INFORMATION TO USERS

This material 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.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeib Road
Ann Arbor, Michigan 48106
74-11,059

SUDDUTH, Durward Lee, 1939-
AN ANALYSIS OF OBSERVER AND PARTICIPANT AGREEMENT USING THE FLANDERS' SYSTEM OF INTERACTION ANALYSIS FOR OBSERVING CLASSROOM VERBAL BEHAVIOR.

The Ohio State University, Ph.D., 1973
Education, teacher training

University Microfilms, A XEROX Company, Ann Arbor, Michigan
AN ANALYSIS OF OBSERVER AND PARTICIPANT AGREEMENT 
USING THE FLANDERS' SYSTEM OF INTERACTION ANALYSIS 
FOR OBSERVING CLASSROOM VERBAL BEHAVIOR 

DISSERTATION 

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

By 
Durward Lee Sudduth, B.S., M.A. 

* * * * * 

The Ohio State University 
1973 

Reading Committee: 
Dr. John B. Hough 
Dr. John C. Belland 
Dr. James K. Duncan 

Approved By 
Dr. John B. Hough 
Adviser 
Faculty of 
Curriculum and Foundations
ACKNOWLEDGMENTS

Anyone who has completed a doctoral program soon comes to realize that many people contribute to its completion . . .

I wish to express my sincere appreciation to Dr. John B. Hough who served as my major adviser. As one of the best teachers I have ever had, he has set an outstanding example of the quality teacher I hope to become. As an adviser, his guidance and encouragement and "being available" demonstrated his ability to put into practice outside the classroom those teacher qualities he inspires in his classroom.

I wish to thank Dr. John C. Belland and Dr. James K. Duncan. Their suggestions and assistance made a positive contribution to the completion of this dissertation as did their positive attitudes as members of the reading committee.

Special thanks are given to Dr. Joseph N. Wetmore and Dr. Lester E. Angene for their constant support and encouragement. The contributions of these colleagues have been significant and are deeply appreciated.

I wish to thank my students at Ohio Wesleyan University, both those who participated directly in this study as well as those who have encouraged me in many ways to strive to become a better teacher. I also thank my former students at Columbus Central High School who initiated me into the classroom and conclusively confirmed my love of teaching.
Finally, I thank my parents, who have always supported and encouraged my educational efforts in every way; my Grandmother, who is a worker and has encouraged me to work at learning; and especially Sue, for sharing with me the good times and the bad in helping me to finish what started a long time ago.

D. L. S.
VITA

October 1, 1939 ........ Born - Columbus, Ohio

1961 ..................... B.S. in Education
The Ohio State University
Columbus, Ohio

1964 ..................... M.A. in Education
The Ohio State University
Columbus, Ohio

1963-1964 ................ Teacher
Columbus Central High School
Columbus, Ohio

1964-1968 ................ Teacher-Counselor
Columbus Central High School
Columbus, Ohio

1968-1969 ................ Part-time Instructor
Department of Education
Ohio Wesleyan University
Delaware, Ohio

Assistant Director
Project Upward Bound
Ohio Wesleyan University
Delaware, Ohio

1969-1970 ................ Project Director
Project Upward Bound
Ohio Wesleyan University
Delaware, Ohio

1969-1973 ................ Instructor
Department of Education
Ohio Wesleyan University
Delaware, Ohio

FIELD OF STUDY

Major Field: Education

Studies in Curriculum and Foundation. Professor John B. Hough

Studies in Counseling and Guidance. Professor Anthony C. Riccio
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Background of the Problem</td>
<td></td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td></td>
</tr>
<tr>
<td>Definition of Terms</td>
<td></td>
</tr>
<tr>
<td>Assumptions, Delimitations, and Limitations</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>II. REVIEW OF RELATED LITERATURE</td>
<td>17</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Coefficients of Reliability</td>
<td></td>
</tr>
<tr>
<td>Observational Research</td>
<td></td>
</tr>
<tr>
<td>Perceptual Variables</td>
<td></td>
</tr>
<tr>
<td>Flanders' System of Interaction Analysis/</td>
<td></td>
</tr>
<tr>
<td>Broadwater Study/ Peer-Microteaching</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>III. METHODS AND PROCEDURES</td>
<td>54</td>
</tr>
<tr>
<td>Context</td>
<td></td>
</tr>
<tr>
<td>Hypotheses</td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Selection of Subjects</td>
<td></td>
</tr>
<tr>
<td>Peer-Microteaching</td>
<td></td>
</tr>
<tr>
<td>Data Collection</td>
<td></td>
</tr>
<tr>
<td>Data Analysis</td>
<td></td>
</tr>
<tr>
<td>Data Analysis</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>IV. PRESENTATION AND ANALYSIS OF RESULTS</td>
<td>79</td>
</tr>
<tr>
<td>V. SUMMARY AND CONCLUSIONS</td>
<td>103</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Discussion of the Findings</td>
<td></td>
</tr>
<tr>
<td>Implications and Recommendations for Further Study</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>124</td>
</tr>
</tbody>
</table>
# TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Comparison of Category Similarities between the Flanders' System of Interaction Analysis and the Observational System for Instructional Analysis</td>
<td>60</td>
</tr>
<tr>
<td>2. Scott Coefficients for Sixteen Subjects</td>
<td>69</td>
</tr>
<tr>
<td>3. Mean Per Cent of Behaviors Recorded in Each Category for Each of the Sixteen Peer-Microteaching Lessons</td>
<td>80</td>
</tr>
<tr>
<td>4. Cohen Coefficients for Four Peer-Microteaching Lessons for the Four Groups</td>
<td>83</td>
</tr>
<tr>
<td>5. Mean Cohen Coefficients for Four Peer-Microteaching Lessons for the Four Groups</td>
<td>84</td>
</tr>
<tr>
<td>7. Mann-Whitney U Values for Pairs of Subjects by Peer-Microteaching Roles for the Sixteen Peer-Microteaching Lessons</td>
<td>87</td>
</tr>
<tr>
<td>8. Mann-Whitney U Test for Teacher-Students and Observer-Students for Sixteen Microteaching Lessons</td>
<td>88</td>
</tr>
<tr>
<td>9. Kruskal-Wallis One-Way Analysis of Variance for Each of the Four Microteaching Groups</td>
<td>90-91</td>
</tr>
<tr>
<td>10. Chi-Square Test for Each Category on the Eighteen Selected Verbal Behaviors in Each Microteaching Lesson</td>
<td>94</td>
</tr>
<tr>
<td>11. Chi-Square Test on the Eighteen Selected Verbal Behaviors for Sixteen Peer-Microteaching Lessons by Associated Category Groupings</td>
<td>95</td>
</tr>
<tr>
<td>12. Chi-Square Values for Selected Verbal Behaviors From the Suduth Study (FSIA) and the Broadwater Study (OSIA)</td>
<td>99</td>
</tr>
</tbody>
</table>
### FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Comparison of Category Similarities Between the Flanders' System of Interaction Analysis and the Observational System for Instructional Analysis</td>
<td>61</td>
</tr>
<tr>
<td>2.</td>
<td>The Observational System for Instructional Analysis</td>
<td>62</td>
</tr>
<tr>
<td>3.</td>
<td>Comparison of Findings Between the Sudduth Study (FSIA) and the Broadwater Study (OSIA)</td>
<td>96</td>
</tr>
</tbody>
</table>
CHAPTER I
STATEMENT OF THE PROBLEM

Background of the Problem

America's educational system has always been subject to criticism. Despite criticism, as documented in Silberman's book *Crisis in the Classroom*, (32) from public and private, and group and individual sources, our public school system has continued to provide one of the best examples of mass education known to the world. In order to continue to improve the quality of education our nation's youth is receiving, more knowledge is needed about what happens in the classroom.

Since ultimate responsibility for student learning in our present educational system rests with the classroom teacher it would seem reasonable to provide the classroom teacher with valid and reliable data upon which he can make instructional decisions which will lead to improved student learning. The classroom teacher is concerned with the qualitative and quantitative nature of his instructional behavior as it influences the learning of his students. Information about a teacher's instructional behavior may be obtained through the use of one of the numerous classroom observational systems currently available. The sixteen volume anthology of observational systems compiled by Simon and Boyer is a definite indication of the significant development of observational systems. (33) Teachers, researchers or supervisory
personnel select the observational system best suited for their goals and use the system as one tool for gathering data about the type and frequency of designated verbal or non-verbal behaviors. The term "behavior" as used in this study will refer to observable acts on the part of either the teacher or students which are significant enough to record using an observational system designed for classroom use. The data recorded and properly analyzed provides the user with feedback upon which future instructional decisions can be based.

Observational systems most frequently used in a classroom setting make use of an outside observer who enters the classroom for the purpose of recording the instructional behavior of the teacher and learners. Availability of audio- or videotaping equipment makes it possible for the classroom teacher to make an electronic recording of his class and then analyze the tape using an observational system. In either situation, a person classifies the behavior of others according to the guidelines of the observational system being used. Once the data obtained through the use of an observational system is summarized and analyzed it may then be used for such purposes as:

a. describing teacher and/or student behavior
b. recording changes in teacher and/or student behavior
c. increasing teacher awareness of verbal or non-verbal behavior
d. evaluating teacher performance
e. evaluating teacher training
f. researching social-emotional climate of the classroom. (29)
One purpose for gathering information about classroom instruction is to provide valid data upon which future instructional decisions can be formulated. Research and evaluation studies as well as theory building also make use of data gathered in direct observation of instructional situations. Therefore, there is a crucial need for reliable and valid data obtained through the use of observational systems if these systems are to be used as educational research or evaluation tools or to collect data on which future instructional decisions can be made.

Thus, validity of data gathered through the use of an observational system is a critical factor in the value of the system and is the central concern of this study. Early development and use of observational systems dealt mainly with construct validity and only indirectly with representative validity. Reliance on the data obtained from the use of observational systems in the past has been predicated upon the assumption that an observer, trained in the use of an observational system, has provided valid data. If educators are going to draw conclusions from research and evaluation studies, or make changes in schools and in teacher preparation programs based, in part, on data gathered through classroom observational systems, it would seem to be a vital precaution to make certain that what observers observe in a classroom is in agreement with what the teacher and the students, as active participants in the learning process, also see occurring. Information of this nature can add significantly to the educational value of classroom observational systems as tools for educational improvement.
Credit for the significant contemporary development and use of observational systems should go to H. H. Anderson who was conducting research into the use of observational systems as early as 1936. (6) Anderson developed a twenty category checklist, trained observers in the use of the checklist, obtaining high inter-observer agreement, and then related the observational data collected by the observers in a preschool setting to his construct of dominative and integrative classroom behaviors. Through this method, Anderson established what might be called the construct validity of his observational system. He proposed a construct based upon certain mental health assumptions between teacher and student dominative and integrative behavior, developed a checklist to record behaviors related to the construct, trained observers in the use of an observational system designed to observe behaviors related to the construct and then used the observational data to support the construct.

Anderson also established a measure of indirect representative validity for his system by associating changes in student behavior to the classroom behavior of the teacher. His observers recorded data supporting the principle that when the teacher tended to manifest dominative behavior, students in the classroom would also manifest dominative behavior while the reverse was observed in the classroom of a teacher considered to be integrative.

Withall (58) followed Anderson's work with a seven category system whose focus was placed upon teacher classroom behavior. Withall modified Anderson's construct of dominative and integrative
behavior into a construct involving the influence of teacher and learner-centered behavior on the social-emotional climate of the classroom. Withall trained observers in the use of his seven category system and, from observational data collected, concluded that a valid measure of the social-emotional climate of a classroom could be obtained by recording teacher verbal behavior. Withall came very close to establishing direct representative validity of this conclusion by having students record their "feelings" at various points in the lesson. Therefore, in addition to the observational data recorded by an observer, Withall also had direct feedback from actual participants in the instructional situation.

Flanders' influence in the field of classroom observation has been significant and he is considered to be one of the major leaders in the development of classroom observational systems. (28:18) Building on the earlier research of Anderson and Withall, Flanders continued to work in the area of investigating classroom social-emotional climate. Acknowledging the influence of the dominative and integrative constructs of Anderson, and the teacher-learner centered constructs of Withall, Flanders developed both his ten category observational system and his theory of the social-emotional climate in the classroom around constructs dealing with direct and indirect teacher influence.

In Flanders' now classic New Zealand and Minnesota studies (1955-1960), trained observers were used to record teacher and student verbal behavior. Scott's coefficient was applied to determine
inter-observer agreement as well as observer agreement with a standard (the term "standard" being used to identify an authority in the use of a particular observational system) in the recording of teacher-student verbal behaviors. Flanders served as the standard in the observer reliability check of the New Zealand observers. (20:56)

The New Zealand and Minnesota studies produced observational data which was used to establish indirect representative validity. Through student attitude inventories it was determined in both the New Zealand and Minnesota studies that students were more inclined to have a positive attitude toward their teachers and their schoolwork when the teacher manifested a more indirect pattern of teacher behavior. This research concluded with similar findings in the 1959-60 Minnesota study in regard to student achievement. A more indirect teaching pattern produced higher student achievement. Thus, verbal behaviors recorded by persons trained to be reliable observers as measured by Scott coefficients of inter-observer agreement, identified classroom teacher behavior patterns which correlated with student attitude and student achievement. By attempting to correlate observational data with obtained measures of student attitude and achievement, Flanders was testing a construct derived from his theory of classroom social-emotional climate and thereby indirectly established the representative validity of his observational system.

If the use of data collected from observational systems can be used to influence teaching patterns in the classroom, the next logical step is to introduce training in the use of observational systems to
college programs involved in the preparation of teachers. A study by Lohman examined the impact of observational system training on students preparing to enter the teaching profession. In this study, preservice teachers who had been trained in Flanders' System of Interaction Analysis had their classroom instructional behavior during student teaching recorded by trained observers. Verbal behavior patterns of student teachers trained in the use of Interaction Analysis were compared to a group that received a different treatment. The observers used in this study were considered to be competent observers having obtained a high degree of inter-observer agreement. The study supported the position that training in the use of a classroom observational system can have significant influence on later student teaching behavior.

Certain common factors appear in the studies cited above:

1. Individuals were trained to recognize instructional behaviors and then relate those behaviors to previously defined categories constituting an observational system.

2. Individuals trained as observers attained a high degree of inter-observer agreement. In some cases, high inter-observer agreement with a standard was also established.

3. Observational data recorded by trained observers in the classroom have been related to selected dependent variables from which indirect measures of representative validity have been inferred.
The above factors do not include direct measures of the representative validity of data collected through the use of an observational system. The first two factors have been used in an attempt to establish construct validity. A more reliable measure of representative validity based upon observational data involved collecting, summarizing and analyzing those behaviors recorded during a given classroom lesson by the teacher and students as well as the classroom observer. The agreement of the active participants in the learning situation as well as the trained observer provides a direct measure of the representative validity of the observational data. In the present study, observational data was collected using the Flanders' System of Interaction Analysis (15) to record classroom instructional behaviors. The data was recorded by an observer as well as the participants in the instructional event being observed, all of whom were trained to use Flanders' System of Interaction Analysis. The categories for recording verbal behaviors are:

Teacher Talk

1. Accepts Feeling
2. Praises or Encourages
3. Accepts or Uses Ideas of Students
4. Asks Questions
5. Lectures
6. Giving Directions
7. Criticizing or Justifying Authority
A thorough description of Flanders' System of Interaction Analysis is presented in Chapter II of this study.

The following examples of classroom verbal behavior are intended to clarify for the reader the nature of the problem being researched in this study by providing specific examples of the representative validity problem as it may exist in the classroom. A classroom observer hears the teacher say, "John, are you really trying?" The observer records a "4" to represent the category of "teacher asks question." The teacher intended the statement as a "7" or the category for "teacher criticizes." John, on the other hand, perceives the statement as encouraging him to try harder and would record a "2" representing the category of "teacher encourages." In this single example the observer and all participants in the instructional event could conceivably record a different verbal category for the same statement. However, normally it is the datum recorded by the observer which would be used for purposes of analysis. What significance does this situation hold when one is concerned with direct representative validity? One must certainly question the representative validity of the observer's categorization of this particular classroom sample of teacher talk. Other situations can easily arise in which a combination
of any two persons (either active participants or observer) may disagree with the conclusion of a third person. An example might include the teacher statement, "John, please sit down." In this situation the observer records a "7" for "teacher criticizes," while the teacher and student record a "6" each having perceived the teacher as having simply given a direction with no criticism implied. Once again it would be the observer's datum which would be conveniently recorded for future analyses.

The representative validity of classroom observational systems remains largely unmeasured. A recent study by Broadwater took a major step toward researching this problem. (61) In that study, Broadwater examined the representative validity of a modified version of the Observational System for Instructional Analysis.

Before the Broadwater study, the issue of the validity of observational data was accounted for by one of two methods and in some instances a combination of both methods. One method has been to establish construct validity from the collected data. This method uses the observational data to test and then support the theory or theories used in the development of the observational system. The other approach has been to establish representative validity through indirect means. In this method, observational data has been related to such dependent variables as classroom social-emotional climate and student achievement. These and similar variables have then been theoretically assumed to be dependently related to the use of selected classroom instructional behaviors.
Using variables assumed to be related to recorded classroom behaviors as a way of finding representative validity of observational data is at best an indirect attempt to establish representative validity of data collected through the use of observational systems. The lack of or failure to directly establish the representative validity of classroom observational instruments represents a significant weakness in research in the area of observational instruments. High inter-observer agreement as well as high agreement with a standard as to classification of teacher-student classroom instructional behavior should not be accepted to mean that the primary participants in the actual learning situation, namely the teacher and the students, would agree with an outside observer's classification of their classroom instructional behaviors. Direct representative validity of classroom observational systems remains largely unmeasured; therefore, the purpose of this study is to research this problem.

Statement of the Problem

This study was designed to investigate the representative validity of one classroom observational system, the Flanders' System of Interaction Analysis. In order to establish an estimate of representative validity it was necessary to obtain observational data not only from the classroom observer but also from the teacher and students involved in the particular instructional event being observed. Data involved in answering questions concerning the representative validity of the Flanders' System of Interaction Analysis was dependent on concurrent
recording of verbal behaviors by subjects serving in the roles of observer, teacher and students. All subjects used the Flanders' System of Interaction Analysis to record verbal behaviors of peer-microteaching sessions in which they either participated or acted as an observer. Specifically, the study was designed to investigate the extent to which subjects trained in the Flanders' System of Interaction Analysis who were placed in the role of observer, teacher or student in peer-microteaching lessons, agreed on the classification of verbal behaviors recorded during each lesson. The subjects classified the peer-microteaching verbal behaviors in which they themselves assumed one of the above roles. Thus, representative validity of the Flanders' System of Interaction Analysis was investigated by analyzing inter-observer agreement of verbal behaviors recorded by the active participants in an instructional situation as well as an observer of the same instructional situation.

**Definition of Terms**

*Representative Validity of Observational Data* (hereafter referred to as representative validity) - the degree to which an observer, teacher, and students agree on the classification of verbal behaviors occurring in an instructional event in which the observer and participants were involved. (In this study the verbal behavior occurred in a peer-microteaching lesson.)

*Construct Validity of Observational Data* (hereafter referred to as construct validity) - the degree to which observational data correlates with a construct generated by a theory and some measurable criteria thought to represent the construct.
Flanders' System of Interaction Analysis - an instrument used to classify the classroom verbal behavior of both teacher and students. All verbal statements which occur in the classroom are categorized in one of three major sections: (1) teacher talk; (2) student talk; and a separate category, (3) silence/confusion, for events which do not fit into teacher talk or student talk. The category of teacher talk is further divided into seven specific types of "teacher talk" and the category of student talk is divided into two specific types of "student talk." All categories are intended to be mutually exclusive, but as a ten category system they are intended to be totally inclusive of all verbal interaction occurring in a classroom.

Behavior - an observable verbal expression of the teacher or student which is significant and clear enough to be recorded using the Flanders' System of Interaction Analysis.

Participants - pre-service teachers who, as subjects in the study, assumed the role of observer, teacher or student in peer-microteaching lessons, and classified the verbal behaviors occurring in those peer-microteaching lessons by using the Flanders' System of Interaction Analysis.

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

c. Observer - a participant who viewed the behaviors of the teacher and students and later (along with the teacher and the students who participated in the microteaching lesson) recorded the behaviors from an audio tape of the lesson.

Peer-Microteaching - as used in this study refers to the teaching of a lesson by an undergraduate education major to other undergraduate education majors. The exercise was a scaled down teaching experience: the number of "students" was two; a maximum time limit of ten minutes was used for each lesson; and an obvious limit existed on the content presented. The microteaching format used in this study did not include the usual evaluation of teacher performance for later improvement, nor was the "teacher" instructed 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 two subjects who were playing the role of "students." In this peer-microteaching format a subject in the role of observer viewed the lesson.

Assumptions, Delimitations, and Limitations

Assumptions - In this study it was assumed that preservice teachers trained to tally at an inter-observer agreement level of .80 as calculated by the Scott coefficient, (54) would be able to record
verbal behaviors used in peer-microteaching lessons in a way that would be highly consistent with the constructs of the Flanders' System of Interaction Analysis as they perceived those verbal behaviors occurring as participants in, or observers of, peer-microteaching lessons. It was further assumed that participants in the study would behave in ways that would be typical of the roles of teacher or student when portraying those roles in peer-microteaching lessons. It was also assumed that the participants could, in listening to an immediate replay of an audio tape recording of the lesson in which they participated, cognitively and emotionally reassume the roles they played during the peer-microteaching lesson.

**Delimitations** - The subjects who participated in this study included only preservice teachers. The teacher and student verbal behaviors recorded were restricted to those which occurred during a peer-microteaching lesson as defined above.

**Limitations** - The preservice education students who served as subjects in this study were not controlled on such variables as sex, age, grade point average, discipline major, I.Q. or cultural background. No personality variables were used in the selection of subjects, nor were any existing interpersonal attitudes between students measured.

**Summary**

This investigation sought to study the representative validity of the Flanders' System of Interaction Analysis. It is designed to serve as a continuation of the area of investigation initiated by
Ernest H. Broadwater in his original study involving the representative validity of the Observational System for Instructional Analysis. (61)

In the present study, preservice teachers trained in the use of Flanders' System of Interaction Analysis tallied verbal behaviors from an audio tape recording of microteaching lessons in which they participated as an observer, teacher or student.

The increased use of observation systems both to improve classroom teaching and learning and in educational evaluation and research has been significant. The results of information gathered from a study on the representative validity of observational data recorded from peer-microteaching lessons by the participants in those lessons may have significant transfer value as to other uses of observational systems. Therefore, information on the representative validity of the Flanders' System of Interaction Analysis may provide useful information to those persons who are seeking to improve the quality of classroom instruction. Providing teachers, administrators, supervisors or researchers with more accurate observational data upon which they can base instructional decisions and/or draw instructional inferences can have a positive influence on improving the quality of classroom instruction.
CHAPTER II

REVIEW OF RELATED LITERATURE

Introduction

Chapter II is devoted to a review of the literature relevant to the instrument employed in this research and the problem studied. Flanders' System of Interaction Analysis is an instrument used to record classroom verbal behavior, and the problem investigated in this study involved the representative validity of observational data obtained through the use of this instrument.

The first section of this chapter contains a review of the statistical testing techniques traditionally employed to establish what users of classroom observational systems refer to as observer reliability. The second section will present a brief historical review of the development of the Flanders' System of Interaction Analysis. This section will also include a synopsis of research using Interaction Analysis with specific emphasis placed upon the methodology used in training observers to use observational systems. Previous attempts to establish the validity of observational data will also be reviewed. The third section of this chapter will review literature pertaining to the effects of certain perceptual factors which influence the objective encoding of classroom behaviors and briefly examine problems of behavior classification in category systems of classroom observational
systems. The fourth section will present a brief review of the observational instrument used in the Broadwater study and its relationship to the instrument used in the present research as well as reviewing the instrument used in the present study -- the Flanders' System of Interaction Analysis. The fourth section in this chapter also will introduce the concept and use of peer-microteaching in teacher education and its relationship to the present study.

**Coefficients of Reliability**

When using classroom observational systems it is often considered desirable to obtain an index of inter-observer agreement. The Scott coefficient (54) is one of the most frequently used tools in measuring inter-observer agreement for nominal scale coding. The Scott coefficient has been used in numerous studies to support both the reliability and the validity of observational data. In these studies, reliability has usually been used to mean interjudge reliability, that is, a measure of the extent to which two or more observers consistently agree in the encoding of the same event. These studies have also used the Scott coefficient to obtain some estimate of construct validity and in some instances, concurrent validity.

An example of the use of the Scott coefficient as an index of inter-observer reliability is described in a detailed description of the training practices used by Flanders as the director of an extensive inservice training project conducted at the University of Michigan. The project involved a two stage training program designed to prepare observers to become reliable in the use of Interaction Analysis. The
Scott coefficient was used as the measure of inter-observer agreement both in training and during actual classroom observation. According to Flanders, "Scott's method is unaffected by low frequencies, can be adapted to per cent figures, can be estimated more rapidly in the field, and is more sensitive at higher levels of reliability." (Emphasis added) (5;161)

The formula for the Scott coefficient is:

\[ \tilde{\pi} = \frac{P_o - P_e}{1 - P_e} \]

Where: \( \tilde{\pi} \) = coefficient of inter-observer agreement
\( P_o \) = observed proportion of agreement
\( P_e \) = proportion of agreement expected by chance.

\( P_o \) is derived by summing the differences in per cent agreement for each category regardless of sign and subtracting from one. \( P_e \) is derived by squaring the per cent of tallies of one observer for each category and summing the squares. Scott's \( \tilde{\pi} \) is "the amount that two observers exceed chance agreement divided by the amount that perfect agreement exceeds chance." (5;161)

Cohen (44) has developed a formula for a similar purpose with a correction factor included to account for what Cohen believes to be a questionable assumption on the part of Scott. Cohen interprets Scott as assuming that "the distribution of proportions over the categories for the population is known and is taken to be equal for the judges." (44;40) Therefore, Cohen computes his \( P_e \) by "finding
the product of the proportion for each of the categories for each observer and then summing the products." (Emphasis added) (44;39)

However, Scott's $\bar{N}$ and Cohen's $K$ are both indices of inter-observer agreement. The formula for the Cohen coefficient is:

$$K = \frac{P_o - P_e}{1 - P_e}$$

Where: $K =$ coefficient of inter-observer agreement
$P_o =$ proportion of units in which the judges agree
$P_e =$ proportion of units for which agreement is expected by chance.

When comparing inter-observer agreement, it is common for one of the observers to be considered an expert in the use of the system and his tallies become a standard against which the other observer is compared. It is assumed that the expert knows the system, its theoretical construct and the definition of categories, and is highly proficient in the use of the system. If this assumption is valid, then high agreement between the tallies of the observer and the expert provides an estimate of the construct validity of the trained observer's data. In situations involving two trained observers, neither of whom are considered experts, both Scott's and Cohen's coefficients yield an estimate of concurrent validity.
Observational Research

In order to verify the need to devote more research to the establishing of representative validity of classroom observation systems, this portion of Chapter II will focus on the development of Flanders' System of Interaction Analysis, the training of observers in the use of observational systems, and the inferences made from the observational data collected. Effort has been made to select studies which are representative of the typical methodology used in training observers and typical of uses made of the observational data. The sixteen volume anthology of observational systems compiled by Simon and Boyer (33) is evidence of the large number of observational instruments available for use. For many of these systems the observer training procedures are basically similar.

The category systems used in the studies reviewed in this chapter were developed by their designers from theoretical constructs. Therefore, when researchers report coefficients of inter-observer agreement they are, in many instances, using the coefficient as both a measure of inter-observer reliability and construct validity. Rarely do research reports indicate that the trained observers were taught by a person who is an expert regarding the underlying theoretical constructs of the system. For this reason, reports of researchers indicating coefficients of inter-observer agreement may actually represent estimates of concurrent validity.

Flanders' System of Interaction Analysis, one of the most widely known classroom observational systems (28), has grown out of
the original work of H. H. Anderson (6) who in 1936 developed a twenty category check list for use in observing teacher and student behavior. Anderson's list was used to measure what he called dominative and integrative teacher and pupil behavior. Anderson trained observers in the use of the check list until they were able to reach high inter-observer agreement. Trained observers then used the check list in a preschool setting to record the frequency and type of student-teacher behavior. Anderson states,

Revised categories of teacher contacts were offered, together with certain mental hygiene assumptions regarding the classroom behavior of teachers and children. It was shown that by these revised classifications of teacher's dominative and integrative behavior data could be reliably recorded. It was further shown that the behavior of the teachers had correlations with the behavior of the children. (6;152)

Thus, Anderson was able to support his theoretical construct of dominative and integrative teacher and student behavior through the use of data obtained from observers using his classroom check list. Therefore, the substantial inter-observer agreement obtained by the observers using Anderson's check list can be considered to be a measure of construct validity if one can assume that the observers did understand the theoretical construct of the system, and/or the observer's coding of behavior was highly consistent with the coding of Anderson or some other person who was highly familiar with the theoretical construct of dominative and integrative behavior.
At approximately the same time that Anderson was conducting his study, Lewin and two colleagues, Lippitt and White, designed a study to analyze the impact of different kinds of adult leadership on group behavior. (25) Lewin, Lippitt and White organized eleven-year-old boys into after-school-clubs and exposed each group to three different types of leadership: authoritarian, democratic and laissez-faire. The pattern Lewin, Lippitt and White defined as "authoritarian leadership" was similar to Anderson's dominative teacher behavior while "democratic leadership" was similar to Anderson's integrative teacher behavior. Conclusions of the Lewin, Lippitt and White study tended to confirm the general findings of the Anderson study in that adult leadership behavior influenced the behavior of the group members and thereby influenced the social-emotional climate of club meetings.

Both the Anderson and Lewin, Lippitt and White studies created considerable interest in the analysis of teacher behavior. Withall's research also made a significant contribution to Flanders' eventual development of the Flanders' System of Interaction Analysis. Flanders gives credit to Withall for his first experience in interaction analysis.

Withall demonstrated that a relatively simple classification of a teacher's verbal behavior into seven categories can produce an index of teacher behavior that is conceptually related to the dominative-integrative ratio of Anderson. Therefore, by modification of his category system, Withall conducted research into the examination of the influence of the classroom teacher on the social-emotional climate of the classroom. He showed that, "a simple classification of teacher
verbal statements into seven categories produced an index of teacher behavior almost identical to the integrative-dominative ratio described by Anderson. Withall trained observers to use his seven category system to record teacher verbal behavior. He obtained observer reliability or "objectivity" as Withall described inter-observer agreement, by having four trained judges independently record four different transcripts of classroom teaching.

Because the mean percentages of the judges' inter-observer agreement ranged from 64% to 75%, Withall concluded that, "a valid measure of social-emotional climate of groups is obtainable by categorizing teacher statements." Withall's conclusion was partially validated through the use of an experimental class of ten students. These students were asked to record their "feelings" at different points in the classroom lesson by pushing a button to indicate whether their feelings were positive or negative. From this portion of the study, Withall concluded that students tended to have a more positive attitude during the "learner-centered" part of the lesson and a more negative attitude during the "teacher-centered" part of the class. Withall's use of the terms "learner-centered" and "teacher-centered" is similar to Anderson's use of integrative and dominative behavior and Lewin, Lippitt and White's use of authoritarian and democratic leadership behavior. Withall's examination of student feelings was a close approximation of representative validity in that he used direct information from participants in the instructional event to confirm the influence of teacher behavior on the social-emotional climate of the classroom as measured by his observational instrument.
One additional influence on Flanders' creation of his observational system is the work of Thelen. (35) Thelen noted that in observing classroom teachers, each teacher has a teaching style of his own. The teacher functions in his classroom according to the model he has in his mind. Thelen describes several models: Socratic discussion; The town meeting; Apprenticeship; Boss-Employee or Army Model; The business deal; The good old team and The guided tour. Thelen states that:

Models like the above help the teacher define the working relationship between himself and his class. They serve to clarify the role of the teacher; his power, his concerns, his style of teaching. The roles of the students are not differentiated in these models--all the students are expected to behave alike in ways which enable the teacher to enforce his role. (35;41)

Thelen concluded that his model theory provides two basic principles for understanding how a teacher can consciously and by design change the classroom atmosphere. First, the teacher is the main determiner of his own behavior and secondly, roles open to students are dependent upon the role the teacher chooses to play. Once again, the influence of teacher behavior on the behavior of students is indicated, and the influence of this relationship on the social-emotional climate of the classroom is implied.

It was from the above studies and research that Flanders developed his classroom observational system called Interaction Analysis. According to Flanders, "Interaction analysis is a label that refers to any technique for studying the chain of classroom events in such a fashion that each event is taken into consideration." (15;5)
Acknowledging the influence on the research of Anderson, Lewin, Lippitt and White, Thelen and Withall, Flanders developed his interaction analysis observation system on a construct of the influence of teacher behavior on the social-emotional climate of the classroom. According to Flanders:

The term 'classroom climate' refers to generalized attitudes toward the teacher and class that the pupils share in common despite individual differences. The development of these attitudes is an outgrowth of classroom social interaction. As a result of participating in classroom activities, pupils soon develop common attitudes about how they like their class, the kind of person the teacher is, and how he will act in certain typical situations. These common attitudes color all aspects of classroom behavior, creating a social atmosphere, or climate, that appears to be fairly stable, once established. (20:3)

Flanders used the research of the studies cited above to arrive at the two teacher behavior patterns which he believes create constrasting classroom climates: (20:6)

<table>
<thead>
<tr>
<th>The Integrative Pattern</th>
<th>The Dominative Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Accepts, clarifies, and supports the ideas and feelings of pupils.</td>
<td>a. Expresses or lectures about own ideas or knowledge.</td>
</tr>
<tr>
<td>b. Praises and encourages.</td>
<td>b. Gives directions or orders.</td>
</tr>
<tr>
<td>c. Asks questions to stimulate pupil participation in decision-making.</td>
<td>c. Criticizes or deprecates pupil behavior with intent to change it.</td>
</tr>
<tr>
<td>d. Asks questions to orient pupils to schoolwork.</td>
<td>d. Justifies own position or authority.</td>
</tr>
</tbody>
</table>

Flanders developed an observational system which reflected the influence of his view of the classroom climate. In its basic ten category format, Flanders' System allows for the recording of teacher direct and teacher indirect verbal behavior and student verbal behavior.
Flanders' System depends upon classroom observers trained to record classroom verbal behavior according to the System's defined categories. Using the Scott coefficient (54) to determine when observers were trained to an acceptable point of inter-observer agreement, Flanders then used these trained observers to gather classroom observational data. The data could then be plotted on a matrix and a teacher's total verbal behavior pattern analyzed in relation to Flanders' construct of indirect and direct teacher verbal behavior.

In addition to examining data related to direct and indirect teacher influence, Flanders has attempted to draw a relationship between teacher influence and student attitude and student achievement. (20) In Flanders' classic New Zealand and Minnesota studies conducted in 1955-1957, and his Minnesota study of 1959-60, he used trained observers to collect observational data which he then correlated with student attitudes about school and student achievement. From the observational data, Flanders identified teachers who tended to be either direct or indirect according to their verbal behaviors as recorded through the use of his observational system. A student attitude inventory was administered to 37 seventh grade English-social studies classes and 38 eighth grade mathematics classes. Sixteen mathematics teachers and fifteen English-social studies teachers were classified on the basis of their natural teaching style as using verbal behaviors which were either direct or indirect. Pre- and post tests were administered to measure student achievement prior to and after being taught a two-week unit specifically designed for the study
and taught by the fifteen English-social studies teachers and the sixteen mathematics teachers.

From these studies, Flanders concluded that, "teachers of classes that scored high on liking the teacher, motivation, fair rewards and punishments, lack of anxiety, and independence used more indirect influence, while the teachers of classes that scored low used less indirect influence." (20:64) He also found that "in those mathematics and social studies classrooms in which the teacher had a higher i/d ratio, (category frequencies 1 + 2 + 3 divided by 6 + 7 renders what Flanders calls the revised indirect/direct ratio) the students scored significantly higher on a measure of achievement controlled for initial ability." (20:96)

Thus in view of the construct upon which Flanders' System is based, the findings would indicate that teacher direct and indirect verbal behavior have varying effects on the social-emotional climate of a classroom and this climate influences both the achievement and attitudes of students. By correlating observational data, gathered by persons trained to be reliable observers, with measures of student attitude and student achievement, Flanders was examining a relationship between his construct of social-emotional classroom climate and potentially related student variables. Through this process Flanders was indirectly establishing the representative validity of his classroom observational system.

Amidon and Flanders investigated the effects of direct versus indirect teacher behavior on the geometry achievement of dependent
prone and independent prone eighth grade students. (36) From their study they concluded that:

Compared with students in general, dependent-prone students are apparently more sensitive to the influence pattern of a geometry teacher. . . We assume that dependent-prone students are more sensitive to the directive aspects of the teacher's behavior. As the teacher becomes more directive, this type of student finds increased satisfaction in more compliance, often with less understanding of the problem solving steps carried out. (36;290)

The specially trained teacher in this study performed his classroom role of "direct" or "indirect" teacher in accordance with Flanders' definition of those two teaching styles as measured by his own observational instrument.

Lohman studied the effect of pre-service training in interaction analysis on the verbal behavior of student teachers. (65) Using two groups of students selected from a larger population of a preceding study, Lohman compared the verbal behaviors of one group of student teachers who had been trained in Flanders' System of Interaction Analysis with another group of student teachers who had been taught to analyze teaching without the guidance of a formal classroom observation system. The observers in the study were trained by an assumed expert and used a thirteen category modification of the Flanders' System. Observer reliability was checked six times using the Scott coefficient with a mean reliability of .76. Observation was carried out in the classrooms in which the student teachers were teaching. Observational data gathered in the study enabled Lohman to affirm that significant differences in the verbal behavior of those student
teachers trained in the use of Interaction Analysis with the group receiving another treatment did occur. Lohman found that:

Student teachers trained in the Flanders' System of Interaction Analysis differed significantly from student teachers not so trained in the use of the following verbal behaviors:
1. They used less direct teacher talk.
2. They lectured less.
3. They gave fewer directions.
4. The extendedness of their direct teacher talk was less.
5. They used more indirect teacher talk.
6. They accepted and clarified student ideas more.
7. They used more extended indirect teacher talk.
8. They used more indirect teacher behavior as opposed to direct teacher verbal behavior.

In addition, there was significantly more student talk and significantly more spontaneous student talk in classes taught by student teachers trained in the Flanders' System of Interaction Analysis. (65;105,108)

Due to the initial training of the observers by an assumed expert in the Flanders' System, the observational data of the Lohman study may be considered to have construct validity.

A study by Hill examined two questions of importance to the present study: What are the effects on the verbal behavior of in-service teachers trained in the use of Flanders' System of Interaction Analysis and what effects do length of training time and mode of feedback have on changing teacher verbal behavior. (63) The 35 in-service teachers who served as subjects in the study were observed before training in Flanders' System, after training in the System and again in a delayed post-training observation. Hill's study used the observational data to conclude that the in-service teachers did change
their verbal behavior but the changes were not related to the training procedures (varying training time 6, 8 or 10 hours) nor to the method used to provide feedback. Hill used Amidon as his standard to establish reliability and the assumed construct validity of the data and the Scott coefficient to establish inter-observer agreement. Amidon is a close associate of Flanders and is a known expert in the use of the Flanders' System.

Carline investigated the relationships between various verbal patterns of teaching behavior and the arithmetic achievement of elementary students using Flanders' System of Interaction Analysis as both the major training instrument and data gathering device. (62) Carline trained the in-service teachers in one school through learning the Flanders' System to increase their indirect verbal behavior while in-service teachers in the control school were not trained in the System but received placebo treatment designed to minimize the Hawthorne effect. The design of Carline's study presented a major variation of Flanders' early investigation into the achievement of students, as related to teachers who, on the basis of their natural teaching style, were classified as either verbally direct or indirect. Carline's study showed that in-service teachers trained in Interaction Analysis became more indirect in their verbal teaching behavior than the control group. However, the data failed to show any significant connection between increased indirect teaching strategies and student achievement. Thus, efforts to establish representative validity by indirect means using the relationship between direct/indirect teacher verbal behavior
and student achievement are not supported by Carline's study. This finding takes on added significance when Carline's study was designed to provide a confirmation of the relationship between indirect verbal teaching strategies and increased student achievement. This finding calls to question other findings which have been based on similar methods of gathering and using observational data to show a positive relationship between teacher verbal behavior and student achievement. Carline's findings also raise a question concerning the problem of changing the "natural" teaching style of the teacher. By inducing change through training a teacher to assume a particular classroom verbal behavior pattern, is the observation sampling recorded by the observer a representative sample of the teacher's real behavior? This problem is complicated by the added effects of the presence of an observer in the classroom.

The three preceding studies have demonstrated how Flanders' System of Interaction Analysis has been used in training preservice and inservice teachers. Each study used data gathered by trained observers to evaluate the effects of training in the use of Interaction Analysis on the verbal behavior used by the trainees in actual classroom teaching. Each study confirmed the position that training in Interaction Analysis does have an effect on teacher verbal behavior. Thus, if teachers, trained in the use of an observational system, can modify their verbal behavior as a result of the training and trained observers record that change, this would seem to be a further confirmation of the system's construct validity. This is also a step closer to
the establishment of representative validity in that the observer and one of the active participants in the instructional situation being observed are functioning within the same framework.

The role of the observer is a critical one in the use of observational systems. Samph conducted a study designed to ascertain the effect of the observer on teacher classroom verbal behavior. (68) Flanders' System of Interaction Analysis was used to record teacher verbal behavior. In a careful design, an observer would record the teacher's verbal behavior: a) in the classroom with the teacher's advanced knowledge; b) by a monitoring system without the teacher's knowledge; c) in the classroom without the teacher's advanced knowledge; and d) by monitoring outside the classroom when the teacher was anticipating the observer's being present. Analysis of the data demonstrated a definite existence of observer effect, and substantiated the conclusion that the presence of a classroom observer does lead to changes in a teacher's verbal behavior. This finding raises the question of the ability of a classroom observer to obtain representative samples of verbal behavior if participants vary their verbal behavior in the presence of an "outsider."

Masling and Stern also investigated the effect of the observer in the classroom. (51) They used seven trained observers and obtained high inter-observer agreement while finding no consistent pattern of observer effect. In this study, the observer was always present in the classroom while recording the observation. An alternate conclusion suggested by the data was:
The effects of the observer are more complex than had been foreseen and affect various aspects of teacher and pupil behavior differentially. The implications for teacher and children to have an observer stationed in the back of the class, surveying all and saying nothing, writing cryptic notes, and noting all that occurs with clipboard, stopwatch, coded categories, and fixed, scientific expression are not well understood. (51;353)

The inconsistency in the reported effect of an observer in the classroom has significant importance for this study. In all of the studies cited above, conclusions have been drawn from data gathered by an observer who was present in the classroom (the only exception being those instances in the Samph study in which classroom verbal behavior was monitored.) The question of the effect of the observer's presence causes additional concern in establishing the reliability and representative validity of data collected through the use of an observational system.

Good and Brophy bring to question yet another point in the confidence which might be placed on the data gathered through the use of an observational system. (22) They state:

Traditional interaction analysis studies treat the classroom as a group as the unit of analysis. This usage involves two key assumptions: (1) the interaction variables being studied are properly conceptualized as interactions between the teacher and the class as a group and (b) teachers are consistent across students in their classroom behavior so that individual differences within a classroom are of little or no importance relative to inter-class difference. (22;15)

Good and Brophy demonstrated that "many coding categories are most properly conceptualized as interactions between teacher and individual students." (22;15) Therefore, the observers in their
study recorded behaviors between a teacher and individual students and they concluded that teacher verbal behavior does vary with individual students. Good and Brophy found sex differences and differences based on the academic standing of the student influenced teacher-student interaction. They also found teacher appraisal and interactive behavior differed with individual students.

Davis and Dollard in their early study of the treatment of inner-city students found a difference in the manner in which a teacher used praise and criticism based upon economic levels of the student. (13) Anderson and Brewer in their early research found that teachers tend to focus dominative and integrative verbal behavior on only a few students thus making the psychological environment in the same room different from each student. (6) Jackson and Lahaderne, in a recent study, arrive at a similar conclusion: "For at least a few students, individual contact with the teacher is as rare as if they were seated in a class of 100 or more pupils, even though there are actually only 30 or so classmates present." (64:13)

These studies tend to document a possible weakness in the use of generalizing teacher verbal treatment of students from data obtained by observational systems which treat the individuals in a classroom as a single group when, as the above studies suggest, the verbal behavior of the teacher is specific to the student. Therefore, the question of representative validity is again raised with respect to the validity of the observer's sampling of verbal behavior being generalized as having the same effect on all students.
The studies reviewed thus far have been representative of studies involving observational systems. Flanders' System of Interaction Analysis, the system being investigated in the present study, has also been the interaction analysis instrument used in several of the studies previously cited. A generalized pattern used to establish construct validity has been to train observers in the use of an observation system to agree with a standard or expert in the use of the system. The validity of observational data has typically been established by trained observers who obtain a high degree of inter-observer agreement between other observers and an expert in the system.

This same approach has been used to indirectly establish representative validity of observational data by using the obtained data and establishing a relationship to a theoretically related dependent measure such as student academic achievement. The lack of research into the representative validity indicates a real need for research in this area. Considering the critical role observers play in providing data from which such conclusions as the influence a certain teacher verbal pattern has on student achievement, a definite need exists to see if in fact the observer is recording the same behavior being perceived by the major participants in the classroom--the teacher and the student.

In reviewing typical observer training programs, little attention has been given to the consistency with which a trained observer uses the individual categories of a system. Little information is available to demonstrate the reliability of an observer to consistently repeat
the same category tally for the same sample of verbal behavior over time.

One of the major inferences based upon observational data is that a certain type of teacher verbal behavior will influence student achievement in a certain direction. However, the Carline study (62) was designed to substantiate this contention and it failed to do so. This failure of a direct effort to support the relationship between teacher verbal behavior and student achievement, a major theoretical construct used to establish a system's validity, indicates a need for further research into the representative validity of observational systems.

Rosenshine has contributed research to the question of teacher verbal behavior and student achievement. (28) He summarizes more than fifty studies in which some measure of observed teacher behavior was correlated to one or more measures of student achievement. His survey reviews studies which related such various teacher behaviors as use of criticism, praise, acceptance of student ideas, warmth, enthusiasm, cognitive structuring, questioning, flexibility and student-teacher interaction to student achievement.

Rosenshine's review of the research led him to be conservative in his reporting of the relationship between teacher behavior and student achievement. As an example, in discussing teacher use of praise, Rosenshine concludes that, "although there is a trend in favour of a positive relationship between teacher approval and pupil achievement, the directions of the correlations are inconsistent from one
study to the next." (28;69) In the area of teacher criticism he states that, "the review of research appears to indicate that there is no evidence to support a claim that a teacher should avoid telling a student that he is wrong, or should avoid giving academic directions. However, teachers who use a great deal of criticism appear consistently to have classes who achieve less in most subject areas." (28;61) In many of the areas cited above, Rosenshine found consistent but non-significant correlations or inconsistent findings between studies. Thus, the numerous variables, which may be associated with student achievement and are being analyzed through the use of observational systems, present a real challenge to researchers to collect data having representative validity in order to provide valid data needed to analyze the relationships between teacher behavior and student achievement. Rosenshine and Furst also confirm the position that:

"Representativeness' has received relatively little study. Insufficient attention has been given to determining whether a sample of observed classroom transactions is a trustworthy representative sample of total behavior. . . . If the purpose is to relate instructional activities in individual classes to outcomes obtained in these classes, then the problem of representativeness are perplexing and in need of further study. (29;169)

The problem of observational validity is further compounded by research such as Samph's (68) into the role of observer influence when present in the classroom showing a teacher can change his verbal behavior pattern when knowingly being observed. Studies into the response of students to teaching patterns such as Amidon and Flanders (36) and Good and Brophy (22) indicate that the role of the observer
is additionally difficult in having to be able to differentiate student perceptions of teacher verbal behaviors. These findings point out the influence of perceptual variables and the need to be able to measure observational data not only from the perspective of the classroom observer but also from the perspective of the teacher and the students, the active participants in the classroom.

Perceptual Variables

The topic of how one individual perceives another individual or situation is complex. Secord and Backman state that, "perception is most appropriately defined in terms of experiences that stem directly from sensory stimulation." (30;13) They go on to provide a guideline of four basic processes affecting perceptual responses:

1. The selectivity of perception and characteristic ways of organizing stimulus patterns. At any given moment, the perceiver responds to only a small portion of the sensory information provided by his environment, and he organizes it in certain ways.

2. The frequency of previous experience with particular stimulus patterns and responses. Later perceptions are affected by these previous experiences.

3. Experiences with stimuli and responses that have been positively or negatively reinforced. This reinforcement history also influences later perceptions.

4. The contemporary factors prevailing at the moment of perception. Certain current conditions, such as hunger, fatigue, or anxiety, may affect what is perceived. (30;14)

As an observer views the verbal action occurring in a classroom he must see the "total action" but as suggested by the first process
identified above, he may in fact see only a small sampling of what is happening. Thus the observer may see the effect of intended teacher praise directed toward a particular student but miss the effect it has on other students also present in the room. The second process would indicate that the observer may be influenced in his recording of verbal behavior by what has transpired earlier in the lesson. This is also the case with the third process. As observers are human, they too are subjected to numerous factors influencing them at the specific moment of observation as suggested in the fourth process listed above.

According to Peak, "Perception depends on the activation of psychological structure by patterns of stimuli which initiate events leading to the identification and classification of the stimuli. The perception of complex social stimuli will as a rule involve a series of perceptual acts." (27;350) Her research report supports the concept that the role of the observer can be a very complex one with the possibility of his recorded perception being influenced by numerous variables.

There are many factors which influence one's perceptions especially when combinations of factors are considered. Taft cites the following factors as possible influences on one's perceptions of others: age, intelligence and academic ability, family background, sibling rank, sex, emotional stability, self-insight, character integration, social skill and popularity.
Similar influencing factors are included in other sources. Norman cites multiple factors influencing a perceiver's view of others, such as insight into oneself, acceptance of self and perceived acceptance by others. (52) Shrauger and Altrocchi discuss the perceiver's personality as it affects perception as being one of a set of three variables influencing the perception of other people. (55) The three sets of variables are: "the attributes of the stimulus person, the nature of the interaction situation, and the characteristics of the perceiver." (55;289) These variables all interact in a complex manner to help shape person perception. Gage also found a positive correlation between accuracy of social perception with effectiveness in interpersonal relations. (48)

Biehler presents four factors which can interfere with an observer's objectivity: (10) The halo effect, defined as a "tendency for prior information or one impression or characteristic to influence all other impressions" (10;378); Cognitive dissonance, once a halo effect has been established the perceiver will "see" things that fit the halo impression and ignore information that does not fit; Projection, the tendency "to interpret the behavior of others in terms of the perceiver's personality --to 'project' his thoughts and feelings upon others." (10;378) The last factor is unconscious likes and dislikes and according to Biehler, "a much larger proportion of our experiences are in the realm of the unconscious; they persist as memories and may influence our behavior but they cannot be examined or understood because we are not aware of them. (10;379) Therefore, the classroom observer
has many factors including unconscious likes and dislikes which may influence his objectivity in recording the verbal behavior he "hears" and "sees" taking place in a classroom.

Combs and Snygg present their position as a perceptual view of human behavior and state that:

The facts of human behavior . . . are not the facts that exist for others but the facts that exist for the behaver. . . . If behavior is a function of perception, then a science of human relationships must concern itself with the meaning of events for the behaver as well as for the observer. Human feelings, attitudes, fears, hopes, wants, likes, and aversions cannot be set aside while we deal with objective events. The subjective aspects of human experience cannot be suspended from operation. Perceptions are the very fabric of which human relationships are made. (12;307-309)

Thus reality is viewed as the perspective of the observer. Gage extends this idea when he states:

Behavior of all kinds is governed by perception as well as by motives. If we assume a certain kind of motive, then satisfaction of it depends on our ability (a) to perceive the situation accurately and (b) to govern our subsequent behavior in accordance with the motivation and the perception. If we may assume sufficient integration of the personality, we may expect behavior appropriate to a given motivation when perceptions of the surrounding environment are accurate. We can test whether perceptions of personalities are accurate in terms of the accuracies of predictions of the behavior of those personalities. (48;15)

Taft includes family background as a factor which can influence the manner in which one perceives and is perceived and this brings into view the importance of previous experiences as a contributing factor in the matter of observer perception. (57) Gage and Cronbach make
reference to social perception as a process influenced more by what the observer brings to a situation than what he observes during the situation. (49) Asch presents a similar position in suggesting if one wants to know why a person perceives a situation the way he does, investigate the previous experiences which are relevant to the current situation rather than looking into the current situation for the answer. (41) This reference to past experience as a significant factor in observer perception is confirmed by Hyman in a discussion of frameworks for observing teaching. He states, "If I were to ask ten principals to describe a particular teacher, there is little doubt that we would get disagreement. The ten principals would probably disagree simply because they would focus on different things. And this brings us to the heart of our concern. When we observe--visually and/or auditorily--we do so from a particular vantage point." (50;313)

Taft in his research into the ability to judge people arrives at the following conclusion:

The main attributes of the ability to judge others seem to lie in three areas: possessing appropriate judgmental norms, judging ability, and motivation. Where J (judge) is similar in background to S (student) he has the advantage of being readily able to use appropriate norms for making his judgment. The relevant judging ability seems to be a combination of general intelligence and social intelligence, with the possibility of an additional specific factor for nonanalytic judgments ("intuition") . . . But probably the most important area of all is that of motivation: if the judge is motivated to make accurate judgments about his subject and if he feels himself free to be objective, then he has a good chance of achieving his aim, provided of course that he has the requisite ability and can use the appropriate judgmental norms. (57;21)
Bruner and Tagiuri in their research into the accuracy of judging found judges who tended to be relaxed made more accurate judgments than judges who tended to be more emotional. (11) Thus motivation and the emotional perspective of the observer can influence the quality of the encoding of classroom verbal behavior.

A major research study by Morse, Bloom and Dunn concluded that, "the appraisers cannot be substituted one for another. Each observer, teacher, pupil, outsider--sees different aspects of the classroom. Classroom research should take this fact into consideration." (66;53) This conclusion points out the need to include the perceptions of more than just the observer in trying to appraise classroom behavior. The perceptions of involved participants could help to improve the comprehensiveness in the observation of the instructional situation. By including the observations of the participants as well as the observer, perceptual problems pertaining to the role of "judge" also extend to these additional "judges."

The preceding review of perceptual variables serves to substantiate the complexity of the situation in which an observer enters a classroom to record the verbal behaviors of others. These studies have pointed out the difficulty an observer might have not only because of external influences but also because of his level of personal motivation and emotional involvement. The review of the literature on perceptual variables brings to question the validity of data obtained through an observer using an observational system to record the verbal behaviors of teachers and students. Thus, having raised the
question, the need for research into the representative validity of data obtained through the use of observational systems is further reinforced.

In addition to a review of the development and use of observational systems and perceptual variables already presented this section also will examine briefly behavior classification involved in category systems of classroom observational instruments. Hough and Duncan state that comprehensive category systems have two major characteristics. (23) Inclusiveness is the first characteristic. This "makes possible the classification of all events that occur in the classroom that are encompassed by the category system." (23;117) The second major characteristic is mutually exclusive categories. Those events occurring in the classroom can be categorized into only one of the categories used by that system. Therefore, in a comprehensive category system, it is necessary for the user to record all behaviors observed into a particular category designating that event.

Good and Brophy point out the problem of classifying behavior for one individual and allowing it to become representative of the group when, according to their theory, the category system should use the individual student as the unit for data gathering. (22) Bruner and Tagiuri contend that the nature and degree of the category distinction required on the part of the user can pose a recording problem. (11) The complexity and involvement of the decision making process as well as the necessity for making fine category distinctions in a limited time span can increase the difficulty of the observer's
This is when the characteristic of mutually exclusive categories becomes critical. Biddle contends that, "all observer rating techniques suffer both from limitation of content and from unreliability of the encoding process." (42;339-340) Thus he is agreeing with Good and Brophy in suggesting that an observer is incapable of observing simultaneously the total behaviors occurring in the instructional situation.

Belland, Belland and Price studied the classification of teacher questions using a system designed specifically for analyzing classroom questions. (60) They compared their instrument with the Observational System for Instructional Analysis in order to evaluate differences in decision-making while recording verbal behavior. Belland, Belland and Price concluded that the complexity of the decision-making process required of the observer can influence his classification of verbal behaviors.

This has special significance for the Observational System for Instructional Analysis in which the observer must always make at least two decisions when recording each different verbal behavior and as many as four decisions in some instances. The Broadwater study (61) reduced the multiple decision making process by modifying the Observational System for Instructional Analysis into a nine category system (plus the Silence and Confusion categories.) This modification helped make the decision process basically the same as the Flanders' System in that the person recording the verbal behaviors had only one decision to make and that was to select the correct verbal category.
Premac poses another problem in selecting proper category designation. (53) The problem he poses involves recording appraisal behavior as his research demonstrates that reinforcement is relative to the individual. Thus when a teacher attempts to praise a student, the student may or may not perceive the teacher's verbal behavior as reinforcing. Because one student does perceive the teacher's behavior as reinforcing does not mean that same verbal behavior will provide positive reinforcement to other class members. Findings by Good and Brophy tend to support this position. (22) Findings in the Broadwater study also tend to support the difficulty of judging appraisal behavior as the category of confirmation was tallied with significant disagreement between the subjects and the researcher. (61;79-80)

From the above studies it can be assumed that the category distinctions and the necessary decisions required of the observer can influence the recorded observational data. Therefore, in order to confirm representative validity of an observational system, the value of having data recorded by the participants as well as the observer is once again confirmed.

**Flanders' System of Interaction Analysis/ Broadwater Study/ Peer-Microteaching**

The observational system investigated in this study for the purpose of establishing its representative validity was the Flanders' System of Interaction Analysis. Subjects were trained by the researcher to use the Flanders' System in order to record classroom verbal behavior during microteaching lessons. A complete description of the system,
its uses and directions for use, can be found in *Analyzing Teaching Behavior*. (15)

This system was selected by the researcher, because of his experience with and exposure to the system. It was also selected because it is the system being used for in-service teacher improvement by the public school system in the community in which the university is located at which the researcher is presently teaching. However, the major reason for selecting Flanders' System of Interaction Analysis is that it is the most widely used observational system in both teacher training and research. Rosenshine states, "the interaction analysis system developed by Flanders is the best known and most frequently used category system." (28;18)

The Flanders' System of Interaction Analysis is a comprehensive category system for the classification of classroom verbal behavior. In its basic form, it is a ten category system with seven categories devoted to teacher verbal behavior and two categories devoted to student verbal behavior. One category is included to account for all other events not included in the first nine categories. The procedure for using the system involves the classification of verbal statements every three seconds and the tabulation of data in special matrices for later analysis. The first four teacher verbal categories are considered by Flanders to be indirect teacher verbal behavior and the next three teacher verbal categories to be direct teacher verbal behavior. The system was developed and designed to analyze Flanders' theory of the influence of teacher verbal behavior on the attitude and
achievement of students. For this reason it is considered to be a social-emotional observational system. The Flanders' System is based "on a set of social skills used by teachers to control and manage class activities. This model is based on a psychology of superior-subordinate relationships, adapted to fit classroom conditions." (20;2)

Rosenshine and Furst, in a discussion of the use of direct observation to study teaching, describe a paradigm for studying teaching in the natural setting of the classroom. (29) Their model contains three elements the sequence of which is not fixed:

The Descriptive-Correlational-Experimental Loop

1. development of procedures for describing teaching in a quantitative manner;

2. correlational studies in which the descriptive variables are related to measures of student growth;

3. experimental studies in which the significant variables obtained in the correlational studies are tested in a more controlled situation. (29;122)

They describe the investigation of Flanders as one of the best examples of the type of research which proceeds from descriptive studies to correlational studies to experimental studies. (29;125) This traces Flanders' development of his observational system around 1957 to his correlational studies of teacher verbal behavior with student achievement in the Minnesota studies of 1960 to the experimental studies of Amidon and Flanders of 1961 dealing with the effects of direct and indirect teacher influence on dependent-prone students.
In describing the four dominant purposes for which observational systems are used, Rosenshine and Furst state, "the system developed by Flanders has been used extensively in teacher training, in research on student growth, and in describing teaching, and has also been used to monitor specific programs." (29:159) Thus the Flanders' System has been used for all four of the dominant purposes served by observational systems.

Rosenshine and Furst also state that:

The existing system most frequently cited as a source of a new instrument is the system developed by Flanders. Several reasons may account for the popularity of this system. First, the instrument is relatively easy to learn to use. Second, several manuals and kits are readily available for training observers. Third, the system includes a matrix to record dyadic sequences, and this structure appeals to many users. Fourth, more research has been conducted with this instrument than with any others. (29:142)

The present study was designed to build on to the knowledge produced by the Broadwater study into the representative validity of observational systems. The major difference between the Broadwater study and the present study is the observational system being investigated. Broadwater used a modified version of the Observational System for Instructional Analysis developed by Hough and Duncan. (61)

Both studies were conducted at the same small, private, midwestern university. The present researcher assisted Broadwater in a limited manner with his original research.

The Observational System for Instructional Analysis was developed to describe patterns of verbal behavior which help to define instructional strategies used by teachers and students in the classroom.
This system was not developed to test any theoretical assumptions about classroom verbal behavior but developed as a descriptive instrument. This is a basic difference between Flanders' System and the Observational System for Instructional Analysis. Both systems are similar in the manner in which classroom verbal behaviors are recorded. Each system has been used to train in-service and pre-service teachers and each system uses trained observers to record classroom verbal behavior.

The findings of the Broadwater study indicated that subjects serving in the three different peer-microteaching roles were in basic agreement regarding the recording of verbal behaviors. Student inter-observer agreement tended to be higher than teacher-observer agreement. Teacher-student inter-observer agreement was lower than observer-student agreement. The category of "confirmation" was recorded with significant disagreement by subjects in all roles. (61;79-80) Morse, Bloom and Dunn in their research on the relationships among pupil, teacher and observer perceptions tend to confirm the position that observer-student agreement occurs more frequently than teacher-student agreement or teacher-observer agreement. (66)

The Broadwater study indicated additional research was needed in order to gain a better understanding of the meaning behind differences in recording verbal behavior according to roles. This research is an effort to continue the investigation into representative validity initiated by Broadwater.

The last ten years have seen a significant increase in the use of microteaching as both a research device and a teaching method effective in work with both pre-service and in-service teachers.
Developed by Allen and others in 1960, microteaching is not an entirely new concept in improving teaching but it has taken on new importance with the current advances and availability of audio- and videotape equipment. Microteaching is basically a scaled down version of regular classroom teaching. Through microteaching, a small group of students can be used as an experimental group for the purpose of single concept teaching. New feedback systems allow the pre-service teacher to microteach a lesson, review the lesson and then re-teach it and receive additional feedback. Allen and Ryan indicate the value of micro-teaching as a research tool since its development at Stanford University. (1) Flanders indicates a national survey conducted by Johnson in 1968, revealed that fifty-three per cent of all teacher-education programs were using microteaching, including four per cent which reported using microteaching extensively. (1;xiii) Peer-microteaching proved to be a very valuable research tool in the present study.

In a study by Ober, it was demonstrated that verbal behaviors of pre-service teacher trainees obtained during peer-microteaching in a methods course were indicative of their verbal behavior during student teaching. (67) Ober compared thirty students with training in Flanders' System (using a modified 13-category version) with thirty students who were not trained in the use of Flanders' System of Interaction Analysis. In the Ober study, subjects used peers as micro-teaching students while the Stanford model used actual school students.
The fact that peer-microteaching behavior has been shown to be an effective predictor of student teaching behavior supports the use of peer-microteaching in the present study. It is assumed that peer-microteaching provides a situation containing certain elements which make peer-microteaching a valid context in which to test the hypothesis being investigated by the present study.

Summary

The basic purpose of this chapter has been to present literature which would support the need for research into representative validity of observational data, and to present problems related to that research. This chapter has reviewed literature related to the study of representative validity in relation to the use of observational systems for gathering data involving classroom verbal behavior. Basic statistical testing methods used to establish a measure of observer reliability were presented.

This chapter has also presented an historical review of the development leading to and including Flanders' System of Interaction Analysis. Representative studies of the use of observational systems in research and problems of observer objectivity and influencing perceptual variables which might affect the establishing of representative validity of observational data were presented.

Included in this chapter was a discussion of the Flanders' System of Interaction Analysis and the Broadwater study, the initial research and design upon which the present study is based. Discussion of peer-microteaching as a realistic research and teaching device concluded this chapter.
CHAPTER III

METHODS AND PROCEDURES

The purpose of this chapter is to present the questions that were researched in this study, and the hypotheses related to those questions. This chapter will also describe the training procedures used on the subjects who participated in the study and the experimental setting. Finally, this chapter will present the method of data collection and analysis.

Context

The major purpose of this study was to investigate the representative validity of the Flanders' System of Interaction Analysis by analyzing inter-observer agreement of verbal behaviors recorded by an observer, teacher, and students, all of whom observed or participated in the same instructional situation. In the questions and hypotheses that follow, for the purposes of this study the term "inter-observer agreement" is used to represent the agreement between two or more people with respect to the classification of verbal behaviors used in peer-microteaching sessions using Flanders' System of Interaction Analysis as the observational instrument. All inter-observer agreement measures were obtained by using the Cohen coefficient. (44) The framework within which the research questions were
asked involved an N of sixteen subjects who were randomly assigned to four groups of four subjects per group. Each subject served in three different roles within his group; observer, teacher and student (each participant served in the role of student twice).

This study seeks to investigate six questions. The first three questions and null hypotheses pertain to the inter-observer agreement obtained during the total sixteen peer-microteaching lessons. The fourth question and null hypothesis pertains to the inter-observer agreement obtained during the four peer-microteaching lessons taught by each of the four groups. In questions one through four inclusive, inter-observer agreement refers to a comparison of the consistency of subject tally agreement for each lesson. The fifth question and null hypothesis refers to inter-observer agreement between the experimenter and subjects with respect to eighteen verbal behaviors selected by the experimenter from each of the sixteen microtaught lessons. The sixth question seeks to compare the findings of a previous study involving the investigation of the representative validity of the Observational System for Instructional Analysis with the present investigation into the representative validity of the Flanders' System of Interaction Analysis.

**Hypotheses**

In this study, research into the representative validity of data obtained through the use of the Flanders' System of Interaction Analysis was guided by the following questions and hypotheses:
Question 1: Will a subject serving in one of three roles (teacher, student or observer) classify verbal behaviors differently than other subjects also serving in one of the three roles when recording the same instructional situation using the Flanders' System of Interaction Analysis? For example, will a subject serving in the role of "student" record the same category for a specific verbal behavior as a subject serving in the role of observer, teacher or another student? Specifically stated: Given six possible combinations of pairs of subjects, will the inter-observer agreement of verbal behaviors of any one pair of subjects differ from the inter-observer agreement of any other combination of pairs of subjects? The following six combinations of pairs are possible when four persons are matched:

T - S₁, T - S₂, T - 0, 0 - S₁, 0 - S₂, S₁ - S₂ where:

T = teacher; S₁ = student one; S₂ = student two; and 0 = observer.

Hypothesis 1: There will be no difference in the inter-observer agreement of the classification of verbal behaviors of any of the six possible pairs of subjects.
Question 2: Will the inter-observer agreement of any two pairs of subjects differ in the classification of verbal behaviors when using the Flanders' System of Interaction Analysis to record the sixteen peer-microteaching lessons? The fifteen two-pair combinations are listed below:

\[
\begin{align*}
T - S_1 & \rightarrow T - S_2 \\
T - S_1 & \rightarrow T - O \\
T - S_1 & \rightarrow 0 - S_1 \\
T - S_1 & \rightarrow S_1 - S_2 \\
T - O & \rightarrow 0 - S_1 \\
T - O & \rightarrow 0 - S_2 \\
T - 0 & \rightarrow S_1 - S_2 \\
T - S_2 & \rightarrow T - 0 \\
T - S_2 & \rightarrow 0 - S_1 \\
T - S_2 & \rightarrow 0 - S_2 \\
0 - S_1 & \rightarrow 0 - S_2 \\
0 - S_1 & \rightarrow S_1 - S_2 \\
0 - S_1 & \rightarrow S_1 - S_2 \\
0 - S_2 & \rightarrow S_1 - S_2 \\
\end{align*}
\]

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

Question 3: Will the inter-observer agreement in the classification of peer-microteaching verbal behaviors of the teacher and students differ from the inter-observer agreement of the observer and students for the same verbal behaviors when using the Flanders' System of Interaction Analysis? The thirty-two two-pair combinations of T - Ss will be compared with the thirty-two two-pair combinations of 0 - Ss:

\[
T - S_1, T - S_2 \rightarrow 0 - S_1, 0 - S_2
\]
Hypothesis 3: There will be no difference in the inter-observer agreement of the classification of verbal behaviors of teacher and students when compared with the inter-observer agreement of the observer and students.

Question 4: The preceding questions examined possible differences in inter-observer agreement for the total sixteen peer-microteaching lessons. The fourth question attempts to examine differences within each of the four groups. Will the inter-observer agreement of verbal behaviors of any one pair of subjects differ from the inter-observer agreement of any other combination of pairs of subjects for each of the four lessons within each of the four groups when using the Flanders' System of Interaction Analysis? Once again, the following six combinations of pairs are possible when four persons are matched:

T - S₁, T - S₂, T - O, O - S₁, O - S₂, S₁ - S₂

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

Question 5: Do subjects serving in the role of teacher, student or observer differ in the classification of specific verbal behaviors which occur in the lessons in
which those subjects participated when compared to a standard while using the Flanders' System of Interaction Analysis to record those verbal behaviors?

In answering this question the researcher served as the standard. This question involved the following five persons:

\[ T \rightarrow S_1 \rightarrow S_2 \rightarrow O \rightarrow R \]

where: \( R \) = researcher.

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

Question 6: This general question was used to examine and compare the findings of the Broadwater investigation into the representative validity of the Observational System for Instructional Analysis (61) with the present researcher's investigation of the Flanders' System of Interaction Analysis. What implications can be drawn between the results of research into the representative validity of two different classroom observational systems?

A comparison will be made of the relationship between the behaviors tallied using Flanders' System of Interaction Analysis and comparable categories in the Observational System for Instructional Analysis. As no two observational systems are identical, a comparison between the categories
of the Observational System for Instructional Analysis and Interaction Analysis must take into account slight inter-category differences. Analysis of data between the two systems will be made on the basis of the following comparison of categories: (See Figures 1 and 2 for category description.)

### TABLE 1

<table>
<thead>
<tr>
<th>FSIA Categories</th>
<th>OSIA Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/7</td>
</tr>
<tr>
<td>2</td>
<td>6/8</td>
</tr>
<tr>
<td>3</td>
<td>1/7</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>2/3</td>
</tr>
<tr>
<td>6</td>
<td>2/4</td>
</tr>
<tr>
<td>7</td>
<td>5/9</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>1/2/3/4</td>
</tr>
<tr>
<td>10</td>
<td>10/11*</td>
</tr>
</tbody>
</table>

*The category comparison is based upon the modified version of the Observational System for Instructional Analysis used in the Broadwater study. In that study, categories 10 and 11 indicated Silence and Confusion respectively. Broadwater did not use the T (teacher) S (student) behavior symbol differentiation procedure, and did not use the Managerial categories. (61:43-46)
**FIGURE 1**

Flanders' System of Interaction Analysis

<table>
<thead>
<tr>
<th>TEACHER TALK</th>
<th>STUDENT TALK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1. ACCEPTS FEELING: accepts and clarifies the feeling tone of the students in a nonthreatening manner. Feelings may be positive or negative. Predicting or recalling feelings is included.</td>
<td></td>
</tr>
<tr>
<td>2. PRAISES OR ENCOURAGES: praises or encourages student action or behavior. Jokes that release tension, but not at the expense of another individual; nodding head, or saying &quot;um hm?&quot; or &quot;go on&quot; are included.</td>
<td></td>
</tr>
<tr>
<td>3. ACCEPTS OR USES IDEAS OF STUDENTS: clarifying, building, or developing ideas suggested by a student. As teacher brings more of his own ideas into play, shift to Category 5.</td>
<td></td>
</tr>
<tr>
<td>4. ASKS QUESTIONS: asking a question about content or procedure with the intent that a student answer.</td>
<td></td>
</tr>
<tr>
<td>5. LECTURING: giving facts or opinions about content or procedures; expressing his own ideas, asking rhetorical questions.</td>
<td></td>
</tr>
<tr>
<td>6. GIVING DIRECTIONS: directions, commands, or orders with which a student is expected to comply.</td>
<td></td>
</tr>
<tr>
<td>7. CRITICIZING OR JUSTIFYING AUTHORITY: statements intended to change student behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.</td>
<td></td>
</tr>
<tr>
<td>8. STUDENT TALK - RESPONSE: talk by students in response to teacher. Teacher initiates the contact or solicits student statement.</td>
<td></td>
</tr>
<tr>
<td>9. STUDENT TALK - INITIATION: talk by students, which they initiate. If &quot;calling on&quot; student is only to indicate who may talk next, observer must decide whether student wanted to talk. If he did, use this category.</td>
<td></td>
</tr>
<tr>
<td>10. SILENCE OR CONFUSION: pauses, short periods of silence, and periods of confusion in which communication cannot be understood by the observer.</td>
<td></td>
</tr>
</tbody>
</table>

Source: (4;14)
**FIGURE 2**

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>

*Modified version of the Observational System for Instructional Analysis as used in the Broadwater Study (61) adapted from *Teaching: Description and Analysis* by John B. Hough and James K. Duncan. (23)
Training

The individuals who served as subjects in this study were all undergraduate students at a small, private midwestern university. All the subjects were enrolled in a required education course for students preparing to teach at the secondary level. While there is only one section of this course offered each term, there is no reason to believe the students enrolled were not representative of the university's student body. This researcher was the course instructor and the entire class was exposed to the same initial training in the use of Flanders' System of Interaction Analysis. The sixteen members of the class who were eventually selected to serve as subjects volunteered to continue with the training and received credit in the course on the same basis as the remaining students who did not participate in this study. None of the students has had previous training in behavioral analysis nor had they participated in microteaching as used in this research. Therefore, the learning of Flanders' Interaction Analysis System and participating in microteaching lessons was a new experience for all the subjects.

Total training time covered a six week period. During the training period each subject received a minimum of three hours of training per week as dictated by the class structure of two class sessions of one and one-half hours each week. The means by which the subjects were trained in Flanders' Interaction Analysis System will be presented in chronological order.
The initial training session was devoted to an orientation to and overview of the Flanders' System of Interaction Analysis. The subjects were informed of the background, uses and basic operation of Flanders' System. Each subject was given a copy of the ten verbal categories with a brief description of each category. The subjects were instructed to memorize the categories by code symbol and the researcher provided verbal examples of each of the categories and answered questions asked by the subjects. Flanders' category descriptions as presented to the subjects is presented in Figure 1.

The subjects were then presented with a set of flash cards numbered from one to ten. The researcher again gave verbal examples of different categories and the subjects would indicate the category by holding up the flash card with the number corresponding to the category. This method proved most useful to the subjects and the researcher as it provided instant feedback. Having determined that the subjects had acquired a basic understanding of the categories and of the methodology of the system, the researcher concluded the first session by playing Exercises 1 - 4, Session 1 of the audio tape from the Interaction Analysis Training Kit -- Level 1.(3) The subjects used their flash cards to practice recording their first recorded classroom audio tape.

The second training session began by pairing the subjects and having a review session of the material covered in the first session. Subjects continued to work in dyads periodically throughout the training sessions. While working in dyads the subjects would review
the categories with one subject defining a category or holding up a flash card and the other subject either stating the number of the defined category or defining the category representing the number on the flash card. After the review session was complete, rules for tallying were presented. The subjects received the following instructions:

1. Tally a category code every three seconds.
2. Should the verbal behavior change in less than three seconds, that change in behavior should also be tallied.
3. When more than one person is speaking at a time so that it is impossible to understand what is being said, tally a category 10.
4. When a period of silence lasting for at least three seconds occurs tally a category 10.

Several examples of tallying were presented and questions from the subjects were answered. The subjects then tallied several brief sample classroom transcripts presented orally by the researcher. Following this practice, each subject was given a tally work sheet which they used to tally Exercises 1–4, Session I of the Interaction Analysis Training Kit. The remainder of the second training session was used in practicing tallying by moving on into Session II of the Training Kit. Some work in dyads and use of the flash cards helped the subjects gain confidence in their tallying and answered questions involving more difficult category discrepancies.
The third training session involved continued practice in tallying and review work as the subjects moved on through the Interaction Analysis Training Kit -- Level I. The basic ground rules about classifying statements were also presented at this session. The subjects received the following ground rules for classifying verbal behaviors:

**Rule No. 1** When not certain to which of two or more categories a statement belongs, choose the category that is numerically farthest from Category 5. This is advisable except when one of the two categories in doubt is Category 10, which is never chosen if there is an alternate category under consideration.

**Rule No. 2** If the primary tone of the teacher's behavior has been consistently direct or consistently indirect, do not shift into the opposite classification unless a clear indication of shift is given by the teacher. The trained observer who is observing a particular action is in the best position to judge whether or not the teacher is restricting or expanding the freedom of action of class members.

**Rule No. 3** The observer must not be concerned with his own biases or with the teacher's intent. Rather, he must ask himself the question, "What does this behavior mean to the students so far as restriction or expansion of their freedom is concerned?"

**Rule No. 4** If more than one category occurs during the three-second interval, then all categories used in that interval are recorded; thus, record each change in category. If no change occurs within three seconds, repeat the previous category number.

**Rule No. 5** Directions are statements that result (or are expected to result) in observable behavior of children. Examples of direction are: "Go to the board, read question 3, go to your seat, etc." Some teacher statements sound like directions, but will not be followed by observed student compliance. These statements often precede the actual direction. For example, "let's get ready now to go to recess." (Orientation, Category 5) "Now, Row 5, get your coats." (Category 6)
Rule No. 6 When the teacher calls on a child by name, the observer ordinarily records a 4.

Rule No. 7 If there is a discernible period of silence (at least 3 seconds), record one 10 for every three seconds of silence, laughter, board work, etc.

Rule No. 8 When the teacher repeats a student answer, and the answer is a correct answer, this is recorded a 2. This tells the student he has the right answer and therefore functions as praise.

Rule No. 9 When the teacher repeats a student idea and communicates only that the idea will be considered or accepted as something to be discussed, a 3 is used.

Rule No. 10 If a student begins talking after another student (without the teacher's talking), a 10 is inserted between the 9's or 8's to indicate the change of students.

Rule No. 11 Statements such as "uh huh, yes, yeah, all right, okay," which occur between two 9's, are recorded as 2 (encouragement). These statements function as encouragement (the student continues talking after the 2) and are therefore classified as 2.

Rule No. 12 A teacher joke, which is not made at the expense of the children, is a 2. If the joke makes fun of a child, then it is coded as a 7.

Rule No. 13 Rhetorical questions are not really questions; they are merely part of lecturing techniques and should be categorized as 5's.

Rule No. 14 A narrow question is a signal to expect an 8. If the student gives a specific predictable answer, this is an 8. If the child expands, documents, or justifies his answer, the observer should begin tallying 9's.

Rule No. 15 When the class members respond to a question or direction in unison with a single discernible response, an 8 rather than a 10 is recorded. (3;71-74)

Succeeding training sessions followed a similar pattern as the training tapes increased in length and difficulty. This researcher introduced several audio tapes designed to overcome difficulties in
certain categories and continued periodic work in dyads and with flash cards. The researcher would tally the training tapes on the blackboard while the subjects tallied on work sheets at their seats. Discrepancies would be discussed and resolved. Subjects were encouraged to ask questions should any problems become apparent. Copies of the training tapes were placed in the audio-visual department of the university's library to enable the subjects to review any tape at their discretion. Several subjects took advantage of this additional learning opportunity.

Selection of Subjects

During the sixth, seventh and eighth training sessions the researcher collected the tally sheets recorded by the subjects from audio tapes prepared by the researcher. Those subjects who obtained Scott coefficients of .80 or better based upon the standard established by the researcher were asked to continue with some additional training. During the next three training sessions, seventeen subjects obtained a Scott coefficient of .80 or better on three consecutive audio tapes, again using the researcher as the standard. The sixteen subjects with the highest Scott coefficients were selected for peer-microteaching. Upon completion of the four peer-microteaching sessions, a final audio tape check was administered ten days after the pre-testing which qualified the final sixteen subjects. This final check showed no reduction in subject Scott coefficients. The researcher again served as the standard for this final check. Table 2 presents the Scott coefficients for the three consecutive inter-observer agreement tally checks which served as the minimum qualification for each subject and the final Scott coefficient inter-observer agreement tally check which followed completion of the four peer-microteaching sessions.
## TABLE 2

**SCOTT COEFFICIENTS FOR SIXTEEN SUBJECTS**

<table>
<thead>
<tr>
<th>Subject</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>Pre Peer-Microteaching</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>Post Peer-Microteaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.92</td>
<td>.89</td>
<td>.82</td>
<td></td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.81</td>
<td>.82</td>
<td>.81</td>
<td></td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.81</td>
<td>.89</td>
<td>.91</td>
<td></td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.82</td>
<td>.86</td>
<td>.85</td>
<td></td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.92</td>
<td>.81</td>
<td>.84</td>
<td></td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.87</td>
<td>.89</td>
<td>.84</td>
<td></td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.85</td>
<td>.82</td>
<td>.86</td>
<td></td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.89</td>
<td>.81</td>
<td>.82</td>
<td></td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>.87</td>
<td>.88</td>
<td>.89</td>
<td></td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.87</td>
<td>.85</td>
<td>.89</td>
<td></td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>.82</td>
<td>.81</td>
<td>.80</td>
<td></td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>.81</td>
<td>.87</td>
<td>.90</td>
<td></td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>.82</td>
<td>.80</td>
<td>.87</td>
<td></td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>.82</td>
<td>.84</td>
<td>.91</td>
<td></td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>.87</td>
<td>.80</td>
<td>.85</td>
<td></td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>.87</td>
<td>.93</td>
<td>.87</td>
<td></td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Peer-Microteaching Lessons

The final sixteen subjects having been selected for this study, they were each presented with the following written instructions to help establish a proper learning set for their individual microteaching lesson:

1. The students you will be teaching will be members of this class, but not necessarily in the same discipline.

2. Insofar as is possible the lesson should be a real teaching experience for you and a real learning experience for the students. Therefore: Choose a content that has the following characteristics:
   (a) The content is likely to be new to the students, not something they already know.
   (b) The content is not so complex that it cannot be taught in 10 minutes MAXIMUM.
   (c) Though the students may not know the content, they would be capable of learning it in 10 minutes without your providing a great deal of background information to get them to the point of being ready to learn the main content (objective) of the lesson.

3. The lesson should allow you to demonstrate a wide variety of teaching behaviors. Therefore: Choose instructional procedures that have the following characteristics:
   (a) Provides for a reasonable balance of teacher and student talk.
   (b) Involves questioning and responding, or class discussion mixed (if desirable) with short periods of information giving. In other words, for the purpose of this lesson (even though it may be very appropriate in other situations) do not plan to lecture for 10 minutes and do not have the students discuss something for 10 minutes. Such lessons would not provide an opportunity for you to demonstrate that you can use a wide variety of teaching verbal behaviors.
4. Among the important teaching behaviors that are not often used in great quantity are clarification of student ideas, acceptance of student feelings and appraisal behavior of praising or correcting students and accepting their ideas. See how many of these behaviors you can use without making the lesson seem artificial.

5. Therefore, when teaching, plan to incorporate in your lesson at least two (2) verbal examples of each of Flanders' categories (1-7) and when acting as the student plan to use both 8's and 9's.

6. Ten minutes is not a very long time in which to really teach something. Therefore: Plan the lesson carefully ahead of time and even "rehearse" the lesson in your own mind so that you can anticipate problems that you may encounter. Plan alternative strategies should your main plan fail to work as you had hoped it would. Finally, know the content well enough so that you do not have to rely on notes as you teach the lesson.

In addition to these written instructions, this researcher added strong verbal emphasis to the importance of attempting to incorporate into each lesson at least two examples of each verbal behavior.

The sixteen subjects were randomly assigned a time block for the presentation of their lesson. Four three-hour sessions were held with four different subjects participating in each session. All four peer-microteaching sessions were completed within one week of the conclusion of the training program.

At each peer-microteaching session, the four subjects were randomly assigned their role order for the evening. Each subject served once in the role of teacher and observer and twice in the role of student.

The environment selected for the presentation of the peer-microteaching sessions was the same classroom as was normally used for the regular class. As the peer-microteaching sessions were held
in the evening hours there was minimal distraction with only the researcher and four subjects present at each session.

**Data Collection**

In each peer-microteaching session the subject serving in the role of teacher taught two other subjects who were serving in the role of student and these three subjects were observed by the fourth subject who was serving in the role of observer. The subject serving as observer watched from a short distance away as the teacher and student subjects proceeded with the microteaching lesson. Each peer-microteaching lesson was recorded on an audio tape and the quality of each tape was consistently clear and audible. Immediately following each lesson, the tape was played back and each of the four subjects tallied the verbal behaviors recorded during the lesson using Flanders' System of Interaction Analysis. Upon completion of the first playback, the tape was re-played. During the second playback the recorder was stopped at eighteen different points and the subjects were instructed to tally the specific verbal behavior occurring as the recorder was stopped. This same procedure was repeated for each microteaching lesson.

Eighteen verbal behaviors were selected in order to allow for two occurrences of each verbal category with the exception of category 10, "Silence or Confusion." Thus, a sampling of verbal behaviors classified as indirect (categories 1 through 4) and direct (categories 5 through 7) teacher verbal behavior and student verbal behavior (categories 8 and 9) was made possible. This was the primary
reason for selecting eighteen verbal stops. A secondary consideration was to provide a verbal sampling distributed over the entire microteaching lesson. Each of these objectives had to remain flexible and at the discretion of this researcher in order to achieve the desired sampling. In some microteaching lessons certain verbal categories simply did not occur -- category 7, "Criticizing or Justifying Authority" is an example of one verbal category sometimes omitted. When a category was omitted, the researcher arbitrarily selected additional verbal stops to insure a total sample of eighteen verbal behaviors. Allowing the researcher the flexibility of selecting the points at which the verbal behaviors were sampled served to strengthen the study because it enabled the researcher to "catch" such categories as 1 and 7 which occurred infrequently in several lessons.

Training of the subjects to tally classroom verbal behaviors using the Flanders' System of Interaction Analysis and using those subjects to perform in classroom roles and then to tally their own verbal behaviors was critical in attempting to obtain valid data. Through the methodology and format used in this research, the subjects provided the source for gathering observational data needed to investigate the representative validity of the Flanders' System of Interaction Analysis.

Data Analysis

The initial review of the data involved a survey of the total number of tallies in order to gain a perspective of the total sampling
of verbal behaviors. This data was obtained by summing the total number of tallies recorded by each subject for each of the sixteen peer-microteaching lessons.

The Cohen coefficient was selected to measure the amount of agreement between the recorded tallies of pairs of subjects. The tallies of each subject were compared with the tallies of each of the other subjects in computing a Cohen coefficient. This process yielded a total of ninety-six Cohen coefficients; six Cohen coefficients for each of the sixteen peer-microteaching lessons. This analysis permitted the researcher to examine and compare subjects according to the role they assumed in each microteaching lesson. It was from this data that the measure of representative validity was drawn in an effort to realize the major goal of this research study.

The Cohen coefficient is used for nominal scales and yields a measure of agreement between two sets of recorded behavior. As discussed in Chapter II, the Cohen coefficient is a sensitive instrument in measuring frequency differences between two sets of recorded data. As the Cohen coefficient is more sensitive to the distributions of the proportions in the various categories within an observational system, its use was significant in this research which analyzed and interpreted differences in recorded verbal behaviors.

The first research question sought to determine if any Cohen coefficients of the six groups of pairs differed from any other groups of pairs over the sixteen peer-microteaching lessons. Observational information used to answer the second research question helped to
clarify and contribute to data collected to answer the first research question by determining which, if any, groups of pairs differed. The Cohen coefficients represented an index of the extent to which two subjects serving in their respective roles agreed upon the classification of the verbal behaviors recorded in the peer-microteaching lessons.

The first research hypothesis examined the Cohen coefficients for pairs of subjects and the inter-observer agreement between any other pair for the sixteen peer-microteaching lessons. The Kruskal-Wallis one-way analysis of variance by ranks was used to test this hypothesis. The data used in this analysis fulfilled the basic assumptions of this statistical test. According to Siegel, the non-parametric Kruskal-Wallis test can compare favorably with the parametric F test. (31) Therefore, the Kruskal-Wallis H was used to determine if differences in the Cohen coefficients for pairs of subjects could have occurred by chance.

The second research hypothesis was set up to test the fifteen sets of pairs of Cohen coefficients for the sixteen peer-microteaching lessons. The Mann-Whitney U test is a non-parametric test and was used in comparing the fifteen sets of pairs of Cohen coefficients to determine if the Cohen coefficients of one pair of recorders differed from a second pair.

The third research question was answered on the basis of comparing the recorded observations of the two subjects serving as students with the subject serving as teacher and the same two student subjects with the subject serving as observer. The Cohen coefficients for these pairings were used to examine inter-observer agreement.
The third research hypothesis again used the Mann-Whitney U test. It was used to determine if the Cohen coefficients of students - teacher differed from the Cohen coefficients of the students - observer.

The fourth research question used the four subjects' Cohen coefficients within each of the four groups to determine if the individual subjects were consistent in recording the verbal behaviors for each of the four peer-microteaching lessons taught by a single four-subject group. This question was answered by using each subject as a member of a single peer-microteaching group four times.

The fourth hypothesis used the Kruskal-Wallis test to compare each of the four peer-microteaching groups. The Kruskal-Wallis H was used to determine if the Cohen coefficients for each of the four subject-taught lessons differed within each of the four peer-microteaching groups.

The fifth research question involved an analysis of the eighteen selected verbal behaviors within each lesson. The researcher who trained the subjects also tallied the eighteen verbal behaviors and thereby served as a standard by which the tallies of the subjects could be compared. From this observational information, a measure of the consistency with which each of the categories was recorded was made possible.

Research hypothesis five, like the fifth research question, did not use Cohen coefficients. Instead, the eighteen selected verbal behaviors were analyzed to determine inter-observer agreement between each subject and the researcher. The eighteen
selected verbal behaviors were summed by individual categories for
the sixteen microlessons and recorded under the role of the person
tallying the verbal behavior. This enabled the researcher to analyze
both frequency of difference according to verbal category and fre-
quency of difference according to role. Subjects' tallies for each
category were recorded as correct or incorrect on the basis of the
standard and plotted in a two by four contingency table by subject
role. A chi-square test was used to determine if frequencies by
roles within a particular category were significant. The chi-square
test demonstrated if the teacher, students and observer differed
significantly from the standard set by this researcher.

The observational data collected in this study were analyzed by
the use of a computer program package designated NPAR for non-
parametric statistical test. These programs were based upon research
in non-parametric statistics by Siegel and designed for use in the
behavioral sciences. (31) A .10 level of significance was used in
testing the null hypotheses of this research study. All tests were
two-tailed tests. The .10 level of significance was selected due to
the exploratory nature of this study and therefore, a greater willing-
ness to risk a Type I error (rejecting a null hypothesis when in fact
it is true) rather than a Type II error (accepting as true a null
hypothesis when in fact it is false.) The .10 level of significance
was also selected in order to facilitate comparison between the find-
ings of the present study with the findings of the Broadwater study.

Research question six was answered by drawing comparisons
between the implications of the representative validity of the
Observational System for Instructional Analysis as reported in
the Broadwater study and the representative validity of the Flanders' System of Interaction Analysis as investigated in the present study. In order to make a comparison between the present research and the Broadwater study possible, careful effort was made to replicate the design methodology originally used in the Broadwater study. (61)

Summary

This chapter has presented the basic research questions and hypotheses investigated in this study. Flanders' System of Interaction Analysis was described along with the ground rules for implementation of the system. The methodology in the training of the subjects and the selection of the subjects was presented followed by a description of the peer-microteaching lessons used in the gathering of data for this study. Finally, means by which the six research questions were answered and the statistical tests used in the treatment and analysis of the five research hypotheses were described.
CHAPTER IV

PRESENTATION AND ANALYSIS OF RESULTS

The purpose of this chapter is to report the research findings obtained in an investigation of the representative validity of the Flanders' System of Interaction Analysis. Each of the research hypotheses will be reviewed and an analysis of the data by which the hypotheses were either accepted or rejected will be presented. Information relevant to the research question involving a comparison of the findings in the present study to the findings of the Broadwater study will also be presented.

All subjects, having been trained by the researcher in the use of Flanders' System of Interaction Analysis, used Flanders' System to record both the frequency and type of verbal behaviors used in peer-microteaching lessons recorded on audio-tape and tallied immediately following the lesson. These tallies, denoting the coded verbal behavior recorded by each subject, were then summed by categories according to the roles played by each of the four subjects. Table 3 presents the mean per cent of the verbal behaviors in each category for each of the sixteen microteaching lessons. The mean per cent of the recorded verbal behaviors are shown for each of the four peer-microteaching lessons which occurred within each of the
### TABLE 3

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>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td></td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>10</td>
<td>23</td>
<td>4</td>
<td>2</td>
<td>23</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>75</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>27</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>14</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>45</td>
</tr>
<tr>
<td>Group II</td>
<td></td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>14</td>
<td>25</td>
<td>7</td>
<td>1</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>20</td>
<td>14</td>
<td>3</td>
<td>2</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>60</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td>47</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Group III</td>
<td></td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>15</td>
<td>8</td>
<td>23</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td>43</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>11</td>
<td>22</td>
<td>6</td>
<td>3</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>5</td>
<td>11</td>
<td>25</td>
<td>7</td>
<td>2</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>Group IV</td>
<td></td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>27</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>6</td>
<td>17</td>
<td>30</td>
<td>6</td>
<td>1</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>15</td>
<td>25</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>2</td>
<td>11</td>
<td>6</td>
<td>9</td>
<td>19</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Total Mean Per Cents</td>
<td></td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>11</td>
<td>31</td>
<td>4</td>
<td>2</td>
<td>12</td>
<td>19</td>
<td>9</td>
</tr>
</tbody>
</table>
four peer-microteaching groups. Roman numerals designate the four peer-
microteaching groups while Arabic numerals designate the four individual
peer-microteaching lessons taught in each group. Total mean per cents
for the sixteen peer-microteaching lessons are presented at the bottom
of Table 3.

Table 3 indicates a fairly similar pattern of verbal behavior
throughout the sixteen microteaching lessons. An exception exists in
Lesson 2, Group I. In this lesson, the mean per cent of Category 5
(Lecturing) accounted for approximately seventy-five per cent of the
total verbal behavior. In Lesson 4, Group I, the mean per cent of
Category 9 behavior (Student Talk - Initiation) was forty-five per cent
and the mean per cent of Category 7 behavior (Criticizing or Justifying
Authority) was less than one half per cent. In Lesson 1, Group III
again there was a mean of less than one half per cent of Category 7
behavior and this same lesson had the highest total mean per cent of
Categories 1, 2 and 3 (Accepts Feeling, Praises or Encourages and Ac-
cepts or Uses Ideas of Students). Lesson 4, Group III was the only
other lesson containing a verbal category (Category 1 — Accepts Feeling)
with a mean of less than one half per cent.

Overall, Teacher Talk (Categories 1 through 7 inclusive)
accounted for approximately sixty-one per cent of the total verbal
behavior and Student Talk represented approximately thirty-one per cent.
Category 10 (Silence or Confusion) accounted for the remaining nine
per cent of the tallied behaviors. It should be noted that all sub-
jects began and ended their encoding with a tally of ten, a procedure
required for the plotting of a matrix used in graphically displaying all tallied behaviors. In addition to its use in recording three or more seconds of silence, Category 10 was also used to separate student talk when one student began talking immediately after another student had finished talking with no intervening teacher talk. (This complies with Flanders' Ground Rule No. 10.) Thus, the rule of beginning and ending each encoding session with a tally of Category 10 and this category's use in separating student talk helped to increase its total per cent of tallied behaviors.

The per cents of tallies for each category for each subject for each lesson were used in computing the Cohen coefficients. These Cohen coefficients were used in testing the first four research hypotheses. By using the Cohen coefficient formula, a coefficient of inter-observer agreement was computed for the total recorded tallies of all possible pairs of subjects for each peer-microteaching lesson. Using this procedure, there were six Cohen coefficients for the four subjects in each peer-microteaching lesson accounting for the four subject roles of teacher, observer and two students in each lesson. These six Cohen coefficients for each of the four peer-microteaching lessons for each of the four peer-microteaching groups are presented in Table 4.

A range of Cohen coefficients can be observed in most lessons with the widest range occurring in Lesson 2, Group IV, and the smallest range occurring in Lesson 4, Group III. The lowest coefficient was observed in the observer-student #2 pairing in Lesson 2, Group IV and the highest coefficient was observed in Lesson 4, Group I in the
TABLE 4
COHEN COEFFICIENTS FOR FOUR PEER-MICROTEACHING LESSONS
FOR THE FOUR GROUPS*  

<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 T S₁ S₂</td>
<td>1 T S₁ S₂</td>
<td>1 T S₁ S₂</td>
<td>1 T S₁ S₂</td>
<td>1 T S₁ S₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₁ .75</td>
<td>S₁ .77</td>
<td>S₁ .73</td>
<td>S₁ .83</td>
<td>S₁ .83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₂ .84 .83</td>
<td>S₂ .78 .75</td>
<td>S₂ .84 .67</td>
<td>S₂ .81 .84</td>
<td>S₂ .81 .84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 .62 .61 .61</td>
<td>0 .65 .79 .82</td>
<td>0 .76 .58 .74</td>
<td>0 .69 .71 .77</td>
<td>0 .69 .71 .77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 T S₁ S₂</td>
<td>2 T S₁ S₂</td>
<td>2 T S₁ S₂</td>
<td>2 T S₁ S₂</td>
<td>2 T S₁ S₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₁ .79</td>
<td>S₁ .63</td>
<td>S₁ .83</td>
<td>S₁ .73</td>
<td>S₁ .73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₂ .63 .80</td>
<td>S₂ .82 .69</td>
<td>S₂ .76 .83</td>
<td>S₂ .85 .87</td>
<td>S₂ .85 .87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 .85 .88 .72</td>
<td>0 .76 .58 .75</td>
<td>0 .69 .75 .82</td>
<td>0 .58 .47 .46</td>
<td>0 .58 .47 .46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 T S₁ S₂</td>
<td>3 T S₁ S₂</td>
<td>3 T S₁ S₂</td>
<td>3 T S₁ S₂</td>
<td>3 T S₁ S₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₁ .74</td>
<td>S₁ .66</td>
<td>S₁ .62</td>
<td>S₁ .69</td>
<td>S₁ .69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₂ .57 .51</td>
<td>S₂ .79 .59</td>
<td>S₂ .67 .63</td>
<td>S₂ .75 .60</td>
<td>S₂ .75 .60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 .80 .78 .56</td>
<td>0 .75 .64 .80</td>
<td>0 .73 .77 .59</td>
<td>0 .66 .59 .81</td>
<td>0 .66 .59 .81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 T S₁ S₂</td>
<td>4 T S₁ S₂</td>
<td>4 T S₁ S₂</td>
<td>4 T S₁ S₂</td>
<td>4 T S₁ S₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₁ .85</td>
<td>S₁ .84</td>
<td>S₁ .88</td>
<td>S₁ .71</td>
<td>S₁ .71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₂ .85 .92</td>
<td>S₂ .78 .76</td>
<td>S₂ .80 .85</td>
<td>S₂ .70 .83</td>
<td>S₂ .70 .83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 .91 .87 .81</td>
<td>0 .88 .82 .82</td>
<td>0 .90 .83 .83</td>
<td>0 .79 .74 .77</td>
<td>0 .79 .74 .77</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Where T = Teacher; S₁ = Student #1; S₂ = Student #2; O = Observer
student-student pair. Lesson 4, Group 1 also had the highest mean coefficient of .87 while Lesson 3, Group 1 and Lesson 2, Group IV had the lowest mean coefficient of .66. Table 5 presents the mean coefficients for the four peer-microteaching lessons for each of the four peer-microteaching groups.

**TABLE 5**

**MEAN COHEN COEFFICIENTS FOR FOUR PEER-MICROTEACHING LESSONS FOR THE FOUR GROUPS**

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
<th>Group IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.71</td>
<td>.76</td>
<td>.72</td>
<td>.78</td>
</tr>
<tr>
<td>2</td>
<td>.78</td>
<td>.71</td>
<td>.78</td>
<td>.66</td>
</tr>
<tr>
<td>3</td>
<td>.66</td>
<td>.71</td>
<td>.67</td>
<td>.68</td>
</tr>
<tr>
<td>4</td>
<td>.87</td>
<td>.81</td>
<td>.85</td>
<td>.76</td>
</tr>
</tbody>
</table>

Group Mean .76 .75 .76 .72

A Kruskal-Wallis one-way analysis of variance was used to test the first hypothesis which stated: There will be no difference in the inter-observer agreement of the classification of verbal behaviors of any of the six possible pairs of subjects. Table 6 presents the Cohen coefficients, ranks, rank sums and Kruskal-Wallis H value found in
TABLE 6
KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE WITH PARTICIPANT AND OBSERVER
COHEN COEFFICIENTS AND RANKS FOR SIXTEEN PEER-MICROTEACHING LESSONS*

<table>
<thead>
<tr>
<th>T-S₁</th>
<th>T-S₂</th>
<th>T-₀</th>
<th>O-S₁</th>
<th>O-S₂</th>
<th>S₁-S₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen</td>
<td>Rank</td>
<td>Cohen</td>
<td>Rank</td>
<td>Cohen</td>
<td>Rank</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.75</td>
<td>42.5</td>
<td>.84</td>
<td>81.5</td>
<td>.62</td>
<td>15.5</td>
</tr>
<tr>
<td>.79</td>
<td>58.5</td>
<td>.63</td>
<td>18</td>
<td>.85</td>
<td>86</td>
</tr>
<tr>
<td>.74</td>
<td>38</td>
<td>.57</td>
<td>5</td>
<td>.80</td>
<td>62.5</td>
</tr>
<tr>
<td>.85</td>
<td>86</td>
<td>.85</td>
<td>86</td>
<td>.91</td>
<td>95</td>
</tr>
<tr>
<td>.77</td>
<td>51.5</td>
<td>.78</td>
<td>55</td>
<td>.65</td>
<td>21</td>
</tr>
<tr>
<td>.63</td>
<td>18</td>
<td>.82</td>
<td>70</td>
<td>.76</td>
<td>47.5</td>
</tr>
<tr>
<td>.66</td>
<td>22.5</td>
<td>.79</td>
<td>58.5</td>
<td>.75</td>
<td>42.5</td>
</tr>
<tr>
<td>.84</td>
<td>81.5</td>
<td>.78</td>
<td>55</td>
<td>.88</td>
<td>92</td>
</tr>
<tr>
<td>.73</td>
<td>35</td>
<td>.84</td>
<td>81.5</td>
<td>.76</td>
<td>47.5</td>
</tr>
<tr>
<td>.83</td>
<td>76</td>
<td>.76</td>
<td>47.5</td>
<td>.69</td>
<td>27.5</td>
</tr>
<tr>
<td>.62</td>
<td>15.5</td>
<td>.67</td>
<td>24.5</td>
<td>.73</td>
<td>35</td>
</tr>
<tr>
<td>.88</td>
<td>92</td>
<td>.80</td>
<td>62.5</td>
<td>.90</td>
<td>94</td>
</tr>
<tr>
<td>.83</td>
<td>76</td>
<td>.81</td>
<td>66</td>
<td>.69</td>
<td>27.5</td>
</tr>
<tr>
<td>.73</td>
<td>35</td>
<td>.85</td>
<td>86</td>
<td>.58</td>
<td>7</td>
</tr>
<tr>
<td>.69</td>
<td>27.5</td>
<td>.75</td>
<td>42.5</td>
<td>.66</td>
<td>22.5</td>
</tr>
<tr>
<td>.71</td>
<td>31.5</td>
<td>.70</td>
<td>30</td>
<td>.79</td>
<td>58.5</td>
</tr>
</tbody>
</table>

Rank
Sums  787.0  869.5  781.5  664.0  725.5  828.5
H  = 2.16 N.S.

*Where T = Teacher; S₁ = Student #1; S₂ = Student #2; O = Observer
H ≥ 9.24 required for p ≤ .10
testing the first hypothesis. Table 6 is organized by the six pairs of subjects according to role over the sixteen microteaching lessons.

The Kruskal-Wallis H of 2.16 is not significant at the .10 level. Therefore, the first null hypothesis cannot be rejected. Thus, the verbal behaviors recorded by each of the possible groups of pairs of subjects were not found to differ significantly from the verbal behaviors recorded by any other group of pairs of subjects.

As the H value of 2.16 did not allow for rejection of the first hypothesis, this finding would normally have eliminated the need to investigate the second and third hypotheses as they tend to extend the line of inquiry being examined by the first hypothesis. However, this study also was designed to replicate the Broadwater study (61) and for this reason, this study did continue with the examination of the second and third hypotheses.

The second hypothesis stated: There will be no difference in the inter-observer agreement of the classification of verbal behaviors of any two pairs of subjects. The Cohen coefficients for pairs of subjects were compared using the Mann-Whitney U test to determine whether or not any pairs of subjects differed from any other pair of subjects over the sixteen peer-microteaching lessons. Table 7 presents the Mann Whitney U values for the fifteen pair combinations possible for the sixteen peer-microteaching lessons.

None of the U values was significant at the .10 level. Once again, the data does not allow for the rejection of the null hypothesis. The verbal behaviors recorded by any pairs of subjects were not found
TABLE 7
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 −−−− T-S₂</td>
<td>110.5</td>
<td>T-S₂ −−−− T-0</td>
<td>112.5</td>
</tr>
<tr>
<td>T-S₁ −−−− T-0</td>
<td>128.0</td>
<td>T-S₂ −−−− S₁−S₂</td>
<td>124.5</td>
</tr>
<tr>
<td>T-S₁ −−−− S₁−S₂</td>
<td>124.5</td>
<td>T-S₂ −−−− S₁−S₂</td>
<td>95.5</td>
</tr>
<tr>
<td>T-S₁ −−−− 0-S₂</td>
<td>106.0</td>
<td>T-S₂ −−−− 0-S₂</td>
<td>103.5</td>
</tr>
<tr>
<td>T-S₁ −−−− 0-S₁</td>
<td>118.0</td>
<td>T-S₂ −−−− 0-S₁</td>
<td>106.5</td>
</tr>
<tr>
<td>T-0 −−−− S₁−S₂</td>
<td>125.5</td>
<td>O-S₁ −−−− S₁−S₂</td>
<td>102.0</td>
</tr>
<tr>
<td>T-0 −−−− 0-S₁</td>
<td>106.5</td>
<td>O-S₁ −−−− 0-S₂</td>
<td>118.0</td>
</tr>
<tr>
<td>T-0 −−−− 0-S₂</td>
<td>126.0</td>
<td>O-S₂ −−−− S₁−S₂</td>
<td>104.0</td>
</tr>
</tbody>
</table>

*T = Teacher; S₁ = Student #1; S₂ = Student #2; O = Observer
U ≤ 83 required for p ≤ .10 (n₁ = 16, n₂ = 16)

to be significantly different than the recorded verbal behaviors of any other pairs of subjects over the sixteen peer-microteaching lessons.

The third hypothesis stated: There will be no difference in the inter-observer agreement of the classification of verbal behaviors of teacher and students when compared with the inter-observer agreement of the observer and students. Table 8 presents the Cohen coefficients for teacher-students and observer-students, the rank sums and the Mann-Whitney U value based upon a comparison of the differences between the thirty-two Cohen coefficients for each group.
TABLE 8
MANN-WHITNEY U TEST FOR TEACHER-STUDENTS
AND OBSERVER-STUDENTS FOR SIXTEEN MICROTEACHING LESSONS*

<table>
<thead>
<tr>
<th>T-S&lt;sub&gt;S&lt;/sub&gt; Cohens</th>
<th>Ranks</th>
<th>O-S&lt;sub&gt;S&lt;/sub&gt; Cohens</th>
<th>Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>.75</td>
<td>28.5</td>
<td>.61</td>
<td>9.5</td>
</tr>
<tr>
<td>.79</td>
<td>40</td>
<td>.88</td>
<td>63.5</td>
</tr>
<tr>
<td>.74</td>
<td>25</td>
<td>.78</td>
<td>37</td>
</tr>
<tr>
<td>.85</td>
<td>60</td>
<td>.87</td>
<td>62</td>
</tr>
<tr>
<td>.77</td>
<td>33.5</td>
<td>.79</td>
<td>40</td>
</tr>
<tr>
<td>.63</td>
<td>12.5</td>
<td>.58</td>
<td>5.5</td>
</tr>
<tr>
<td>.66</td>
<td>15</td>
<td>.64</td>
<td>14</td>
</tr>
<tr>
<td>.84</td>
<td>57</td>
<td>.82</td>
<td>49</td>
</tr>
<tr>
<td>.73</td>
<td>22.5</td>
<td>.58</td>
<td>5.5</td>
</tr>
<tr>
<td>.83</td>
<td>53.5</td>
<td>.75</td>
<td>28.5</td>
</tr>
<tr>
<td>.62</td>
<td>11</td>
<td>.77</td>
<td>33.5</td>
</tr>
<tr>
<td>.88</td>
<td>63.5</td>
<td>.83</td>
<td>53.5</td>
</tr>
<tr>
<td>.83</td>
<td>53.5</td>
<td>.71</td>
<td>19.5</td>
</tr>
<tr>
<td>.73</td>
<td>22.5</td>
<td>.47</td>
<td>2</td>
</tr>
<tr>
<td>.69</td>
<td>17</td>
<td>.59</td>
<td>7.5</td>
</tr>
<tr>
<td>.71</td>
<td>19.5</td>
<td>.74</td>
<td>25</td>
</tr>
<tr>
<td>.84</td>
<td>57</td>
<td>.61</td>
<td>9.5</td>
</tr>
<tr>
<td>.63</td>
<td>12.5</td>
<td>.72</td>
<td>21</td>
</tr>
<tr>
<td>.57</td>
<td>4</td>
<td>.56</td>
<td>3</td>
</tr>
<tr>
<td>.85</td>
<td>60</td>
<td>.81</td>
<td>45</td>
</tr>
<tr>
<td>.78</td>
<td>37</td>
<td>.82</td>
<td>49</td>
</tr>
<tr>
<td>.82</td>
<td>49</td>
<td>.75</td>
<td>28.5</td>
</tr>
<tr>
<td>.79</td>
<td>40</td>
<td>.80</td>
<td>42.5</td>
</tr>
<tr>
<td>.78</td>
<td>37</td>
<td>.82</td>
<td>49</td>
</tr>
<tr>
<td>.84</td>
<td>57</td>
<td>.74</td>
<td>25</td>
</tr>
<tr>
<td>.76</td>
<td>31</td>
<td>.82</td>
<td>49</td>
</tr>
<tr>
<td>.67</td>
<td>16</td>
<td>.59</td>
<td>7.5</td>
</tr>
<tr>
<td>.80</td>
<td>42.5</td>
<td>.83</td>
<td>53.5</td>
</tr>
<tr>
<td>.81</td>
<td>45</td>
<td>.77</td>
<td>33.5</td>
</tr>
<tr>
<td>.85</td>
<td>60</td>
<td>.46</td>
<td>1</td>
</tr>
<tr>
<td>.75</td>
<td>28.5</td>
<td>.81</td>
<td>45</td>
</tr>
<tr>
<td>.70</td>
<td>18</td>
<td>.77</td>
<td>33.5</td>
</tr>
</tbody>
</table>

Rank Sums 1129 951

U = 423.0 N.S.
U ≥ 607 for p ≤ .10 (n<sub>1</sub> = 32, n<sub>2</sub> = 32)

*Where T-S<sub>S</sub> = Teacher with all students; O-S<sub>S</sub> = Observer with all students.
The Mann-Whitney U value based upon the above comparison was not significant at the .10 level. Therefore, the third hypothesis cannot be rejected. The verbal behaviors recorded by the teacher-subject and the two student-subjects did not differ significantly from the verbal behaviors recorded by the observer-subject and the two student-subjects.

The first three hypotheses examined the inter-observer agreement of the classification of verbal behaviors according to the roles assumed by the subjects in the sixteen peer-microteaching lessons. The fourth hypothesis examined inter-observer agreement of the classification of verbal behaviors by four subjects for four different microteaching lessons. In examining this hypothesis the role of the subject was no longer the critical factor but rather it was the consistency with which the same four subjects, regardless of role, recorded verbal behaviors during four different peer-microteaching lessons in which the subjects had either participated or observed.

The fourth hypothesis stated: There will be no differences in inter-observer agreement of the classification of verbal behaviors 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 this hypothesis using the six Cohen coefficients for each of the four peer-microteaching lessons within each of the four peer-microteaching groups. The Cohen coefficients, ranks, rank sums and Kruskal-Wallis H values are presented for each peer-microteaching group in Table 9.
TABLE 9
KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE FOR EACH OF THE FOUR MICROTEACHING GROUPS

<table>
<thead>
<tr>
<th>Group I</th>
<th>Microteaching Lessons</th>
<th>Group II</th>
<th>Microteaching Lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Roles</td>
<td>Cohens Ranks</td>
<td>Roles</td>
</tr>
<tr>
<td></td>
<td>S1-O</td>
<td>.61 4.5</td>
<td>.88 22</td>
</tr>
<tr>
<td></td>
<td>S2-O</td>
<td>.61 4.5</td>
<td>.72 8</td>
</tr>
<tr>
<td></td>
<td>S1-T</td>
<td>.75 10</td>
<td>.79 12</td>
</tr>
<tr>
<td></td>
<td>S2-T</td>
<td>.84 17</td>
<td>.63 7</td>
</tr>
<tr>
<td></td>
<td>S1-S2</td>
<td>.83 16</td>
<td>.80 13.5</td>
</tr>
<tr>
<td></td>
<td>T-O</td>
<td>.62 6</td>
<td>.85 19</td>
</tr>
</tbody>
</table>

Rank Sums 58 81.5 39.5 121
H = 12.39
p ≤ .0062

Rank Sums 78.5 52 55.5 114
H = 8.20
p ≤ .0420
TABLE 9 -- Continued
KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE FOR EACH OF THE FOUR MICROTEACHING GROUPS

<table>
<thead>
<tr>
<th>Group III</th>
<th>Microteaching Lessons</th>
<th>Group IV</th>
<th>Microteaching Lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roles</td>
<td>Cohens Rank</td>
<td>Roles</td>
<td>Cohens Rank</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------</td>
<td>----------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>$S_1-O$</td>
<td>1</td>
<td>$S_1-O$</td>
<td>1</td>
</tr>
<tr>
<td>$S_2-O$</td>
<td>10</td>
<td>$S_2-O$</td>
<td>15.5</td>
</tr>
<tr>
<td>$S_1-T$</td>
<td>8.5</td>
<td>$S_1-T$</td>
<td>20.5</td>
</tr>
<tr>
<td>$S_2-T$</td>
<td>21</td>
<td>$S_2-T$</td>
<td>.85 23</td>
</tr>
<tr>
<td>$S_1-S_2$</td>
<td>5.5</td>
<td>$S_1-S_2$</td>
<td>.87 24</td>
</tr>
<tr>
<td>$T-O$</td>
<td>12.5</td>
<td>$T-O$</td>
<td>.69 3</td>
</tr>
<tr>
<td>Rank Sums</td>
<td>121</td>
<td>Rank Sums</td>
<td>121</td>
</tr>
</tbody>
</table>

$H = 13.09$
$p ≤ .0044$

$H = 3.31$ N.S.
$H ≥ 6.25$ required for $p ≤ .10$
Significant Kruskal-Wallis H values were found for three of the four peer-microteaching groups. Only Group IV failed to yield a significant H value. Group I rendered an H value significant at the .006 level, Group II rendered an H value significant at the .042 level and Group III rendered an H value significant at the .004 level of confidence.

As each of the H values for Groups I, II and III had a p value greater than .10, the minimum level of confidence previously stated as being acceptable in this study, the fourth null hypothesis is rejected. In three of the four microteaching groups, the same four subjects differed significantly with respect to their inter-observer agreement when coding four different peer-microteaching lessons.

The fifth hypothesis stated: Subjects serving in the role of teacher, student, or observer will not differ from the researcher in their classification of selected verbal behaviors used in peer-microteaching lessons. This hypothesis did not use total tally data coded at three second intervals. Instead, eighteen verbal behaviors selected by the researcher from each microteaching lesson were used. Each subject tallied each of the eighteen selected verbal behaviors during a playback of an audio-tape immediately following each peer-microteaching lesson. The subjects' tallies were compared to the tallies of the researcher who served as the standard. The eighteen recorded verbal behaviors were then summed according to correctness/incorrectness (based on researcher as the standard) by category and the role of the subject.
Table 10 presents the correct and incorrect tallies for each category according to role, the chi-square value for each category and the total tallies and chi-square value for the total sample of 1,152 recorded tallies. None of the chi-square values were significant at the .10 level of confidence. Therefore, the fifth hypothesis cannot be rejected.

The researcher used the chi-square test to further analyze the selected recorded verbal behaviors. Table 11 presents this data. Comparing the recorded behaviors in the same manner as in Table 10, the researcher found no significant differences at the .10 level of confidence for any of the groups of categories examined.

Categories 1 + 2 + 3 + 4 divided by categories 5 + 6 + 7 renders a ratio designated by Flanders as the I/D or indirect/direct teacher talk ratio. Categories 1 + 2 + 3 divided by 6 + 7 renders the revised I/D or revised indirect/direct teacher talk ratio. (15;102) Categories 8 and 9 are student talk categories. Even with this additional analysis of the data, no chi-square values approached the level of significance required by this study. There was no significant difference in the classification of selected verbal behaviors used in peer-microteaching lessons as recorded by subjects serving in the role of teacher, student or observer and the researcher.

The sixth research question helped to guide comparisons between the Broadwater study into the representative validity of the Observational System for Instructional Analysis (61) and the present investigation into the representative validity of the Flanders' System
### TABLE 10
CHI-SQUARE TEST FOR EACH CATEGORY ON THE EIGHTEEN SELECTED VERBAL BEHAVIORS IN EACH MICROTEACHING LESSON

<table>
<thead>
<tr>
<th>Categories</th>
<th>Tallies by Role</th>
<th>T</th>
<th>S&lt;sub&gt;1&lt;/sub&gt;</th>
<th>S&lt;sub&gt;2&lt;/sub&gt;</th>
<th>O</th>
<th>x&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C*</td>
<td>24</td>
<td>22</td>
<td>24</td>
<td>25</td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>I*</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>33</td>
<td>35</td>
<td>32</td>
<td>33</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>28</td>
<td>31</td>
<td>27</td>
<td>24</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>10</td>
<td>7</td>
<td>11</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>28</td>
<td>28</td>
<td>29</td>
<td>28</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>29</td>
<td>28</td>
<td>32</td>
<td>33</td>
<td>3.65</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>26</td>
<td>23</td>
<td>22</td>
<td>24</td>
<td>4.26</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>24</td>
<td>28</td>
<td>26</td>
<td>24</td>
<td>4.83</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>26</td>
<td>28</td>
<td>28</td>
<td>24</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>32</td>
<td>28</td>
<td>31</td>
<td>32</td>
<td>3.66</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Total Tallies</td>
<td>C</td>
<td>250</td>
<td>251</td>
<td>251</td>
<td>247</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>38</td>
<td>37</td>
<td>37</td>
<td>41</td>
<td></td>
</tr>
</tbody>
</table>

*Where C = Correct tally; I = Incorrect tally; T = Teacher; S<sub>1</sub> = Student #1; S<sub>2</sub> = Student #2; O = Observer

x<sup>2</sup> ≥ 6.25 required for p ≤ .10 with three degrees of freedom.
TABLE 11

CHI-SQUARE TEST ON THE EIGHTEEN SELECTED VERBAL BEHAVIORS FOR SIXTEEN PEER-MICROTEACHING LESSONS BY ASSOCIATED CATEGORY GROUPINGS

<table>
<thead>
<tr>
<th>Category Groupings</th>
<th>T</th>
<th>Tallies by Role</th>
<th></th>
<th></th>
<th>(X^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,3 and 4</td>
<td>C*</td>
<td>113 116 112 110</td>
<td></td>
<td></td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>18 15 19 21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2 and 3</td>
<td>C</td>
<td>85 88 83 82</td>
<td></td>
<td></td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>17 14 19 20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5,6 and 7</td>
<td>C</td>
<td>79 79 80 81</td>
<td></td>
<td></td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>11 11 10 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 and 7</td>
<td>C</td>
<td>50 51 48 48</td>
<td></td>
<td></td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>4 3 6 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 and 9</td>
<td>C</td>
<td>58 56 59 56</td>
<td></td>
<td></td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>9 11 8 11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Where C = Correct tally, I = Incorrect tally

\(X^2 \geq 6.25\) required for \(p \leq .10\) with three degrees of freedom.

of Interaction Analysis. As previously stated, the present study was carefully designed and implemented to replicate the research conducted by Broadwater. Figure 3 presents the five hypotheses investigated in both studies and whether the hypotheses were accepted or rejected. Immediately following Figure 3 is an explanation of those hypotheses rejected by the Broadwater study.
FIGURE 3

Comparison Of Findings Between
The Sudduth Study (FSIA) And The Broadwater Study (OSIA)

<table>
<thead>
<tr>
<th>Sudduth Study (FSIA)</th>
<th>HYPOTHESES</th>
<th>Broadwater Study (OSIA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Rejected</td>
<td>HYPOTHESIS 1</td>
<td>Not Rejected</td>
</tr>
<tr>
<td>Not Rejected</td>
<td>HYPOTHESIS 2</td>
<td>Rejected(^1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Significant for only one</td>
</tr>
<tr>
<td></td>
<td></td>
<td>paired set of subjects out</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of 15 possible paired sets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of subjects)</td>
</tr>
<tr>
<td>Not Rejected</td>
<td>HYPOTHESIS 3</td>
<td>Rejected(^2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p equal to but not less</td>
</tr>
<tr>
<td></td>
<td></td>
<td>than .10)</td>
</tr>
<tr>
<td>Rejected(^3)</td>
<td>HYPOTHESIS 4</td>
<td>Rejected(^3)</td>
</tr>
<tr>
<td>(Group 1: (p \leq .01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2: (p \leq .05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3: (p \leq .01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4: N.S.)</td>
<td></td>
<td>(Only Group 4 significant:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p \leq .01))</td>
</tr>
<tr>
<td>Not Rejected</td>
<td>HYPOTHESIS 5</td>
<td>Rejected(^4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Significant for only one</td>
</tr>
<tr>
<td></td>
<td></td>
<td>category out of nine)</td>
</tr>
</tbody>
</table>

\(^1\)The Broadwater study rejected Hypothesis 2 on the finding of
a Mann-Whitney U value of 60 which was significant beyond the .10 level
of confidence. This significant U value was obtained in comparing the
Cohen coefficients of the teacher and observer with the Cohen coefficients
of the sets of two students. Therefore, in the Broadwater study, the
inter-observer agreement of the classification of verbal behaviors
between the teacher-observer pair and the student-student pair was
found to differ significantly. (61;65-66)
The Broadwater study rejected Hypothesis 3 on the finding of a Mann-Whitney U value of 607 which was significant at the .10 level of confidence. This significant U value was obtained in comparing the Cohen coefficients of the thirty-two teacher-student pairings with the thirty-two observer-student pairings. Therefore, in the Broadwater study, the inter-observer agreement of the classification of verbal behaviors between the teacher-students and observer-students was found to differ significantly. (61;66-68)

The Broadwater study also rejected Hypothesis 4 when the fourth peer-microteaching group obtained a Kruskal-Wallis H value significant at the .10 level. This was the only peer-microteaching group out of four groups to render significant inter-observer lack of agreement. The H value obtained was significant beyond the .01 level of confidence. Therefore, in the Broadwater study, the same four subjects within the fourth peer-microteaching group differed significantly in their recording of verbal behaviors occurring in four different peer-microteaching lessons. (61;68-70)

The Broadwater study rejected Hypothesis 5 on the finding of a chi-square value of 16.78 which was significant beyond the .01 level of confidence. This significant chi-square value was obtained in comparing the tallies of ten selected verbal behaviors of the subjects by role to the tallies of the same verbal behaviors by the researcher who served as the standard. The significant chi-square value was obtained in Category 6, which in the Observation System for Instructional Analysis is the category of Confirmation. This
was the only category (out of nine categories) for which a significant chi-square value was obtained. Therefore, in the Broadwater study, the subjects, in recording Category 6 behavior as compared to the researcher, differed significantly. (61;71-72)

Table 12 presents the chi-square values for the selected verbal behaviors obtained in the Broadwater study and the present study on the basis of the comparison of the category similarities presented in Table 1, Chapter III. Figure 2, Chapter III presented a category description of the modified version of the Observational System for Instructional Analysis used by Broadwater. The Broadwater study used ten selected verbal behaviors while the present study used eighteen selected verbal behaviors. (Categories for silence and confusion were not included in testing Hypothesis 5 in either study.)

Implications drawn from a comparison of the findings of the two studies will be discussed in Chapter V. The similarities of the two observational systems and the replication of Broadwater's original research design in the present study have facilitated this comparison between Broadwater's investigation of the representative validity of the Observational System for Instructional Analysis and the present investigation into the representative validity of the Flanders' System of Interaction Analysis.
TABLE 12

CHI-SQUARE VALUES FOR SELECTED VERBAL BEHAVIORS FROM THE SUDDUTH STUDY (FSIA) AND THE BROADWATER STUDY (OSIA)

<table>
<thead>
<tr>
<th>Sudduth Study</th>
<th>Broadwater Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSIA Categories</td>
<td>Chi-Square Values**</td>
</tr>
<tr>
<td>1</td>
<td>4.00</td>
</tr>
<tr>
<td>2</td>
<td>.97</td>
</tr>
<tr>
<td>3</td>
<td>3.29</td>
</tr>
<tr>
<td>4</td>
<td>1.03</td>
</tr>
<tr>
<td>5</td>
<td>3.65</td>
</tr>
<tr>
<td>6</td>
<td>4.26</td>
</tr>
<tr>
<td>7</td>
<td>4.83</td>
</tr>
<tr>
<td>8</td>
<td>2.11</td>
</tr>
<tr>
<td>9</td>
<td>3.66</td>
</tr>
</tbody>
</table>

*Significant beyond .01 level of confidence
**Chi-square value ≥ 6.25 required for p ≤ .10 with three degrees of freedom.
Summary

This chapter has presented the research findings and the analysis of data with respect to the stated null hypotheses used to guide this investigation of the representative validity of the Flanders' System of Interaction Analysis. The four statistical tests used in the analysis of the data were Cohen coefficients, Kruskal-Wallis one-way analysis of variance, Mann-Whitney U test and the chi-square test. Information and data used to draw comparisons between the Broadwater study and the present study in order to answer the final research question were also presented.

The first hypothesis was not rejected on the basis of statistical measures involving a comparison of Cohen coefficients for each pair of subjects as analyzed by the Kruskal-Wallis one-way analysis of variance. The findings indicated no significant difference in the inter-observer agreement of the classification of verbal behaviors for sixteen peer-microteaching lessons in which all subjects had either participated or observed. No pair of subjects demonstrated measurable differences in recording verbal behaviors from any other pair of subjects.

The second hypothesis was not rejected on the basis of the application of the Mann-Whitney U test in analyzing the Cohen coefficients for comparing all pairs of subjects. There was no significant difference in the inter-observer agreement of the classification of verbal behaviors of any two pairs of subjects.

The third hypothesis tested the inter-observer agreement in the classification of verbal behaviors between the teacher and students
and the observer and students. The Cohen coefficients for the two subjects serving in the roles of students were compared with the teacher and observer. The Mann-Whitney U test indicated no significant differences and therefore the third hypothesis was not rejected. There was no significant difference in the verbal classification of the teacher-students' tallies and the observer-students' tallies.

The fourth hypothesis was rejected on the basis of significant Kruskal-Wallis H values in three of the four peer-microteaching groups. The same four subjects did differ significantly in their classification of verbal behaviors occurring in four different peer-microteaching lessons in which each of the subjects had either been an active participant or served as an observer. Therefore, the same persons can differ significantly in the classification of verbal behaviors over a series of four peer-microteaching lessons even though such differences may not be associated with the role (teacher, student or observer) played by the person.

The chi-square values obtained in testing the fifth hypothesis were not significant and therefore, the hypothesis could not be rejected. There was no significant difference in the classification of selected verbal behaviors used in peer-microteaching lessons between subjects serving in the role of teacher, student or observer and the researcher. Even with additional analysis of data no significant results were observed.

Research findings in the Broadwater study enabled that investigator to reject four of the five research hypotheses used in
that study. As these same hypotheses were investigated in the present study, findings from the Broadwater study were presented in this chapter. Both studies involved research into the representative validity of two similar observational instruments. Implications resulting from a comparison of the findings of the two studies will be presented in the following chapter.
CHAPTER V
SUMMARY AND CONCLUSIONS

Introduction

This study investigated the representative validity of the Flanders' System of Interaction Analysis. In establishing an estimate of the representative validity of this system, observational data was obtained from observers trained in the use of Flanders' System and from participants, trained in the use of the Flanders' System, serving in the roles of teacher and students in instructional situations.

Observational data obtained by observers using various observational systems has been used to provide information for educational research, evaluation of the learning process, teacher training programs and to monitor specific educational programs. Despite an obvious need to have valid observational data, limited research is available to provide information on how participants (namely teacher and students) would record those same classroom behaviors recorded by an observer. In the past, the ability of the observer to provide valid observational data has been assumed because of the observer's previous training in the use of an observational system and through the process of obtaining high inter-observer agreement between two or more trained observers, or between observers and a standard. This study, by having
the participants as well as an observer of an instructional situation record the verbal behaviors of that same instructional situation, has provided a more direct measure of the representative validity of observational data.

The observational system used for an investigation of representative validity in this study was the basic ten category system developed by Flanders. (15) Through an investigation of this system, the representative validity of the specific categories within the system was also investigated. To establish an estimate of the representative validity of the total system as well as the system's specific categories required obtaining recorded observational data from an observer and the teachers and students who served as active participants in the instructional situations. Through concurrent recording of observational data by subjects serving in the roles of teacher, students and observers an estimate of the representative validity of the Flanders' System of Interaction Analysis was obtained.

Subjects were first trained by the researcher in the basic constructs and procedural operations of Flanders' System. The subjects were then trained by the researcher to use Flanders' System of Interaction Analysis to record classroom verbal behavior. All subjects obtained a Scott coefficient of inter-observer agreement of .80 or better when compared with the researcher. The subjects were then assigned to four different groups with four subjects in each group. These subjects then participated in peer-microteaching lessons in which each subject served in the role of teacher and observer once, and student twice.
During each peer-microteaching lesson, the subject serving in the role of teacher taught two subjects serving in the role of student while a subject serving in the role of observer observed the instructional situation from a short distance away. Each peer-microteaching lesson was recorded on an audio tape recorder. Upon conclusion of each peer-microteaching lesson, the audio tape was replayed and the four subjects recorded the verbal behaviors of the peer-microteaching lesson using Flanders' System of Interaction Analysis. Immediately following the initial audio tape playback, the tape was re-played one additional time. During the second playback, the researcher stopped the tape at eighteen selected points and the four subjects recorded the verbal behavior occurring at the stop. This same procedure was used for each of the four lessons taught by subjects in each of the four peer-microteaching groups.

Cohen coefficients were calculated to measure inter-observer agreement, and a frequency distribution for each of the ten verbal categories was produced for the sixteen peer-microteaching lessons. Nonparametric statistical tests were used to analyze the data obtained from the recorded observations of the subjects who participated in this study. The .10 level of significance was used in testing the null hypotheses. All tests were two-tailed tests.

The findings indicate:

1. Persons serving in the role of teacher, student or observer in peer-microteaching situations were in general agreement in the classification of verbal
behaviors occurring in peer-microteaching situations when using the Flanders' System of Interaction Analysis to record verbal behavior.

2. There was no significant difference in the inter-observer agreement between any pairs of subjects when compared by peer-microteaching roles for sixteen peer-microteaching lessons.

3. Teacher-student Cohen coefficients were found to be higher than observer-student Cohen coefficients but the difference of inter-observer agreement between the two groups was not significant.

4. Subjects in three of the four peer-microteaching groups differed significantly with respect to their inter-observer agreement when coding four different peer-microteaching lessons in which the subjects had either participated as teacher or student or served as observer.

5. No significant differences occurred between the tallies of the researcher and subjects according to peer-microteaching roles or role of observer for the eighteen selected verbal behaviors.

6. General agreement exists between the findings of the present study and the Broadwater study confirming a tendency of observer and teacher and students to agree on the classification of verbal behaviors when
using the Flanders' System of Interaction Analysis and the Observational System for Instructional Analysis to record behaviors occurring in a peer-microteaching situation.

**Discussion of the Findings**

The observational data used in this investigation of the representative validity of the Flanders' System of Interaction Analysis was obtained by having subjects act as teacher, student or observer in a peer-microteaching situation. A basic assumption upon which data collection rested was that peer-microteaching would provide a realistic instructional situation and that verbal behaviors recorded during peer-microteaching lessons would be representative of behaviors occurring in a normal classroom instructional situation. From the findings of the Ober study (67) and the work of Allen at Stanford University (1) as well as other reports of the use of microteaching in teacher-education programs, (1;xiii) the use of peer-microteaching in the present study would seem to provide a fairly authentic teaching-learning situation in which to examine the hypotheses being investigated in this study. Comments by the subjects following completion of the peer-microteaching lessons indicated they perceived the microteaching experience as being highly representative of an actual classroom instructional situation.

Another critical element in this study of the representative validity of the Flanders' System of Interaction Analysis was the time involved between the actual teaching of the lesson and the
coding of the audio tape recording of the verbal behaviors occurring
during the lesson. It was assumed that the subjects could, in
listening to an immediate replay of an audio tape recording of the
lesson in which they participated, cognitively and emotionally
reassume the roles they portrayed during the peer-microteaching
lesson. Therefore, in order to minimize changes in the attitudes
of the subjects and to reduce the possibility of a subject's for­
getting, the peer-microteaching lessons were tallied by the subjects
immediately upon conclusion of the lesson.

In the review of literature pertaining to perceptual variables
which could influence a person's objectivity in recording the behaviors
of others, Taft (57), Norman (52) and Shrauger and Altrocchi (55)
discuss numerous factors which tend to condition an observer's
perception of others. Many of those factors such as emotional
stability, self-insight and motivation to judge accurately were
not controlled in this study. However, the fact that all the sub­
jects attended the same small, liberal arts college in the Midwest,
were advanced students enrolled in a teacher preparation program
leading to teacher certification, and were enrolled in the same
education class with the same instructor, would tend to suggest
that subjects did possess certain common characteristics.

The first null hypothesis that:

There will be no difference in the inter-observer
agreement of the classification of verbal behaviors
of any of the six possible pairs of subjects,
could not be rejected. This finding indicates that the sixteen
subjects who each assumed the role of teacher and observer once, and
the role of student twice, did not differ measureably in their
perception of peer-microteaching verbal behaviors for the sixteen
peer-microteaching lessons. Thus, when using the Flanders' System
of Interaction Analysis to record verbal behaviors occurring in a
peer-microteaching situation, sixteen subjects were in general
agreement as to the classification of verbal behaviors, when compared
according to the three roles they portrayed in the peer-microteaching
situations.

The review of research on perception variables emphasized the
uniqueness of individual perception. Therefore, the general agree­
ment of the subjects in classifying verbal behavior while serving
in varying roles suggests that persons trained in a particular
classroom observational system may be able to overcome individual
perceptual differences as influenced by roles played in instructional
settings. Through the use of a common set of guidelines used to
describe a set of behaviors which may exist in a learning situation,
persons acting in particular roles can arrive independently at a
general agreement as to the classification of those behaviors.

The second null hypothesis that:

There will be no difference in the inter-observer
agreement of the classification of verbal behaviors
of any two pairs of subjects,

was not rejected. An examination of fifteen possible combinations
of pairs resulted in the conclusion that there were no significant
inter-observer differences in recorded behaviors among any of the
pairs of subjects regardless of the role portrayed.
The composite student-student coefficients were the second highest role coefficients reported in Table 6. The Cohen coefficients for pairs of active participants in the learning situation tended to be higher than the Cohen coefficients for the observer and one of the active participants. However, none of the differences were significant. Research by Peak (27) and Shrauger and Altrocchi (55) suggest that active involvement of the participants may, through the nature of the interaction, account for the higher inter-observer agreement among the active participants. Taft's position (57) that motivation is the most important attribute in the ability to judge others may also support the higher inter-observer agreement among the active participants as they were the ones whose behaviors were being recorded. Gage's view (48) of the role of motivation and its influence on perception might also confirm this position. Thus, they were judging themselves and in so doing were providing a measure of the representative validity of the Flanders' System of Interaction Analysis.

The third null hypothesis that:

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

was not rejected. The teacher-students Cohen coefficients were higher than the observer-students Cohen coefficients but the difference was not significant. This relationship is contrary to the
findings of Morse, Bloom and Dunn (66) in which observer student agreement occurred more frequently than teacher-student or teacher-observer agreement. The Morse, Bloom and Dunn study was conducted in an actual classroom instructional situation while the present study was conducted using peer-microteaching as the instructional setting. As the subjects knew one another on a fairly personal basis, it is possible they may have been more in harmony of purpose when acting in their roles of teacher and students and therefore in the roles of active participants they were more involved because of personal motivation. This personal motivation then may have influenced the teacher and students' recording of the verbal behaviors resulting in closer agreement than with the less personally involved observer. Bruner and Tagiuri (11) indicate that the judge who is more relaxed and less emotional tends to be a better judge. Perhaps the personal involvement of the subject serving in the role of teacher or student caused him to be more highly motivated in recording his own behavior and negated the influence of the emotional state of being one of the active participants in the peer-microteaching lesson. This would seem to agree with the position of Taft (57) and Gage (48) with respect to the importance of motivation on the perceptual viewpoint of the observer.

The fourth null hypothesis that:

There will be no differences in the inter-observer agreement of the classification of verbal behaviors between subjects for each of the four peer-microteaching lessons within each of the four peer-microteaching groups,
was rejected. This hypothesis was rejected on the basis of significant differences in inter-observer agreement occurring in three of the four peer-microteaching groups. This finding indicates that there was a significant difference in the inter-observer agreement among the same four subjects recording four peer-microteaching lessons in three of the four peer-microteaching groups. This finding therefore introduces a possible challenge to the representative validity of the Flanders' System of Interaction Analysis. The observational data indicates that the representative validity can differ between peer-microteaching lessons with the same group of subjects.

The three previous hypotheses supported a fairly constant measure of representative validity when examining the inter-observer agreement of subjects according to role. The fourth hypothesis examines the inter-observer agreement of subjects without regard to role. In this situation persons trained in the use of Flanders' System of Interaction Analysis failed to agree on the classification of verbal behaviors between four lessons. This disagreement indicates a challenge to the representative validity of the System.

Another possible explanation of the significant inter-observer lack of agreement within the three peer-microteaching groups is the degree of individual coding reliability. While each subject did achieve a .80 level of inter-observer reliability with the researcher during three recording checks before the peer-microteaching lessons and on one final recording check after the
lessons, this does not insure that some factor, such as forgetting, might not have interfered with the tallying of the individual subject.

This finding may indicate perceptions of the subjects as individuals influenced their coding reliability to a more significant degree than those perceptual variables related to role. Roles changed for each peer-microteaching lesson, but the four subjects involved in the peer-microteaching lessons remained constant.

Thus, the perceptual variables presented by Taft (57), Norman (52), Shrauger and Altrocchi (55) and Gage (48) take on new significance in relation to the rejection of the fourth hypothesis. The factors which can interfere with an observer's objectivity as presented by Biehler (10) and Hyman (50) must again be considered in terms of the effects of what the individual brings to a given situation from past experience. Individual perceptual variables may be constant with each subject even though that individual changes role. Thus, when a subject assumes a given role, the perceptual variables peculiar to that individual may be reflected in his recorded observations because they are a part of the individual and not because of the role he is portraying.

Another possible cause of the significant inter-observer disagreement which called for the rejection of the fourth hypothesis is the variation in the type of lesson that was taught. This is a factor that was not controlled in this study as each subject selected both the content and the method of presentation he would use in assuming the role of teacher. This freedom of choice allowed each
subject did result in a wide range of teaching material. However, the requirement for each subject serving in the role of teacher to include at least two examples of each category of teacher verbal behavior tended to have a general conforming influence on the method of presentation.

The fifth null hypothesis that:

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

was not rejected. The testing of the fifth hypothesis involved the tallying of eighteen selected verbal behaviors by the subjects and then comparing the subjects' tallies with the researcher's tallies. There was general agreement between the tallies of the subjects and the tallies of the researcher for the eighteen selected verbal behaviors. This general agreement between the subjects and the researcher is an indication of both representative and construct validity. This hypothesis compared the tallies of the subjects according to role with the tallies of the researcher. In the preceding hypothesis which compared inter-observer agreement according to role, there was also a positive measure of the representative validity of the observational data when analyzed by role.

The general agreement between the subjects' tallies as compared to the researcher's tallies was sustained when additional examinations of various category groupings were made. The studies
of Good and Brophy (22) and Premac (53) indicating that reinforcement is an individualized matter, help support the positive findings of the representative validity in that the subjects tended to agree on the classification of all verbal behaviors including those verbal categories used to indicate appraisal behaviors. It should be noted however, that appraisal behaviors accounted for a mean of only fifteen per cent of the total verbal behaviors.

Thus, in four of the five research hypotheses examined in the present study, data indicated no significant findings. There are a number of alternative explanations for research findings of "no significance." First, the finding could have been produced by imprecise measurement. Second, the finding could have been produced by some unmeasured or random variable not studied. Third, there could have been little or no difference in the classification of verbal behaviors by the subjects according to role portrayed in the peer-microteaching lessons. This latter case seems to be the most plausible in the present study. Therefore, the findings would seem to support the use of the Flanders' System of Interaction Analysis as a valid instrument to use in the gathering of classroom observational data involving teacher-student verbal behaviors. The fact that the Flanders' System has been used to correlate teacher verbal behavior with student attitudes and student achievement (20) and the support of these correlations by additional research (28) would indicate the value of the use of this instrument to gather valid classroom observational data. The Carline study (62), which failed to confirm a correlation between teacher indirect verbal
behavior and student arithmetic achievement, indicates a need for continued research in the area of teacher verbal strategies. However, research tends to support the use of the Flanders' System of Interaction Analysis as a valid instrument for recording classroom verbal behavior. The findings of the present study would seem to confirm the representative validity of the Flanders' System.

The sixth research question involved a comparison of the present investigation regarding the representative validity of the Flanders' System of Interaction Analysis with the findings of the Broadwater study (61) into the representative validity of a modified version of the Observational System for Instructional Analysis. Figure 3 presents a comparison of the findings for the five hypotheses examined by both studies.

Neither study rejected the first hypothesis dealing with the difference in the inter-observer agreement of the classification of verbal behaviors of any of the six possible pairs of subjects. This finding gives initial confirmation of the representative validity of the two observation systems. In both studies, there was general inter-observer agreement among all subjects.

The second hypothesis pertaining to the inter-observer agreement of the classification of verbal behaviors of any two pairs of subjects was rejected in the Broadwater study in which the inter-observer agreement between teacher-student and observer-student was found to differ significantly. These were the only pairs of subjects to differ significantly out of fifteen possible pairs of combinations.
This same hypothesis was not rejected in the present study. Broadwater admitted that the one case of significant lack of inter-observer agreement may have occurred by chance. In the present study, the same pairs of subjects were among the higher pairs in terms of inter-observer agreement.

The third hypothesis investigating the inter-observer agreement of the classification of verbal behaviors of teacher and students when compared with the inter-observer agreement of the observer and students was rejected in the Broadwater study and not in the present study. The Broadwater study found higher inter-observer agreement between observer-students than between teacher-students. The reverse was the case in the present study. It should also be noted that the $p$ value upon which Broadwater based his rejection of the third hypothesis was at the minimum level Broadwater set as necessary for the rejection of the hypothesis. Despite support in the literature (Morse, Bloom and Dunn) (66) toward a tendency for higher inter-observer agreement to occur between observer-student rather than teacher-student, lack of support in the present study brings this position into question.

The fourth hypothesis pertaining to the inter-observer agreement between subjects for each of the four peer-microteaching lessons within each of the four peer-microteaching groups was rejected by the Broadwater study on the basis of significant differences of inter-observer agreement in one of the four peer-microteaching groups. This same hypothesis was rejected by the present study when three of the four peer-microteaching groups were found to have significant
differences of inter-observer agreement for each of the four peer-microteaching lessons. Broadwater indicated the one peer-microteaching group which produced observational data supporting rejection of the fourth hypothesis was the group "in which most of the negative personal judgment behaviors were reported." (61:84) The fourth hypothesis was the only hypothesis rejected by both studies. Thus, the findings of both studies may raise questions as to the reliability of observers rather than the representative validity of the system each study was designed to investigate.

The fifth hypothesis pertaining to the classification of selected verbal behaviors as tallied by the subjects and compared to the tallies of the respective researchers was rejected by the Broadwater study in that the category of confirmation (Category 6, Observational System for Instructional Analysis) was tallied by the subjects in significant disagreement with the researcher. This was the only category out of nine categories for which Broadwater found significant differences between the subjects' tallies of selected verbal behaviors and his own tallies. The present study failed to reject the fifth hypothesis.

Table 12 presents the chi square values for the selected verbal behaviors for both the Broadwater study and the present study. The comparable Flanders' System category (Category 2, Praises or Encourages) for which Broadwater found a significant difference in the Observational System for Instructional Analysis obtained the least significant chi square value in the present study. However, other than Category 2, those categories associated with appraisal behaviors tended to have
high chi square values in the present study, but not at the required level of significance.

This calls to review the research of Good and Brophy (22) indicating teachers reinforce students as individuals rather than as a total class. Premac (53) also demonstrated how the same conditions may assume differing reinforcing roles for different individuals. Therefore, when appraisal behaviors are involved, individual perceptual differences as suggested by Taft (57) and others may have been operating rather than, or in addition to, role influence.

Two general observations can be drawn from the findings of both studies. Data in both studies support the contention that while there were significant differences observed in both studies the findings would seem to support a greater tendency toward inter-observer agreement than disagreement. This would indicate a positive measure of the representative validity of both the Flanders' System of Interaction Analysis and the Observational System for Instructional Analysis.

The second general observation is that most significant findings occurred in those areas in which individual perceptual differences may have a greater influence on the observational judgments of system users than conditions dictated by the role of the individual. This position is supported by the fact that the fourth hypothesis was the only hypothesis rejected by both studies. The fourth hypothesis compared inter-observer agreement of subjects without regard to role. The third hypothesis resulted in opposite findings in the two studies.
The Broadwater study rejected the third hypothesis and found greater inter-observer agreement between observer-students than between teacher-students. The present study, while failing to reject the hypothesis, found greater inter-observer agreement between teacher-students than between observer-students. The second hypothesis, rejected in the Broadwater study, had similar results in that the role pairings rendering significant results in the Broadwater study were among the role pairings having the highest inter-observer agreement in the present study. Finally, Broadwater states in discussing the rationale for rejecting the fifth hypothesis the position that, "individual differences in perception . . . seem to be operating here rather than the role influence observed in hypotheses two and three." (61;87)

Implications and Recommendations for Further Study

The results of research into the representative validity of the Flanders' System of Interaction Analysis based upon the hypotheses investigated in this study would seem to imply that a high degree of inter-observer agreement exists among subjects serving in the roles of teacher, student or observer in the recording of peer-microteaching lessons. The present study served to extend the original Broadwater investigation into the representative validity of the Observational System for Instructional Analysis to the Flanders' System of Interaction Analysis. Despite differences in the findings of the two studies, data from the two investigations tend to be supportive.
Results of both studies indicate a need for additional research into the question of representative validity of observational systems. The present study would seem to indicate a major need to investigate the role that lesson content plays in creating different learning conditions in the classroom situation. This need is seen as a result of the rejection of the hypothesis concerning inter-observer agreement for different peer-microteaching lessons.

This same finding indicates additional study is needed in the area of perceptual variables as they relate to those decision making processes required in using category discriminations common to classroom observational systems. Further study is needed also to determine the influences of role needs and limitations on the objective recording of instructional situations.

In addition to perceptual variables and role dimensions, the effects of appraisal behavior needs additional research. As Good and Brophy (22) suggest, perhaps this might be done by using an observational system which can focus on the relationship of the teacher and individual students within the instructional setting. By reducing some of the global aspects of observational systems, researchers may be able to identify not only problems associated with appraisal behaviors but other factors which can contribute to faulty encoding of instructional behaviors. Attention toward further refinement of appraisal behavior categories would seem to be warranted in view of the differences associated with recording this type of instructional behavior.
It would seem appropriate to continue investigation of the use of microteaching as a source for obtaining representative samples of classroom behavior. Certain aspects of peer-microteaching need to be studied in order to ascertain if peer-microteaching changes the simulated advantages usually attributed to the microteaching situation. The problems associated with self evaluation also need examination as this condition exists in using peer-microteaching to collect data on the representative validity of observational systems.

**Summary**

This exploratory investigation into the representative validity of the Flanders' System of Interaction Analysis has demonstrated that subjects serving in the role of teacher, student, and observer tend to agree on the general classification of verbal behaviors recorded while participating in peer-microteaching lessons. Because of this general inter-observer agreement, the observational data collected during peer-microteaching lessons is considered to have representative validity. This conclusion must be tentative until further investigation resolves the question of what causes inter-observer agreement to differ significantly when the same subjects record different peer-microteaching lessons. The possibility of either representative validity of the Flanders' System or recorder reliability in the use of Flanders' System or both affecting inter-observer agreement in the classification of verbal behaviors should receive further investigation.
Studies of the representative validity of observational systems should continue in order to strengthen the value of this tool for educational research and instructional analysis. Research into the representative validity of observational systems needs to expand into actual classroom situations in order that findings from exploratory investigations such as the present study may be more fully tested. It is the hope of this researcher that this investigation may have contributed in some positive manner toward the improvement of the study of learning situations.
BOOKS


17. ________. Project Director, *Helping Teachers Change Their Behavior,* University of Michigan, 1963.


BOOKS — Continued


BOOKS -- Continued


ARTICLES IN JOURNALS


46. _______. "Personal-Social Anxiety as a Factor in Experimental Learning Situations." Journal of Educational Research, 45, (October, 1951), 100-110.
ARTICLES IN JOURNALS -- Continued

47. ______. "Using Interaction Analysis in the In-Service Training of Teachers." Journal of Experimental Education. XXX, No. 4 (June, 1962), 313-316.


UNPUBLISHED MATERIALS


