and is intended to have
that person expand upon or restate his earlier point. This should
produce increased understanding on the part of the person seeking the
clarification or assist another person's understanding.

Category 2: **Response to Solicitation** - is a behavior which
follows a question (that also includes a response to a person seeking
clarification). The response is then elicited by a solicitation or
the seeking of clarification. This category is not for initiating
ideas, feelings, or information but is a response category. If a
person responds to a behavior that seeks clarification, the behavior
is recorded as a response.

Category 3: **Initiation of Information** - is the presenting of new
information, ideas, or feelings. This category is not a response cate-
gory, but rather a category where the individual initiates information,
ideas, or feelings.

Category 4: **Solicitation** - is the asking of a question or the
giving of a direction to obtain a response from another individual.
While this question asking may include different kinds of questions,
it still has the effect of causing another person to respond. It
should be noted, however, that the classifying of a behavior as a
solicitation is not dependent upon there being a response to the solicitation.

Category 5: **Corrective Feedback** - is a behavior which communicates to another person that his behavior is incorrect and that it is incorrect because of some accepted source outside the teacher; e.g., custom, convention or empirical data. Telling a person he is wrong is not made for personal reasons. If the communication has strong negative overtones, it is classified under the negative personal judgment category.

Category 6: **Confirmation** - is behavior that communicates to an individual that he is correct or right. Typically this behavior follows correct responses to cognitive memory or convergent questions. Should the behavior have strong positive overtones, it is classified as a positive personal judgment rather than confirmation.

Category 7: **Acceptance** - is a behavior which is intended to inform another person that his behavior is neither right nor wrong but rather that the individual has the right to hold that position. Thus, while respecting the individual by accepting his behavior, it should be clear that no value judgment of "goodness" or "badness" is present in behavior classified under this category.

Category 8: **Positive Personal Judgment** - is a behavior intended to let another person know that his behavior is desirable, liked or appreciated. The personal nature of the praise in this category distinguishes it from the behaviors which occur in category six.

Category 9: **Negative Personal Judgment** - is a behavior which tells an individual his behavior is undesirable, inappropriate,
disliked or unappreciated. This category includes behaviors which are personally critical of an individual. It is the personal rejection intent in this category of behavior that should help distinguish this category from category five.

Category 10: **Silence** - When no verbal behaviors occur within a five second interval, this category is recorded. When persons are engaged in some nonverbal activity such as thinking or writing or perhaps engaging in physical activity, the silence category is used.

Category 11: **Confusion** - If two or more persons are talking at once or other classroom noises make it impossible to understand the communication, then this category is recorded. Note this category has no necessary relationship to preciseness of communication, understanding of material or compliance with directions.

The category system presented to the subjects is shown in Figure 1.

After defining and discussing the categories until the subjects understood the behaviors included in each category, a short audio tape was played. The tape contained examples of behaviors codeable under the categories of the Observational System for Instructional Analysis. As the subjects heard behaviors on the tape, they held up a flash card with the category code number they felt corresponded to the behavior. The use of these flash cards was most effective early in the training. By holding up a flash card for a specific category, the subjects and the researcher received instant feedback on the student's classification success. Later in the training, the flash cards were
used when specific problems appeared in the classification of particular kinds of behavior. The flash cards proved to be an effective training aid by helping to focus on specific behavior classification problems.

FIGURE 1

The Observational System for Instructional Analysis*

<table>
<thead>
<tr>
<th>Behaviors</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarification</td>
<td>1</td>
</tr>
<tr>
<td>Response to Solicitation</td>
<td>2</td>
</tr>
<tr>
<td>Initiation of Information</td>
<td>3</td>
</tr>
<tr>
<td>Solicitation of Response</td>
<td>4</td>
</tr>
<tr>
<td>Corrective Feedback</td>
<td>5</td>
</tr>
<tr>
<td>Confirmation</td>
<td>6</td>
</tr>
<tr>
<td>Acceptance</td>
<td>7</td>
</tr>
<tr>
<td>Positive Personal Judgment</td>
<td>8</td>
</tr>
<tr>
<td>Negative Personal Judgment</td>
<td>9</td>
</tr>
<tr>
<td>Silence</td>
<td>10</td>
</tr>
<tr>
<td>Confusion</td>
<td>11</td>
</tr>
</tbody>
</table>

*This is a shortened list of categories from the Observational System for Instructional Analysis. The whole category system is described in Teaching: Description and Analysis by John B. Hough and James K. Duncan.
At the end of the first meeting subjects were told they would be responsible for memorizing the code number for each category. At the following meeting the practice of working in dyads was introduced. Subjects worked in dyads for short blocks of time throughout the training. Two activities occurred in these dyads: first, subjects would quiz each other by one person stating the category and the other responding with the code number for that category; secondly, subjects would ask each other questions about the system, e.g., how do you define category five, or what is the tallying procedure for a change of behavior which happens within five seconds. Questions which could not be answered within the group were directed to the class or the experimenter. This strategy was successful in meeting its two objectives as well as providing another instructional strategy in the over-all training design. The subjects in these dyads did not give examples of behaviors for their partner to code.

The tape containing examples of each of the categories was used again in the second meeting to help the subjects practice coding behaviors. The rules for tallying were introduced in this session. The students received the following instructions:

1. Record a category code every five seconds.

2. If the same behavior continues for more than five seconds, place a slash mark next to the initial code classification for each five second interval that the behavior continues beyond the first five seconds.

3. Should the behavior change in less than five seconds, that change in behavior should also be recorded.
4. When more than one person is speaking at the same time and it is not possible to determine which person is communicating to the group, record category 11.

5. When a full five seconds passes and there has been silence, record category 10.

The subjects were then shown an example of how this tallying would be done. A short audio tape was used for the demonstration. The subjects then practiced with another short tape.

Training after the second session proceeded in the following order: the tapes the subjects tallied went from the simple (few behavior changes and behaviors in the same category) to complex (rapid behavior change with most of the categories being used); the tapes increased in length from less than a minute to a full ten minutes in duration. During these training sessions dyads were still used and flash cards helped in identifying specific problem areas. For example, subjects had difficulty in distinguishing behaviors between category two and three. By using flash cards, the specific point when this problem occurred could be identified and then increased practice could be given by introducing examples containing those behaviors. Tapes used for this training were from the Interaction Analysis Training Kit, The Ohio State University, and those developed by the experimenter.

After listening to a tape which the subjects thought was difficult, the tallies would be presented for the subjects on an overhead projector or written on the board. Then the tape would be played again. This would permit the students to look at the tallies
as they heard the tape. If the students felt they wanted to tally a tape a second time, the tape would be played again. Subjects were encouraged to ask questions and all questions were answered as well as possible during class or immediately following class.

At the sixth meeting the subjects tallied a tape from which they computed their own Scott coefficients. This provided feedback to the subjects in terms of their tallying ability. At this time a few students reported they had an inter-observer agreement coefficient of .80 or better for their data. Training and inter-observer agreement checks using the Scott coefficient continued for the next four meetings. As the practice continued, it became evident that more than sixteen subjects from the group were frequently obtaining a Scott coefficient of .80 or better.

Selection of Subjects

At the eleventh and twelfth meetings the trainer collected the tallies made by the subjects on the tapes played during those meetings. Subjects with Scott coefficients of .80 or better on three consecutive tapes were selected for peer-microteaching. Actually as more than the necessary sixteen subjects performed that well, the sixteen with the highest Scott coefficients were selected. The Scott coefficients for these sixteen subjects and their Scott coefficient for a tape recorded within a week after the completion of the peer-microteaching lessons is presented in Table 1. This final check, eight days after the testing which preceded the microteaching, showed no reduction in subject Scott coefficients.
## TABLE 1

**SCOTT COEFFICIENTS FOR SIXTEEN SUBJECTS**

<table>
<thead>
<tr>
<th>Subjects</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>Inter-observer Agreement Check After Peer-Microteaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>.87</td>
<td>.84</td>
<td>.81</td>
<td>.80</td>
</tr>
<tr>
<td>2.</td>
<td>.82</td>
<td>.81</td>
<td>.86</td>
<td>.83</td>
</tr>
<tr>
<td>3.</td>
<td>.84</td>
<td>.85</td>
<td>.83</td>
<td>.84</td>
</tr>
<tr>
<td>4.</td>
<td>.88</td>
<td>.85</td>
<td>.82</td>
<td>.88</td>
</tr>
<tr>
<td>5.</td>
<td>.82</td>
<td>.87</td>
<td>.85</td>
<td>.89</td>
</tr>
<tr>
<td>6.</td>
<td>.83</td>
<td>.86</td>
<td>.84</td>
<td>.83</td>
</tr>
<tr>
<td>7.</td>
<td>.86</td>
<td>.83</td>
<td>.87</td>
<td>.84</td>
</tr>
<tr>
<td>8.</td>
<td>.85</td>
<td>.87</td>
<td>.84</td>
<td>.90</td>
</tr>
<tr>
<td>9.</td>
<td>.89</td>
<td>.82</td>
<td>.83</td>
<td>.83</td>
</tr>
<tr>
<td>10.</td>
<td>.88</td>
<td>.84</td>
<td>.87</td>
<td>.90</td>
</tr>
<tr>
<td>11.</td>
<td>.82</td>
<td>.84</td>
<td>.84</td>
<td>.82</td>
</tr>
<tr>
<td>12.</td>
<td>.87</td>
<td>.84</td>
<td>.82</td>
<td>.80</td>
</tr>
<tr>
<td>13.</td>
<td>.83</td>
<td>.88</td>
<td>.81</td>
<td>.82</td>
</tr>
<tr>
<td>14.</td>
<td>.82</td>
<td>.82</td>
<td>.85</td>
<td>.83</td>
</tr>
<tr>
<td>15.</td>
<td>.81</td>
<td>.86</td>
<td>.86</td>
<td>.87</td>
</tr>
<tr>
<td>16.</td>
<td>.81</td>
<td>.91</td>
<td>.85</td>
<td>.82</td>
</tr>
</tbody>
</table>
The Peer-Microteaching Lessons

Having selected the sixteen subjects who would participate in the peer-microteaching lessons, they each received instructions concerning their responsibility for presenting a ten minute lesson. They were instructed that the lesson should be short enough to be taught in ten minutes and should be of a level of difficulty appropriate for the group. In preparing for the peer-microteaching lesson they would teach, subjects were strongly encouraged to use all the behaviors within the category system, and subjects were discouraged from presenting lessons which would be limited to one or a few of the behaviors in the system.

Subjects were informed that there were eight possible time blocks in which the peer-microteaching lessons could be taught. They were instructed to sign up for every possible time they could attend the lessons. Later, four students were randomly assigned to each of the four peer-microteaching sessions. All peer-microteaching sessions were concluded within one week after the completion of training.

When the four reported to their assigned peer-microteaching group, they received a sheet of paper assigning them to a given role for the first lesson. For the first lesson one subject was assigned the role of teacher, another the role of observer and two the role of student. After the first lesson, subjects received a second sheet of paper for their role in that lesson. This procedure was followed for the four lessons until each person had been in the role of teacher and observer once and the role of student twice. An investigation of the
representative validity of the Observational System for Instructional Analysis was thus made possible by analyzing observational data recorded by persons in different classroom roles.

The environment in which the peer-microteaching lessons were taught provided conditions appropriate for the activity. The usual classroom in which subjects normally met for the educational psychology class was used for the peer-microteaching lessons. This was a familiar surrounding for these subjects. Tapes used in recording the lessons were clear and audible and the tape recorder was the one previously used in training. With only the four subjects and the researcher present the classroom was virtually distraction free.

This section has defined the revised form of the Observational System for Instructional Analysis used in this study. The training of subjects to record classroom instructional behaviors using that system has been presented. Described in this section was the selection and placement of subjects by roles in microteaching groups.

Data Collection

In the peer-microteaching lesson the subject in the role of teacher taught the two subjects who were in the role of student and the subject in the role of observer watched from a short distance away. Each of these lessons was recorded on an audio tape recorder. As the lesson proceeded the researcher selected ten behaviors which would be checked by the subjects during the playback. Immediately following each lesson, the tape was played and all four subjects recorded the
behaviors using the revised form of the Observational System for Instructional Analysis. The tape was stopped at ten selected points during the playback and the subjects placed a check mark next to the category they had just tallied. This procedure was followed for all microteaching lessons.

A primary aim in selecting these behaviors was that a sampling from both the substantive (categories 1 through 4) and appraisal categories (categories 5 through 9) could be obtained. A secondary consideration was to sample behaviors distributed over the total time of the lesson. Both of these considerations had to remain flexible and at the discretion of the researcher to achieve the desired sampling. For example, some lessons contained a minimum of behaviors in the appraisal categories. Fixed sampling intervals or rigid adherence to one behavior per minute would have resulted in few or no appraisal behaviors being selected.

By training subjects to tally with the Observational System for Instructional Analysis and then placing those subjects in the various peer-microteaching roles, data was obtained to provide information needed in investigating the representative validity of data gathered by using the Observational System for Instructional Analysis. This strategy places persons in the three classroom roles of teacher, student, and observer which comprise the typical classroom when an observer is present. These subjects then tallied peer-microteaching behaviors and thus recorded the needed observational data to investigate the representative validity of the Observational System for Instructional Analysis.
Each subject's recorded behaviors served as the data base in answering the research hypotheses. The tallies of each subject were used with the tallies of each of the other subjects for computing a Cohen coefficient. This process yielded six Cohen coefficients for each microteaching lesson. These six coefficients were derived as a result of pairing the subjects' recorded tallies in the peer-microteaching lesson. These subjects were paired with respect to their role for the particular lesson. These Cohen coefficients (paired by roles) were then analyzed for the sixteen peer-microteaching lessons and thus provided observational data in terms of roles played by the subjects. These latter data were used as a measure of the representative validity, which was the objective of the research.

**Data Analysis**

The Cohen coefficient is used for nominal scales (which is the level of measurement of observational data) and it yields a measure of agreement between two sets of recorded behaviors. Each Cohen coefficient was computed using the tallies of one subject and the tallies of a second subject. This study used the Cohen coefficient as a research measure in order to maximize the sensitivity to differences between the recorded behaviors of the subjects. This increased sensitivity comes by more closely measuring frequency differences between the two sets of recorded data. This is the position presented by Cohen in Chapter Two. This coefficient then is more sensitive to the distributions of the proportions in the various categories within
an observational system and was, therefore, used in this research which is focusing on those differences in recorded behaviors.

The first research question was answered to determine if any of the six groups of pairs of Cohen coefficients differed from any other groups of pairs over the sixteen lessons. This general over-all answer, however, needed further clarification to determine exactly which groups of pairs differed.

The second research hypothesis was designed to answer the latter point. Again by using the tallies recorded by each subject in his respective role, a Cohen coefficient for each of the pairs was computed. What was being analyzed was the extent to which Cohen coefficients for two different sets of pairs of subjects differ over the sixteen lessons. Recall that each Cohen coefficient represents an index of the extent to which two subjects in their respective roles agreed on the classification of behaviors in the peer-microteaching lesson.

The use of two students in the peer-microteaching lessons provided further data for subsequent analysis. Because each student recorded the peer-microteaching behaviors, it was possible to compare the Cohen coefficients for both students with the observer, and both students with the teacher to determine if the students' agreement with the teacher differed with the students' agreement with the observer. This provided data to answer the third research question.

Since the same four persons were involved in four peer-microteaching lessons, the question was asked - would these
persons' Cohen coefficients differ over the four lessons? Thus, with respect to the four lessons within each of the four peer-microteaching groups, did those individuals within a peer-microteaching group differ in their recording of these four lessons? This fourth question was answered by having each person serve as a member of a single peer-microteaching group four times.

A fifth question was not answered by using the Cohen coefficients. This question was answered by analyzing ten selected behaviors within each lesson. By stopping the recorder at ten places during the playback of the microteaching lessons, it was possible to analyze these selected behaviors using the tallies of the researcher and those of the four subjects. This procedure made it possible for the individual categories within the observational system to be analyzed. A measure of the consistency with which each of these categories was recorded was then possible. It should be noted that by having the researcher, who also trained the subjects, record these ten selected behaviors, it was possible to introduce a standard for the classification of behaviors.

The mean per cent of behaviors in each category was computed. This provided a data pool for the subsequent comparisons of the lessons and the respective behaviors recorded.

Recorded peer-microteaching behaviors by the teacher, observer and two students were used in testing the research hypotheses.

The first research hypothesis sought to determine if the Cohen coefficients of any pair of subjects differed from any other pair for
the sixteen microteaching lessons. In testing this hypothesis, a
Kruskal-Wallis one-way analysis of variance by ranks was used. The
data used for the analysis met the assumptions of this statistical
technique, i.e. independent samples, ordinal level of measurement and
an underlying continuous distribution. Kruskal-Wallis $H$ will deter-
mine if any differences in the Cohen coefficients for the sixteen
lessons could have occurred by chance.

The second research hypothesis tests fifteen sets of pairs of
Cohen coefficients over the sixteen microteaching lessons. A Mann-
Whitney U test was used to compare these pairs of Cohens. The
Mann-Whitney U, like the Kruskal-Wallis, is a nonparametric test.
It, too, requires an ordinal level of measurement and assumes inde-
pendent samples. This Mann-Whitney U test will show if the Cohen
coefficients from one pair of recorders differs from a second pair.

The Mann-Whitney U test was also used in testing the third
hypothesis. This test determined if the Cohen coefficients of the
students with the teacher differ from those of the students with the
observer.

In testing the fourth hypothesis the Kruskal-Wallis one-way
analysis of variance by ranks was computed for each of the four
microteaching groups. Here the Kruskal-Wallis $H$ was used to determine
if the Cohen coefficients for the four lessons differed within each of
the four microteaching groups.

Research hypothesis number five did not use Cohen coefficients in
the analysis. The ten selected behaviors which were checked during
each lesson were analyzed to determine the subject and researcher's
agreement for the individual categories within the observational instrument. Here the behaviors were summed by individual categories for the sixteen lessons. These frequencies were recorded under the role of the person recording the behavior. The result of this procedure was that a frequency of the ten selected behaviors over the sixteen lessons could be analyzed by the category of behavior and the role of the person recording the behavior. To determine if the frequencies by roles within a category were different, a chi-square test was used. This test was used with independent samples with an ordinal level of measurement. Thus, the chi-square test could show if the researcher, teacher, students and observer differed on the frequency with which a specific category was recorded for the sixteen lessons.

In this exploratory study, a .10 level of significance was used in testing the null hypotheses. All tests were two tailed tests.

Summary

This chapter has presented the research questions and hypotheses of this study. The categories of the Observational System for Instructional Analysis and tallying procedures for the system were presented. The training of subjects to record classroom instructional behavior was described. The process of selecting the subjects was reported along with their placement in peer-microteaching groups. The means by which the research questions were answered and the hypotheses tested by the treatment and analysis of data was detailed in the final section. This included the designation of and rationale for the use of the statistical tests to be used in the data analysis.
CHAPTER IV

PRESENTATION AND ANALYSIS OF RESULTS

The purpose of this chapter is to report the research findings. In presenting the findings, each of the research hypotheses will be reviewed and the analysis of the data with respect to each of the hypotheses will be presented.

Each subject used the Observational System for Instructional Analysis to tally the frequency and type of behaviors used in the peer-microteaching lesson, from a tape recording, immediately following the lesson. These recorded behaviors (tallies) were then summed by categories for the four subjects. The mean per cent of the behaviors recorded in each category for each of the sixteen lessons is reported in Table 2. The mean per cent of the behaviors are shown for each of the four peer-microteaching lessons which occurred within each of the four peer-microteaching groups. The four different peer-microteaching groups are designated by Roman numerals while the individual lessons within each group are shown by Arabic numerals. At the bottom of Table 2 are the total means for the eleven categories for the sixteen lessons.

Table 2 shows the mean per cent of the behaviors recorded for each category for each of the microteaching lessons. Also that table displays the mean per cent for each category for all of the lessons.
TABLE 2

MEAN PER CENT OF BEHAVIORS RECORDED IN EACH CATEGORY FOR EACH OF THE SIXTEEN PEER-MICROTEACHING LESSONS

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Categories</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>13</td>
<td>67</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>64</td>
<td>13</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>13</td>
<td>63</td>
<td>13</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>15</td>
<td>57</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>51</td>
<td>8</td>
<td>21</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>30</td>
<td>11</td>
<td>16</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>33</td>
<td>22</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>10</td>
<td>25</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group III</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>20</td>
<td>63</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>60</td>
<td>18</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>75</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>17</td>
<td>59</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>11</td>
<td>58</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>51</td>
<td>16</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>12</td>
<td>64</td>
<td>18</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>66</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Mean Per Cents</td>
<td>1</td>
<td>21</td>
<td>52</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Ninety per cent of all the recorded behaviors occurred in categories 2, response to solicitation; 3, initiation of information and 4, solicitation. The appraisal behaviors, which include categories 5 through 9, accounted for only 9 per cent of the recorded behaviors.

Further investigation of Table 2 reveals that Group II used almost all of the category 5, corrective feedback, behaviors accounted for in the mean total. Group IV used almost all of the category 9, negative personal judgment, behaviors. Lesson 1 of Group III was the only lesson which failed to have any appraisal behaviors recorded. Category 10, silence, and category 11, confusion, were recorded only in the second lesson of Group II. The behavior of clarification accounted for only 1 per cent of the behaviors used in the sixteen lessons.

The tallies of each subject for each lesson were used in computing the Cohen coefficients. Using the Cohen formula, a coefficient of inter-observer agreement was computed for the recorded behaviors of all possible pairs of subjects for a given lesson. This procedure yielded six Cohen coefficients for the four subjects in a single lesson. These six Cohen coefficients for each of the four lessons for each of the four peer-microteaching groups are reported in Table 3. These Cohen coefficients were used in testing the first four research hypotheses.

The six Cohen coefficients for each of the sixteen microteaching lessons are presented in Table 3.
<table>
<thead>
<tr>
<th>Lesson</th>
<th>Group I</th>
<th>Lesson</th>
<th>Group II</th>
<th>Lesson</th>
<th>Group III</th>
<th>Lesson</th>
<th>Group IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T S₁ S₂</td>
<td>1</td>
<td>T S₁ S₂</td>
<td>1</td>
<td>T S₁ S₂</td>
<td>1</td>
<td>T S₁ S₂</td>
</tr>
<tr>
<td>S₁</td>
<td>78</td>
<td>S₁</td>
<td>83</td>
<td>S₁</td>
<td>83</td>
<td>S₁</td>
<td>75</td>
</tr>
<tr>
<td>S₂</td>
<td>84 82</td>
<td>S₂</td>
<td>95 81</td>
<td>S₂</td>
<td>91 85</td>
<td>S₂</td>
<td>84 77</td>
</tr>
<tr>
<td>0</td>
<td>86 84 73</td>
<td>0</td>
<td>95 80 97</td>
<td>0</td>
<td>77 75 75</td>
<td>0</td>
<td>74 90 80</td>
</tr>
<tr>
<td>2</td>
<td>T S₁ S₂</td>
<td>2</td>
<td>T S₁ S₂</td>
<td>2</td>
<td>T S₁ S₂</td>
<td>2</td>
<td>T S₁ S₂</td>
</tr>
<tr>
<td>S₁</td>
<td>91</td>
<td>S₁</td>
<td>68</td>
<td>S₁</td>
<td>76</td>
<td>S₁</td>
<td>93</td>
</tr>
<tr>
<td>S₂</td>
<td>71 81</td>
<td>S₂</td>
<td>69 88</td>
<td>S₂</td>
<td>67 89</td>
<td>S₂</td>
<td>92 95</td>
</tr>
<tr>
<td>0</td>
<td>80 84 80</td>
<td>0</td>
<td>71 88 86</td>
<td>0</td>
<td>70 90 89</td>
<td>0</td>
<td>81 95 88</td>
</tr>
<tr>
<td>3</td>
<td>T S₁ S₂</td>
<td>3</td>
<td>T S₁ S₂</td>
<td>3</td>
<td>T S₁ S₂</td>
<td>3</td>
<td>T S₁ S₂</td>
</tr>
<tr>
<td>S₁</td>
<td>83</td>
<td>S₁</td>
<td>90</td>
<td>S₁</td>
<td>90</td>
<td>S₁</td>
<td>79</td>
</tr>
<tr>
<td>S₂</td>
<td>82 88</td>
<td>S₂</td>
<td>80 88</td>
<td>S₂</td>
<td>86 90</td>
<td>S₂</td>
<td>79 79</td>
</tr>
<tr>
<td>0</td>
<td>71 70 63</td>
<td>0</td>
<td>84 83 83</td>
<td>0</td>
<td>80 71 80</td>
<td>0</td>
<td>72 85 77</td>
</tr>
<tr>
<td>4</td>
<td>T S₁ S₂</td>
<td>4</td>
<td>T S₁ S₂</td>
<td>4</td>
<td>T S₁ S₂</td>
<td>4</td>
<td>T S₁ S₂</td>
</tr>
<tr>
<td>S₁</td>
<td>82</td>
<td>S₁</td>
<td>81</td>
<td>S₁</td>
<td>74</td>
<td>S₁</td>
<td>60</td>
</tr>
<tr>
<td>S₂</td>
<td>70 83</td>
<td>S₂</td>
<td>74 86</td>
<td>S₂</td>
<td>84 88</td>
<td>S₂</td>
<td>71 79</td>
</tr>
<tr>
<td>0</td>
<td>87 85 77</td>
<td>0</td>
<td>74 90 88</td>
<td>0</td>
<td>83 91 89</td>
<td>0</td>
<td>80 71 87</td>
</tr>
</tbody>
</table>

*Where T = Teacher
S₁ = Student #1
S₂ = Student #2
₀ = Observer
In Table 3, the third lesson in Group I shows lower Cohen coefficients for the observer with the teacher and each student than the Cohen coefficients of the teacher with the students or the students with each other. The second lesson in Group II shows the Cohen coefficients with the teacher far lower than those with any of the other subjects. Table 2 shows that this lesson had 16 per cent of all behaviors falling in category 5. That one lesson contained almost half of all the corrective feedback recorded for all the microteaching lessons. In Group III the coefficients of the teacher in the second lesson were lower than any of the other subjects' Cohen coefficients.

Individual Cohen coefficients which are quite high or low can be observed in most lessons, e.g., the teacher and student one in Group I, lesson 2, or the observer and student 2 in the third lesson with the same group.

In testing the first hypothesis, there will be no difference in the inter-observer agreement of any of the six possible classifications of pairs of subjects, a Kruskal-Wallis one-way analysis of variance was used. The Cohen coefficient, ranks, rank sums and Kruskal-Wallis value found in testing Hypothesis I are presented in Table 4. The table is organized by the roles of the subjects over the sixteen microteaching lessons.

The Kruskal-Wallis H of 6.99 reported in Table 4 is not significant at the .10 level. Therefore, the first null hypothesis cannot be rejected.
<table>
<thead>
<tr>
<th>T-S₁</th>
<th>T-S₂</th>
<th>T-O</th>
<th>0-S₁</th>
<th>0-S₂</th>
<th>S₁-S₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>29</td>
<td>84</td>
<td>58.5</td>
<td>86</td>
<td>66.5</td>
</tr>
<tr>
<td>91</td>
<td>88</td>
<td>71</td>
<td>11.5</td>
<td>80</td>
<td>37.5</td>
</tr>
<tr>
<td>83</td>
<td>52</td>
<td>82</td>
<td>47</td>
<td>71</td>
<td>11.5</td>
</tr>
<tr>
<td>82</td>
<td>47</td>
<td>70</td>
<td>7</td>
<td>87</td>
<td>69.5</td>
</tr>
<tr>
<td>83</td>
<td>52</td>
<td>95</td>
<td>93.5</td>
<td>95</td>
<td>93.5</td>
</tr>
<tr>
<td>68</td>
<td>4</td>
<td>69</td>
<td>5</td>
<td>71</td>
<td>11.5</td>
</tr>
<tr>
<td>90</td>
<td>83.5</td>
<td>80</td>
<td>37.5</td>
<td>84</td>
<td>58.5</td>
</tr>
<tr>
<td>81</td>
<td>43.5</td>
<td>74</td>
<td>18.5</td>
<td>74</td>
<td>18.5</td>
</tr>
<tr>
<td>83</td>
<td>52</td>
<td>91</td>
<td>88</td>
<td>77</td>
<td>26.5</td>
</tr>
<tr>
<td>76</td>
<td>21</td>
<td>67</td>
<td>3</td>
<td>70</td>
<td>7</td>
</tr>
<tr>
<td>90</td>
<td>83.5</td>
<td>86</td>
<td>66.5</td>
<td>80</td>
<td>37.5</td>
</tr>
<tr>
<td>74</td>
<td>18.5</td>
<td>84</td>
<td>58.5</td>
<td>83</td>
<td>52</td>
</tr>
<tr>
<td>75</td>
<td>22</td>
<td>84</td>
<td>58.5</td>
<td>74</td>
<td>18.5</td>
</tr>
<tr>
<td>93</td>
<td>91</td>
<td>92</td>
<td>90</td>
<td>81</td>
<td>43.5</td>
</tr>
<tr>
<td>79</td>
<td>31.5</td>
<td>79</td>
<td>31.5</td>
<td>72</td>
<td>15</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
<td>71</td>
<td>11.5</td>
<td>80</td>
<td>37.5</td>
</tr>
</tbody>
</table>

Rank Sums 722.5 686.0 604.0 890.5 795.5 957.0

H=6.99 N.S.

*Where T=Teacher, S₁=Student #1, S₂=Student #2, 0=Observer
Hypothesis one stated:

There will be no difference in the inter-observer agreement of any of the six possible classifications of pairs of subjects.

Thus, these recorded peer-microteaching behaviors were not found to differ significantly. The behaviors were recorded by subjects in different peer-microteaching roles. They used the Observational System for Instructional Analysis to record the behaviors.

The second research question used the Cohen coefficients for pairs of subjects in their peer-microteaching roles. Hypothesis 2 sought to determine if any pairs of Cohen coefficients for given peer-microteaching roles differed over the sixteen microteaching lessons. Table 5 shows the only microteaching roles for which pairs of Cohen coefficients differed. The Mann-Whitney U used to test for that relationship is displayed in Table 5. Included with the Cohen coefficients are the ranks and rank sums used to compute the Mann-Whitney U.

The U of 60 for the Mann-Whitney U test was significant beyond the .10 level of confidence. Here the Cohen coefficients of the teacher and observer were found to differ significantly from the Cohen coefficients of the sets of two students. This finding resulted in the rejection of the second null hypothesis. That hypothesis was:

There will be no difference in the inter-observer agreement of any two pairs of subjects.
**TABLE 5**

MANN-WHITNEY U VALUES FOR PAIRS OF SUBJECTS*

BY PEER-MICROTEACHING ROLES

FOR THE SIXTEEN PEER-MICROTEACHING LESSONS

<table>
<thead>
<tr>
<th>Roles</th>
<th>U Value</th>
<th>Roles</th>
<th>U Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-S₁</td>
<td>129</td>
<td>T-S₂</td>
<td>124.5</td>
</tr>
<tr>
<td>T-S₁</td>
<td>109</td>
<td>T-S₂</td>
<td>97</td>
</tr>
<tr>
<td>T-S₁</td>
<td>99.5</td>
<td>T-S₂</td>
<td>108</td>
</tr>
<tr>
<td>T-S₁</td>
<td>120</td>
<td>T-S₂</td>
<td>85.5</td>
</tr>
<tr>
<td>T-S₁</td>
<td>91</td>
<td>o-S₁</td>
<td>112.5</td>
</tr>
<tr>
<td>T-o</td>
<td>84.5</td>
<td>o-S₁</td>
<td>124.5</td>
</tr>
<tr>
<td>T-o</td>
<td>92</td>
<td>o-S₂</td>
<td>89</td>
</tr>
<tr>
<td>T-o</td>
<td>60</td>
<td>o-S₁</td>
<td>124.5</td>
</tr>
<tr>
<td>T-S₂</td>
<td>105</td>
<td>o-S₂</td>
<td>89</td>
</tr>
</tbody>
</table>

p<.002

*T=Teacher, S₁=Student #1, S₂=Student #2, O=Observer

The Cohen coefficients of the teacher and students and the Cohen coefficients of the observer and students are reported in Table 6. A Mann-Whitney U test was used to determine if there were differences between the Cohen coefficients of teacher and students on one hand and observer and students on the other. This resulted in the comparison of thirty-two teacher - student Cohen coefficients with thirty-two observer - student Cohen coefficients. This analysis was performed to test the third hypothesis:
## TABLE 6

**MANN-WHITNEY U TEST FOR TEACHER-STUDENTS**

**AND OBSERVER-STUDENTS FOR SIXTEEN MICROTEACHING LESSONS***

<table>
<thead>
<tr>
<th>Cohens T-Ss</th>
<th>Ranks</th>
<th>Cohens O-Ss</th>
<th>Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>21</td>
<td>84</td>
<td>38.5</td>
</tr>
<tr>
<td>91</td>
<td>58</td>
<td>84</td>
<td>38.5</td>
</tr>
<tr>
<td>83</td>
<td>34.5</td>
<td>70</td>
<td>6</td>
</tr>
<tr>
<td>82</td>
<td>30</td>
<td>85</td>
<td>42</td>
</tr>
<tr>
<td>83</td>
<td>34.5</td>
<td>80</td>
<td>25.5</td>
</tr>
<tr>
<td>68</td>
<td>4</td>
<td>88</td>
<td>47.5</td>
</tr>
<tr>
<td>90</td>
<td>53.5</td>
<td>83</td>
<td>34.5</td>
</tr>
<tr>
<td>81</td>
<td>29</td>
<td>90</td>
<td>53.5</td>
</tr>
<tr>
<td>83</td>
<td>34.5</td>
<td>75</td>
<td>15.5</td>
</tr>
<tr>
<td>76</td>
<td>18</td>
<td>90</td>
<td>53.5</td>
</tr>
<tr>
<td>90</td>
<td>53.5</td>
<td>71</td>
<td>9</td>
</tr>
<tr>
<td>74</td>
<td>13</td>
<td>91</td>
<td>58</td>
</tr>
<tr>
<td>75</td>
<td>15.5</td>
<td>90</td>
<td>53.5</td>
</tr>
<tr>
<td>93</td>
<td>60</td>
<td>95</td>
<td>63</td>
</tr>
<tr>
<td>79</td>
<td>22</td>
<td>85</td>
<td>42</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
<td>71</td>
<td>9</td>
</tr>
<tr>
<td>84</td>
<td>38.5</td>
<td>73</td>
<td>12</td>
</tr>
<tr>
<td>71</td>
<td>9</td>
<td>80</td>
<td>25.5</td>
</tr>
<tr>
<td>82</td>
<td>30</td>
<td>63</td>
<td>2</td>
</tr>
<tr>
<td>70</td>
<td>6</td>
<td>77</td>
<td>19</td>
</tr>
<tr>
<td>95</td>
<td>62</td>
<td>97</td>
<td>64</td>
</tr>
<tr>
<td>69</td>
<td>5</td>
<td>86</td>
<td>44</td>
</tr>
<tr>
<td>80</td>
<td>25.5</td>
<td>83</td>
<td>34.5</td>
</tr>
<tr>
<td>74</td>
<td>13</td>
<td>88</td>
<td>47.5</td>
</tr>
<tr>
<td>91</td>
<td>58</td>
<td>75</td>
<td>15.5</td>
</tr>
<tr>
<td>67</td>
<td>3</td>
<td>89</td>
<td>50</td>
</tr>
<tr>
<td>86</td>
<td>14</td>
<td>80</td>
<td>25.5</td>
</tr>
<tr>
<td>84</td>
<td>38.5</td>
<td>89</td>
<td>50</td>
</tr>
<tr>
<td>84</td>
<td>38.5</td>
<td>80</td>
<td>25.5</td>
</tr>
<tr>
<td>92</td>
<td>61</td>
<td>88</td>
<td>47.5</td>
</tr>
<tr>
<td>79</td>
<td>22</td>
<td>77</td>
<td>19</td>
</tr>
<tr>
<td>71</td>
<td>9</td>
<td>87</td>
<td>46</td>
</tr>
</tbody>
</table>

**Rank Sums**  

| Rank Sums | 945  | 1117  |

**U=607**  

**z=1.28**  

**p < .10**

*Where T-Ss = Teacher with both students, O-Ss = Observer with both students.*
There will be no difference in the inter-observer agreement of teacher and students when compared with the inter-observer agreement of the observer and students.

Table 6 shows the Mann-Whitney U test for the difference between teacher and student Cohens compared with the observer and student Cohens. The reported Mann-Whitney U value of 607 is significant at the .10 level. This significant p value allows the third research hypothesis to be rejected.

The fourth research hypothesis was:

There will be no differences in inter-observer agreement between subjects for each of the four peer-microteaching lessons within each of the four peer-microteaching groups.

A Kruskal-Wallis one-way analysis of variance was used to test the hypothesis by using the Cohen coefficients for the four lessons within each microteaching group. Table 7 presents the six Cohen coefficients with their corresponding ranks for each lesson. The Kruskal-Wallis H for each microteaching group is presented in Table 7.

The significant Kruskal-Wallis H for the fourth microteaching group allows a conservative rejection of the fourth hypothesis. This significant H was beyond the .01 level of confidence. Thus, the same four persons within the fourth microteaching group differed significantly in their recording of behaviors occurring in four different lessons. Rejecting an hypothesis as a result of one out of four cases may raise question in the mind of the reader. A rationale for this decision will be presented in Chapter Five.
**TABLE 7**

**KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE FOR EACH OF THE FOUR MICROTEACHING GROUPS**

<table>
<thead>
<tr>
<th>Group</th>
<th>Roles</th>
<th>Microteaching Lessons</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cohens Ranks</td>
<td></td>
<td>Cohens Ranks</td>
<td></td>
<td>Cohens Ranks</td>
</tr>
<tr>
<td>S₁-O</td>
<td>84</td>
<td>18</td>
<td>71</td>
<td>4.5</td>
<td>82</td>
<td>13</td>
</tr>
<tr>
<td>S₂-O</td>
<td>73</td>
<td>6</td>
<td>80</td>
<td>9.5</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>S₁-T</td>
<td>78</td>
<td>8</td>
<td>91</td>
<td>24</td>
<td>83</td>
<td>15.5</td>
</tr>
<tr>
<td>S₂-T</td>
<td>84</td>
<td>18</td>
<td>84</td>
<td>18</td>
<td>70</td>
<td>2.5</td>
</tr>
<tr>
<td>S₁-S₂</td>
<td>82</td>
<td>13</td>
<td>81</td>
<td>11</td>
<td>88</td>
<td>23</td>
</tr>
<tr>
<td>T-O</td>
<td>86</td>
<td>21</td>
<td>80</td>
<td>9.5</td>
<td>71</td>
<td>4.5</td>
</tr>
<tr>
<td>Rank Sums</td>
<td>84</td>
<td>76.5</td>
<td>59.5</td>
<td>80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

H=1.17 N.S.

<table>
<thead>
<tr>
<th>Group</th>
<th>Roles</th>
<th>Microteaching Lessons</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cohens Ranks</td>
<td></td>
<td>Cohens Ranks</td>
<td></td>
<td>Cohens Ranks</td>
</tr>
<tr>
<td>S₁-O</td>
<td>95</td>
<td>22.5</td>
<td>69</td>
<td>2</td>
<td>80</td>
<td>6.5</td>
</tr>
<tr>
<td>S₂-O</td>
<td>97</td>
<td>24</td>
<td>86</td>
<td>14.5</td>
<td>83</td>
<td>11</td>
</tr>
<tr>
<td>S₁-T</td>
<td>83</td>
<td>11</td>
<td>68</td>
<td>1</td>
<td>90</td>
<td>20.5</td>
</tr>
<tr>
<td>S₂-T</td>
<td>80</td>
<td>6.5</td>
<td>88</td>
<td>17.5</td>
<td>83</td>
<td>11</td>
</tr>
<tr>
<td>S₁-S₂</td>
<td>81</td>
<td>8.5</td>
<td>88</td>
<td>17.5</td>
<td>88</td>
<td>17.5</td>
</tr>
<tr>
<td>T-O</td>
<td>95</td>
<td>22.5</td>
<td>71</td>
<td>3</td>
<td>84</td>
<td>13</td>
</tr>
<tr>
<td>Rank Sums</td>
<td>95</td>
<td>55.5</td>
<td>79.5</td>
<td>70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

H=2.78 N.S.
### TABLE 7--Continued

<table>
<thead>
<tr>
<th>Group III Roles</th>
<th>Cohens Ranks</th>
<th>Microteaching Lessons</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{1-O} )</td>
<td>91 23.5</td>
<td>67 1</td>
<td>86 15</td>
<td>84 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{2-O} )</td>
<td>75 5.5</td>
<td>89 18</td>
<td>80 9.5</td>
<td>89 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{1-T} )</td>
<td>83 11.5</td>
<td>76 7</td>
<td>90 21</td>
<td>74 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{2-T} )</td>
<td>75 5.5</td>
<td>90 21</td>
<td>71 3</td>
<td>91 23.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{1-S_{2}} )</td>
<td>85 14</td>
<td>89 18</td>
<td>90 21</td>
<td>88 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T-O )</td>
<td>77 8</td>
<td>70 2</td>
<td>80 9.5</td>
<td>83 11.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rank Sums: 68 67 79 86

**H=0.83 N.S.**

<table>
<thead>
<tr>
<th>Group IV Roles</th>
<th>Cohens Ranks</th>
<th>Microteaching Lessons</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
<th>Cohens Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{1-O} )</td>
<td>84 16</td>
<td>92 21</td>
<td>79 10.5</td>
<td>71 2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{2-O} )</td>
<td>80 13.5</td>
<td>88 19</td>
<td>77 7.5</td>
<td>87 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{1-T} )</td>
<td>75 6</td>
<td>93 22</td>
<td>79 10.5</td>
<td>60 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{2-T} )</td>
<td>90 20</td>
<td>95 23.5</td>
<td>85 17</td>
<td>71 2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_{1-S_{2}} )</td>
<td>77 7.5</td>
<td>95 23.5</td>
<td>79 10.5</td>
<td>79 10.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T-O )</td>
<td>74 5</td>
<td>81 15</td>
<td>72 4</td>
<td>80 13.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rank Sums: 68 124 60 48

**H=11.35**

**p < .01**
The testing of the fifth hypothesis did not include the use of the Cohen coefficients but rather used the ten selected behaviors from each lesson. These were the behaviors that were checked during the playback of each lesson. The checked behaviors were then summed by the respective category and the role of the subject during the lesson.

In addition to the four subjects, teacher, two students and observer, the experimenter also recorded the ten selected (checked) behaviors in each lesson. These tallies are displayed in Table 8. The chi-square value for each category is given in the right hand column of that table. Hypothesis 5 reads:

Subjects in the role of teacher, student or observer plus the researcher will not differ in their classification of selected behaviors used in the peer-microteaching lessons.

The significant chi-square value beyond the .01 level for category 6 allows for a conservative rejection of Hypothesis 5. Though the four subjects and the experimenter differed in the recording of category 6, confirmation, it should be noted that the difference here is primarily a function of the high frequency of tallies by the experimenter in this category.

For those categories defined as substantive, categories 1 through 4, agreement was extremely high. Thus, the data does not suggest questionable representative validity of these categories of the Observational System for Instructional Analysis.
### TABLE 8

CHI-SQUARE TEST FOR EACH CATEGORY ON THE TEN SELECTED BEHAVIORS IN EACH MICROTEACHING LESSON

<table>
<thead>
<tr>
<th>Categories</th>
<th>Experimenter T</th>
<th>S₁</th>
<th>S₂</th>
<th>0</th>
<th>x²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>.22</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>17</td>
<td>.02</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>44</td>
<td>45</td>
<td>42</td>
<td>.25</td>
</tr>
<tr>
<td>4</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>.10</td>
</tr>
<tr>
<td>6</td>
<td>26</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

* p< .01

One tally recorded in category 11 is not reported.
No tallies were recorded in category 10.

T=Teacher, S₁=Student #1, S₂=Student #2, 0=Observer

The nature of the higher chi-square values in categories 6, 7, and 8, suggests something different. This may be clarified by referring to Table 9. In this table, for each checked tally in one of those categories by the experimenter, the tally made by the subject is also reported. Table 9 then displays the categories recorded by the
four subjects at the time the experimenter was recording a specific behavior in Category 6, 7, or 8.

Inspection of Table 9 shows that the totals for all persons recording the selected microteaching behaviors of categories 6, 7, and 8 are equal. However, there is substantial variation in the recording of these categories by subjects. The table demonstrates that when the experimenter was coding categories 6, 7, or 8 that subjects were very often not recording the same category as the experimenter.

From Table 9 it can be seen that the classification of behaviors into category 6 presents a considerable area of disagreement between the experimenter and the subjects. Actually, when the experimenter recorded category 6, the subjects more often recorded positive personal judgments, category 8, than they did confirmation, category 6.

The tallies shown for category 7 do not reflect the type of classification problem shown for category 6. Here no other category is consistently recorded by the subjects when the experimenter is recording category 7.

The tallies reported under category 8, however, are similar to the pattern shown for category 6. Behaviors classified in category 8 by the experimenter are frequently classified as being in category 6 by the subjects. This would, however, be expected and the reason for this is discussed in Chapter Five.

Both Table 8 and Table 9 show substantial consistency in the classification of categories 6, 7, and 8 by the subjects.
### FREQUENCY OF CATEGORIES 6, 7, AND 8

**RECORDED WITHIN THE TEN SELECTED BEHAVIORS FOR EACH LESSON**

<table>
<thead>
<tr>
<th>Recording of*</th>
<th>Category 6</th>
<th>Category 7</th>
<th>Category 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Experimenter</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>T</td>
<td>12</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>S₁</td>
<td>11</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>S₂</td>
<td>8</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>5</td>
<td>14</td>
</tr>
</tbody>
</table>

*Where T=Teacher, S₁=Student #1, S₂=Student #2, O=Observer*
Summary

This chapter has presented the results of this exploratory study. The results of the research showed the first hypothesis could not be rejected. That is, no pair of individuals demonstrated measurable differences in recording behaviors from any other pair(s) for all microteaching lessons.

The second research hypothesis was rejected. The Mann-Whitney U test was significant beyond the .10 level of confidence. The Cohen coefficients of the teacher and observer differed from the Cohen coefficients for the two students over the sixteen microteaching lessons. Thus, a question is raised regarding the representative validity of the Observational System for Instructional Analysis data by the significant differences between recorded behaviors of subjects in those roles.

The Cohen coefficients of both students were compared with the teacher and the observer for the sixteen lessons, a significant Mann-Whitney U value resulted and caused the rejection of the third hypothesis. Table 6 showed the students agreed significantly more with the observer than they agreed with the teacher. Again, this represents a difference in recorded behaviors between the subjects who recorded peer-microteaching behaviors in which they themselves participated.

A significant difference in the recorded behaviors of the subjects in the fourth microteaching group caused the rejection of Hypothesis 4. This result showed that differences in recorded
peer-microteaching behaviors can occur between lessons for the same group of subjects. Thus, the same persons can differ significantly in recording microteaching behaviors over a series of four lessons. It was noted that this fourth group contained almost all of the negative personal judgment behavior for all the peer-microteaching lessons.

The fifth hypothesis tested both construct and representative validity of the categories in the observational system used in this study. A significant difference in the classification of category 6, confirmation, caused this hypothesis to be rejected. The representative validity between subjects and experimenter in tallying this category was therefore brought into serious question. Agreement for categories 7 and 8, acceptance and positive personal judgment, respectively, was shown to be less (though not significantly so) than for any of the other categories.
CHAPTER V

SUMMARY AND CONCLUSIONS

Introduction

This study investigated the representative validity of the Observational System for Instructional Analysis. Establishing an estimate of representative validity necessitated obtaining observational data not only from observers but from persons in the role of teacher and student as well. The following paragraphs will cite the more salient factors in this study of representative validity.

Recorded instructional behavior using observational systems has provided a data base for research, instruction, and evaluation in the classroom. The need to have valid data is obviously important; however, no research to date has provided information on how the participants, teacher and students, would have recorded behaviors recorded by observers. Rather, the ability of the observers to make valid observations has been assumed because of the observer's prior training in the use of the observational instrument. By having the teacher, students, and observer record classroom instructional behaviors, this study provided a more direct estimate of the representative validity of observational data.

This study was designed to investigate the representative validity of an eleven category simplification of the Observational
System for Instructional Analysis. In doing so, the representative validity of specific categories of that system was also investigated. To establish an estimate of representative validity required obtaining observational data from the observer as well as the teacher and students involved in a given lesson. Therefore, answering questions concerning the representative validity of the Observational System for Instructional Analysis is dependent on concurrent recording of observational data by subjects in typical roles of observer, teacher and student.

Subjects were trained to record classroom instructional behavior using the Observational System for Instructional Analysis. These subjects were then placed in peer-microteaching lessons in which each subject assumed the role of teacher and observer once and student twice.

In each of the peer-microteaching lessons the subject in the role of teacher taught the two subjects who were in the role of student and the subject in the role of observer watched from a short distance away. Each of these lessons was recorded on an audio tape recorder. Immediately following each lesson, the tape was played and all four subjects recorded the behaviors using the simplified form of the Observational System for Instructional Analysis. During the playbacks, the tape was stopped at ten points selected by the investigator during recording, and the subjects placed a check mark next to the category they had just tallied. Cohen coefficients were calculated to measure over-all agreement and a frequency distribution of categories was produced on
each set of checked tallies. Nonparametric statistics were used to analyze the data.

The findings indicate:

1. Persons in the various peer-microteaching roles were in general agreement regarding behaviors recorded. However, behaviors in the lessons were primarily substantive and substantive behaviors were tallied with higher inter-observer agreement than were the appraisal behaviors.

2. One significant difference (out of fifteen possible) occurred between the teacher - observer Cohens and the student - student Cohens in the peer-microteaching lessons. Agreement between the two students was the highest while agreement between the teachers and observers was markedly lower ($p<.002$).

3. Teacher - student Cohens were found to be lower than observer - student Cohens across the sixteen lessons.

4. The group exhibiting the highest percentage of negative personal judgment behavior produced significantly different Cohens across the four peer-microteaching lessons in which that group engaged.

5. The category of confirmation was tallied with significant disagreement by the teacher, students, observer and the experimenter. While the high
frequency of behaviors tallied by the experimenter in this category produced the significant effect, all other categories were tallied within chance agreement.

**Discussion of the Findings**

Before discussing each research hypothesis, several broad observations seem appropriate. The review of the literature in Chapter Two presented numerous variables which could contribute to the variance in recorded behaviors. Many of those variables were not controlled for in this study. For example, no index of motivation was obtained from the subjects at any time during the study. Individual differences such as perceptual history or variables similar to those identified by Taft\(^{(50)}\) were not included in the research design. However, the subjects did possess certain common characteristics such as age, race and school year. Thus, throughout all the findings, there were many variables which were not controlled.

The first hypothesis that:

There will be no difference in the inter-observer agreement of any of the six possible classifications of pairs of subjects,

could not be rejected. This indicates that the sixteen individuals who each assumed the role of teacher and observer once and student twice, did not measurably differ in their perception of the micro-teaching behaviors for the sixteen lessons. These findings would seem
to indicate that persons in the various peer-microteaching roles are in
general agreement regarding behaviors which they recorded while taking
the role of observer, teacher, or student. However, other data from
the study qualify the above statement.

These sixteen lessons contained a high percentage of substantive
behaviors, actually 90% over all the lessons while appraisal behavior
accounted for only 9%. From the data on the recording of behaviors in
specific categories, it is evident that these substantive categories
are tallied with a high degree of representative validity for all
roles. Thus, the agreement between subjects in the microteaching
roles may in part reflect the numerous substantive behaviors and the
high agreement in classifying those behaviors.

The second null hypothesis in the study was rejected. That hypothesis was:

There will be no difference in the inter-observer
agreement of any two pairs of subjects.

In testing that hypothesis there were fifteen possible
combinations of pairs of subjects. Therefore, the finding that two
pairs of roles differed for the sixteen lessons is certainly within
chance probability. While this significant finding may have occurred
by chance, the following discussion will explore a rationale for
suggesting that this was not a chance occurrence.

The teacher - observer Cohen coefficients differed from the
student Cohen coefficients for the sixteen lessons. The work by
Bruner and Tagiuri(7) on the influence of a person's role in his
perception may help to explain this finding. In testing this hypothesis then, it was found that the two subjects in the same role, the students, recorded microteaching behaviors with significantly higher agreement than the two subjects who had dissimilar roles, i.e. the teacher and observer. This could mean, as Bruner and Tagiuri suggest, that persons in different roles do cue on different aspects of the same stimulus.

The statement by Hyman\(^{(11)}\) that when we observe we do so from a particular vantage point is important here. This finding suggests that the subjects assumed their microteaching roles when they recorded the lessons. Further, in doing that, they showed that perception from the different roles is different. Not only are the roles different but the subjects in the role of student have a more common frame of reference from which to classify the instructional behaviors than do the subjects in the roles of teacher and observer.

While the review of the research on perception stressed the uniqueness of individual perception, this finding demonstrates that those subjects in the role of students had greater agreement in classifying behavior. Conclusions by Good and Brophy\(^{(12)}\) that individual students are reinforced differently was not found to influence the recorded perception of the students over all the lessons. It should be stated, however, that the subjects continuously rotated roles throughout the sixteen lessons. Certainly this would minimize between student variance, even though individuals were differentially reinforced throughout the lessons. The importance of this latter
point cannot be overemphasized. Good and Brophy have studied classrooms where the students could be expected to develop more stable perceptions of the teacher. This was not true in the fluid peer-microteaching situations.

In rejecting the third hypothesis:

There will be no difference in the inter-observer agreement of teacher and students when compared with the inter-observer agreement of the observer and students,

further indication of the need for studies of representative validity of observational system data was indicated. Here both students were paired with the teacher and also with the observer.

Bruner and Tagiuri\(^{(7)}\) indicated that a judge of behavior who is in an excited emotional state will be less reliable in reporting the behavior he observes. Though it would not be reasonable to say that a teacher recording his own behavior would be in an excited emotional state, it would seem likely that the teacher would be more emotionally involved in the recording of the microteaching behavior than the observer. Thus, the teacher then may have experienced emotional stress in recording his behavior. He may have been trying to have his behavior fit some preconceived notion of intent or personal style. These factors could be acting to produce the observed results. The observer who was not recording his own behavior may have been less emotionally involved than the teacher.
The fourth hypothesis that:

There will be no differences in inter-observer agreement between subjects for each of the four peer-microteaching lessons within each of the four peer-microteaching groups,

was rejected. The microteaching group which accounted for this difference was the group in which most of the negative personal judgment behaviors were reported. The finding that the representative validity of observational data can differ between lessons with the same group presents a complex challenge for establishing a representative validity index.

From the data it can be stated that extensive use of negative personal judgment behaviors coincided with decreased agreement in recorded microteaching behaviors. Heider noted the relationship between perception and evaluation and went on to state that evaluation can influence perception.\(^{(13)}\) Therefore, an individual being criticized may start to record behavior of the other individual differently from a situation in which that individual was not being criticized. Also, a person criticizing another individual can record the latter's behavior in ways not observed by another person recording that behavior. These statements are similar to Jones' when he reported on persons responding differently when the recipient of criticism as opposed to the observer of the criticism.\(^{(15)}\)

The rejection of hypothesis four stresses the need for further study of the representative validity of observational data. Here
persons trained in the use of the Observational System for Instructional Analysis failed to agree on the classification of behavior between four lessons. This disagreement in these lessons shows that representative validity differs for different lessons. Perhaps further study of the representative validity concept will give evidence as to what kinds of lessons are most likely to result in the lowering of the representative validity of recorded observational data. Perhaps as indicated here, those lessons high in appraisal behaviors will be recorded with lower representative validity. Note in rejecting this hypothesis, that it was only one of the four peer-microteaching groups which showed a significant difference.

The testing of the fifth hypothesis involved the tallies of the teacher, student, observer and the experimenter. Hypothesis five was:

Subjects in the role of teacher, student or observer plus the researcher will not differ in their classification of selected behaviors used in the peer-microteaching lessons.

Category 6, confirmation, was tallied by the teacher, student, observer and experimenter with significant disagreement. Two factors are central in rejecting this hypothesis. Firstly, only one of the nine categories tested reached significance and it was this one significant finding which caused the rejection of the hypothesis. Secondly, the reason for this result lies with the higher frequency of the experimenter's tallies over any of the subjects' in their peer-microteaching roles. A further discussion of the possibilities which exist here may point to plausible reasons for the finding.
It may be possible that the subjects didn't understand the behaviors associated with category 6; given this is possible, there is further need for concern. First of all, these subjects had already demonstrated their ability to tally at a .80 level as measured by the Scott coefficient. This does not insure, however, that these subjects were able to achieve reliability for all categories within the system. A second position may be that in a structured training session, subjects record the way they are expected to record, but when they are in a field situation they rely upon internal referents to record behaviors. The latter alternative would seem exciting to pursue, but is definitely beyond the scope of the present investigation.

Another possible explanation may be found in the motivation of the person recording the behavior. This particular problem could be extremely complex in nature, e.g. anticipating a 'desired' classification of the experimenter; not placing the needed effort into the classification of the behaviors; seeking to record their own behavior or the behavior of another in some desired way. While suggesting these differences are a function of measurable perceptual variables, the possibility that classification errors were made by the individuals, no matter what the reason, cannot be discounted.

Good and Brophy state that teachers reinforce students as individuals rather than classes. Premac showed how the same conditions take on different reinforcing values for individuals. Both findings are relevant to the data presented in Table 9. There, behaviors classified as confirmation by the experimenter were
frequently not labeled confirmation by the subjects. Actually, these behaviors were more frequently labeled positive personal judgment by the subjects. Information is not available as to the exact nature of those behaviors or the specific person initiating or receiving the confirmation. Individual differences in perception as presented by Secord and Backman (22) and others seem to be operating here rather than the role influence observed in hypotheses two and three.

The rejection of hypothesis five by the significant finding in category 6 causes a closer look at the category system itself. Belland, Belland and Price, (52) Bruner and Tagiuri, (7) and Biddle (33) have already been cited for their work bringing attention to the nature of category systems. Data presented in Table 9 indicate that in Bruner and Tagiuri's terms the nature and degree of the distinction required by the recorder is too great for category 6. This finding suggests that further similar studies of the individual categories of an observational instrument would yield data relevant to the representative validity of the categories themselves.

Because of the weight typically given to appraisal behaviors in analyzing the total instructional matrix, this significant lack of representative validity for category 6 is quite important. Adding to this, the lower representative validity for categories 7 and 8, more information is needed on the representative validity of appraisal behaviors.

A further complexity with respect to this hypothesis still remains. The issue is, is hypothesis five a test of construct or
representative validity? Most probably the answer is both types of validity are involved and cannot be separated here. With respect to representative validity, as presented, the subjects' agreement with the experimenter was low. Another procedure would have been necessary to determine how, in fact, the subjects agreed with each other when they recorded behaviors in category 6. The construct validity question may be addressed by a discussion in the implication section to follow, i.e. a redefinition of the appraisal categories.

Implications and Recommendations for Further Study

In testing the research hypotheses, there appeared few significant relationships when compared with the total number being analyzed. This suggests there is high agreement among the teacher, students, and observers in the recording of peer-microteaching behaviors. Of the relatively few significant findings, a pattern seems to exist in the following two areas: the recording of the subject in the role of teacher and the classifying of appraisal behaviors by all subjects. Though there were only a few significant findings does not mean they are not important.

Future research may be able to clarify the nature of the recording by the teacher. Certainly some of this work must be done in the classroom. There the teacher's perception can readily be compared with the perception of the students who are under the extensive control of this teacher. An observer's recording of the teaching can provide further data to assist in ascertaining the teacher's view of classroom behavior.
A combination of the strategy used in this research with the approach taken by Good and Brophy\(^{(12)}\) may be most desirable. A classroom teacher and his students could learn to record classroom behaviors using an observational instrument. Initially, groups within the class may be identified, by sex or race or eventually by individuals, and that observational data from those sources could be contrasted with the teacher's data. The results would provide information to address the teacher's perception more explicitly.

If recurring relationships were found in similar studies, simple techniques could be introduced for classroom teachers which would assist them in the understanding of their pupils' perception of teacher behavior. Remember the significant effect in hypothesis three reflects a cumulative perception of both students. Perhaps a cumulative effect for a cultural subgroup within the class would produce a variance in classifying classroom instructional behaviors. Certainly further study of the representative validity of observational data is needed with respect to the teacher - students and observer - students recorded behaviors.

The appraisal behavior categories studied here showed a greater amount of disagreement in the classifying of behaviors than did the substantive behaviors. A redefining of the appraisal behavior categories in the Observational System for Instructional Analysis may solve the observed lack of agreement for these behaviors. The slight change in category classification may be as follows:

1. Categories five and six would be defined in terms of content, e.g. indication that a response was incorrect
by virtue of a stated or understood external authority would be classified in category five; an indication of a correct answer under the same conditions would be classified in category six.

2. Categories eight and nine would be limited to personal exchanges not involving content e.g., any personal comments would be classified in category eight or nine depending upon whether it was positive or negative.

Such direct alterations of definitions in this category system may be more difficult to incorporate in other systems. Specifically if the other system does not contain concise categories which reflect this dichotomous nature, the alterations necessary may be more extensive. Thus, if the appraisal behaviors in other category systems allow for more subjective interpretation on the part of the recorder, more variance in individual report is likely to occur. Obviously, what is needed then, is more clarification of these appraisal categories in observational systems. The proposed changes in definition for the Observational System for Instructional Analysis may result in decreased observer difference for these categories. Research on the classification agreement present in the categories of other systems should logically precede any changes in category definition.

A complex problem remains in ascertaining the differences observed in recorded observational data. How much of the observed difference reflects the different perspective of the persons recording
the data? Is the variance a function of role, subgroup membership, individual differences or human error? A ventured answer to this complex problem would be that all are factors. The relative extent to which each account for the observed variance is presently unknown, but is certainly information which would be valuable to the researcher. The magnitude of such factors would determine their academic importance with respect to the use of classroom observational instruments.

In viewing these findings in retrospect, a few brief observations seem appropriate. The earlier concern that the microteaching environment may limit the behavior of the subjects does not seem to have occurred. The behaviors displayed in the lessons were not dissimilar in category frequency from behaviors recorded in school classrooms. Whether the actual perception of individuals in recording the microteaching behaviors was influenced by the microteaching situation becomes a question for later study.

Studies are needed to investigate the extent to which the representative validity of observational data recorded in microteaching lessons corresponds to observational data derived from the classroom. The significance of such a study to the representative validity of observational instruments is obvious. It may be learned that the microteaching situation is too artificial. That could mean that those relationships which we expect to exist in the classroom as a function of time are not able to be studied in a microteaching environment.

Subjects themselves in this study may not possess the perceptual sets of experienced classroom teachers. Thus, the recording of ones
own behavior may lack initial sophistication in performing those behaviors as well as in the recording of them. It has already been suggested that subjects in the role of students have not developed the stability of certain expectations and nuances which develop in on-going instructional settings. Did the observers in this study assume perceptual sets similar to persons who usually observe classroom instruction? It is known that these subjects lack the experience and formal training of veteran classroom observers. These subjects may not have the expectations and pedagogical orientations which influence the recording of classroom observers.

In this study there was one novelty effect of the experiment that may have had some effect on the participants in the research. This point was that these subjects were involved in recording their own behavior. This fact in itself is divergent from previous studies using observational instruments to record instructional behavior. A disciplined study of variables involved in this factor would certainly be complex, however, the evidence gleaned from such a study would have far reaching applications.

Finally, many of the numerous variables shown to influence the perception and recording of behavior were uncontrolled in this research. It must be accepted that answers to the research questions may have resulted from these many and often difficult to measure uncontrolled variables.
Summary

Future studies on the representative validity of the Observational System for Instructional Analysis and other observational systems should produce information to further our understanding of the nature of observational data. Replication of the present study is needed. Sensitivity to the variables mentioned earlier as well as research in the school classroom are important for future studies. The application of such investigations may function to further assist the classroom teacher in becoming more aware of how he or she is perceived by students and perhaps even learn how given behaviors are perceived by individual students.

This exploratory study into the representative validity of the Observational System for Instructional Analysis has shown that subjects in the role of teacher, student, and observer do not disagree on the over-all classification of behavior recorded in peer-microteaching lessons. This finding of the general agreement of these subjects indicates that the data recorded by subjects in peer-microteaching roles is representatively valid.

Considering the total number of areas of possible disagreement, there were few specific areas in which disagreement was measured. Subjects in the role of students agree more with each other than do those subjects in the role of teacher and observer. Classification agreement between the teacher and the students is lower than the agreement between the observer and the students. Subject agreement in the classification of appraisal behaviors was found to be lower than the classification of substantive behaviors.
BIBLIOGRAPHY

BOOKS


ARTICLES IN JOURNALS


44. Meyer, W. J. and Thompson, G. G. "Sex Differences in the Distribution of Teacher Approval and Disapproval Among Sixth-Grade Children". Journal of Educational Psychology, 47, (1956), 385-396.


UNPUBLISHED MATERIALS


