DOCUMENT RESUME

ED 101 002		TM 004 127
AUTHOR	Casteel, J. Doyle; Stahl, Robert J.	
TITLE	The Social Science Observation Record: Theoretical Construct and Pilot Studies. Research Monograph No. 7.	
INSTITUTION	Florida Univ., Gainesville. P. K. Yonge Lab. School.	
REPORT NO	RM-7	
PUB DATE	73	
NOTE	134p.; For related documents, see TM 004 125-126	
AVAILABLE FROM	P.K. Yonge Laboratory School, University of Florida, Gainesville, Florida 32611 (free)	
EDRS PRICE	MF-$0.75 HC Not Available from EDRS. PLUS POSTAGE	
DESCRIPTORS	*Classroom Observation Techniques; Inservice Education; *Interaction Process Analysis; Matrices; Microteaching; *Models; Preservice Education; *Records (Forms); Reliability; *Social Sciences; Student Behavior; Teacher Behavior; Values	
IDENTIFIERS	*Social Science Observation Record; SSOR	

ABSTRACT

The Social Science Observation Record (SSOR) is a systematic observation system designed to abstract and describe class discussion, specifically those discussions directed toward value clarification as an aspect of subject-matter instruction. The objective and yet diverse nature o category observation systems is described and an explanation is synthesized of how pre-service and in-service teachers may use category observational systems to study their behavior as teachers. A model of class discussion relevant to student understanding and to value clarification is presented and specific categories are identified and described of student and teacher behaviors through which student understanding and value clarification behaviors can be planned for and employed during class discussion. The knowledge and skills necessary for teachers to organize and interpret the data are discussed. Empirical studies are reported with regard to the modification of teacher and student behaviors, and methods by which observers have been trained to collect data with reliability using the SSOR are described. (Author/RC)

The SOCIAL SCIENCE OBSERVATION RECORD

Theoretical Construct and Pilot Studies

P. K. Yonge Laboratory School
Research Monograph No. 7

ED101002

THE SOCIAL SCIENCE OBSERVATION RECORD:

THEORETICAL CONSTRUCT AND PILOT STUDIES

BY

J. DOYLE CASTEEL

DEPARTMENT OF SECONDARY EDUCATION

AND

ROBERT J. STAHL

P. K. YONGE LABORATORY SCHOOL

P. K. YONGE LABORATORY SCHOOL

RESEARCH DEVELOPMENT

COLLEGE OF EDUCATION
UNIVERSITY OF FLORIDA
GAINESVILLE 32611

ACKNOWLEDGMENTS

The Social Science Observation Record (SSOR) was developed over a five-year-period—1969 through 1973. Completion of this monograph is the result of interaction with, and the assistance of, a large number of people.

When original efforts were made to describe those student behaviors which could be associated with value clarification, Dr. John Jarolimek (University of Washington) and Dr. Ambrose Clegg (Kent State University) offered encouragement and advice.

Since then, three sections of the College of Education at the University of Florida have been most helpful. These sections and the names of those who have assisted the authors are the following:

Department of Secondary Education: Dr. Eugene Todd (Chairman of the Department), Dr. John Gregory (Assistant Professor of Education), and Dr. Eugene Timmerman (Assistant Professor of Education).

Institute for Development of Human Resources: Especially Dr. Ira J. Gordon (Director of the Institute) and Dr. Robert S. Soar (Research Associate, Professor of Education).

P.K. Yonge Laboratory School: Especially Dr. J.B. Hodges (Director of the Laboratory School), Sandi Damico (Research Associate), Ruth Duncan (Coordinator for Dissemination), Robert Gasche (Director of Learning Resources), and Dr. Kirby Stewart (Coordinator of Research).

Donald R. Sloan developed the computer program for handling and analyzing SSOR data which eased the work of the authors and provided more accurate treatment of the data.

Too numerous to mention are the names of the graduate and undergraduate students from the College of Education who participated in various SSOR-related studies over the past three years. Secondary school students who participated in microteaching experiences were drawn from Buchholz High School, Gainesville High School, Howard Bishop Middle School, The Lincoln Center for Human and Industrial Arts, Westwood Middle School, and P.K. Yonge Laboratory School. All schools are located in Gainesville, Florida

PREFACE

Every teacher has experienced the frustration of having had a carefully planned classroom discussion session fail to yield the outcomes he desired, and of not knowing exactly what "went wrong." Although group discussion is used more than any other instructional process in elementary and secondary classrooms, many teachers have difficulty in identifying the elements which contribute to the success or failure of their class discussions. These teachers have not found an adequate vehicle for helping themselves analyze exactly what occurs in a group discussion. Consequently, they are unable to obtain the data they need in order to use the discussion process with increasing effectiveness. The studies described in this monograph focuses upon this problem.

The SSOR is the product of five years development and testing. It is a systematic observation instrument useful to teachers in planning, implementing, and analyzing classroom verbal and non-verbal behaviors. The seventeen categories and accompanying definitions describe behaviors directed toward values clarification as well as subject-matter instruction. Thus, it describes classroom behaviors associated with intellectual operations associated with both the cognitive and affective domains.

Findings of the studies based on use of the instrument indicate that:

1. Verbal and non-verbal behaviors identified by the instrument accurately categorize what occurs during classroom discussions.

2. Teachers who know the instrument can modify their planning for discussion in terms of categories of statements they want to hear, can shape their teacher-questioning behaviors in directions they perceive as desirable.

3. Teachers who receive feedback data in terms of the system express positive feelings about their experience.

4. People wanting to know the system can become reliable observers with less than 18 hours of training.

P. K. Yonge Laboratory School is in a position to offer workshops presenting the system and providing practice in its application. Workshops can be made available to instructional leaders for use in their work with teachers. Ideally, workshops can be offered to groups of teachers interested in refining their skills in the area of verbal and non-verbal interaction. Workshops can also be provided for individuals desiring to learn the system for research purposes. If interested in sponsoring or attending a workshop, please contact either of the authors or me.

Dr. J. B. Hodges, Director
Professor of Education
P. K. Yonge Laboratory School
College of Education
University of Florida
Gainesville, Florida 32611

CONTENTS

LIST OF FIGURES

Chapter I

INTERACTION OBSERVATION SYSTEMS: AN OVERVIEW AND CLASSROOM IMPLICATIONS

An Introduction

Classroom teachers are familiar with observation instruments designed to "rate" them as teachers. They have had experience, for example, with rating instruments as these are often used for purposes of annual faculty evaluations. Rating scales tend to be loosely defined, high inference instruments whose users may not be adequately trained to collect data. The use of such instruments by subjective but influential observers has aroused legitimate concern that data so obtained lack sufficient objectivity to be of value but, nevertheless, carry adequate professional weight to be personally threatening to the classroom teacher. Ambiguous rating scales containing numerous high inference categories that cannot be used reliably are not to be confused with systematic observation instruments (Rosenshine, 1970b).

Systematic observation instruments are also referred to as category systems. Category systems enable those trained in their use to collect objective data and to study instructional behavior analytically. Category instruments are designed to be descriptive, non-evaluative, and objective (Medley and Mitzel, 1963; Simon and Boyer, 1970). This monograph presents one such category instrument--the Social Science Observation Record (SSOR). More specifically the purposes of this monograph are:

1) To describe the objective and yet diverse nature of category observation systems,

2) To synthesize an explanation of how pre-service and in-service teachers may use category observational systems to study their behavior as teachers,

3) To present a model of class discussion relevant to student understanding and to value clarification,

4) To identify and describe specific categories of student and teacher behaviors through which student understanding and value clarification behaviors can be planned for and employed during class discussion,

5) To develop the knowledge and skills necessary for teachers to organize and interpret data,

6) To report empirical studies with regard to the modification of teacher and student behaviors, and

7) To describe how observers have been trained to collect data with reliability using the Social Science Observation Record (SSOR) system.

Objectives 1 and 2 are pursued in the remainder of this chapter. Objectives 3 and 4 are the focus of Chapter II, and objective 5 of Chapter III. Pilot empirical studies are reported in Chapter IV. Finally, two between-observer reliability studies are reported in Chapter V.

Observation Systems in Historical Perspective

Historically, those who have constructed observation instruments have sought to devise means for collecting empirical data descriptive of what occurs in the classroom (Rosenshine and Furst, 1973; Medley and Mitzel, 1963). Pursuant to this goal, instruments intended to describe the verbal, non-verbal, climatic, logical, cognitive, affective, and managerial dimensions of the classroom have been devised (Simon and Boyer, 1970; Rosenshine, 1970a; Rosenshine and Furst, 1973). Generally, relationships between students and their teachers within the classroom environment have been emphasized. Typically, these student-teacher relationships have been broken down analytically into elements believed to be variables which influence student learning (Flanders, 1965; Morine et al., 1971; Amidon and Flanders, 1967a). These elements when identified, defined, and organized, have become category observation systems by which descriptive data are collected, organized, stored, and analyzed (Simon and Boyer, 1970; Rosenshine and Furst, 1973). The central thrust of these efforts has been to describe classroom events by isolating variables in the form of categories that are used in order to analyze teacher behavior as it relates to student learning (Flanders, 1965; Medley and Mitzel, 1963; Simon and Boyer, 1970; Rosenshine, 1970b; Rosenshine and Furst, 1973).

Whereas the developers of observation instruments have consistently valued objective data collection as a tool for the analysis of student-teacher interaction, different interests, alternative objectives, conflicting convictions, and competing beliefs as to what variables are most critical to student learning have eventuated in an equally divergent collection of instruments. (For an overview of seventy-nine of these systems, see Mirrors of Behaviors [Simon and Boyer, 1967, 1970].) Some instruments identify factors associated with classroom climate (Withall, 1951; Amidon and Hough, 1967). Some stress indirect and direct categories of teacher behaviors (Flanders, 1960). Some stress student and teacher cognitive behaviors (Brown et al., 1967; Smith and Meux, 1970). Other systems focus on student behaviors associated with the study of social issues and the analysis of public policies (Massialas, 1969; Oliver and Shaver, 1966). Still others specify additional emphases in terms of such factors as content field (Bellack et al., 1966), cognitive processes (Smith and Meux, 1970), and teacher experimentalism (Brown, 1968). Hence, a variety of systems are available for collecting objective data about classroom behavior.

Many category systems are designed solely to describe the verbal interaction of students and teachers under classroom or micro-teaching simulation conditions. Systems stressing verbal behaviors are referred to as verbal interaction observation systems. The Social Science Observation Record (SSOR) is a verbal interaction observation system.

Developers of interaction observation systems have used predetermined and carefully defined categories of verbal and non-verbal behaviors to describe teacher and student verbal behaviors. To the degree that the data collected are descriptive of classroom behavior, what has been observed may be reconstructed and analyzed; hence, these systems are sometimes refererred to as "mirrors" of behavior (Simon and Boyer, 1970). The accuracy with which data reflect classroom events is limited by the number of categories incorporated into the observation system;by the skill of observers who collect data; by the conceptual tools available for reconstructing, interpreting and analyzing the data; and by the adequacy of the categories as [redacted] descript[redacted] of observed behaviors. The Social Science Observation Record contains seventeen verbal and non-verbal categories serving as descriptors of observable classroom behaviors. The system also provides mechanisms by which the data may be reconstructed or "stored" for analysis and interpretation.

Interaction observation instruments are intended to collect empirical data systematically. The accuracy with which an observer can collect data and the degree to which the meaning of categories can be communicated are critical factors within these stystems. Category systems are characterized by precisely defined categories. This precision reduces the need for coders to infer, and thereby frees them from the necessity of deciding which category to assign to classroom events.

These definitions also enable those who know them to communicate with one another. Thus, observers can provide classroom teachers with reality data by which they may interpret and make decisions about their behavior in the classroom.

The descriptive nature of interaction observation systems is to be contrasted with the evaluative nature of rating systems. One of the primary purposes of a rating system is to evaluate or "grade" a behavior or the person deemed responsible for the behavior (Remmers, 1963). The primary function of descriptive systems is to collect accurate, objective data for subsequent analysis. Those who have designed category systems relevant to interaction analysis have constructed and recommended them in terms of their descriptive and analytical power. The Social Science Observation Record was developed as a descriptive instrument.

Classroom interaction observation instruments consistently focus on one or more of the following: student cognitive behavior, student affective behavior, and student or teacher managerial behavior (Simon and Boyer, 1970). Silence or both silence and confusion are frequently added to these systems as catchalls for all the behaviors which do not fit the other categories in the system. Cognitive category systems tend to stress such student behaviors as recall, defining, inferring, interpreting, applying, opining, and evaluating (Simon and Boyer, 1970). Affective category systems tend to stress student feelings and emotions and such factors as teacher warmth. (Simon and Boyer, 1970). Category systems emphasizing control or management tend to be keyed to lecturing and disciplining as teacher behaviors (Rosenshine and Furst, 1973; Simon and Boyer, 1970).

Some instruments embrace teacher behaviors and student affective performances (Aschner et al., 1967; Fuller, 1970). At least one instrument used categories designed to collect data relevant to affective behaviors such as awareness, responding, affirming, and valuing as student behaviors (Kaplan, 1970). Others incorporate teacher and student cognitive performances (Brown, 1968; Brown et al., 1967; Massialas, 1969). The SSOR contains categories of behaviors which have been associated with the cognitive, affective, and control dimensions of the classroom. Therefore, the Social Science Observation Record incorporates categories relevant to student understanding, valuing, and feeling, and to teacher control and management as well as non-verbal behavior in the classroom.

Interaction analysis systems can be learned by classroom teachers, supervisors, and other educational personnel. Learning may be defined as understanding the categories of a system. Learning may be defined as the ability to understand and interpret data as it is explained by a trained analyst. Learning may be defined as the ability of a teacher to organize his own data for his own interpretive purposes. Learning may be defined to include the ability to code data reliably for one's own use. The SSOR has been taught and learned in each of the senses mentioned above. No matter what level of learning is achieved, there is evidence to suggest that the classroom teacher who "learns" an observation system is more effective than one who does not (Soar, 1968).

Summary

Category systems are instruments designed to collect data relevant to the behavior of teachers and students in the classroom. Interaction observation systems collect samples of classroom behavior in terms of the categories they include. These samples of behavior can be subjected to analysis and interpretation. These systems are developed for purposes of collecting objective data and are not recommended as evaluative devices. Interaction analysis systems stress divergent aspects of classroom behavior. Classroom teachers can learn observation systems for use as consumers, interpreters, or collectors of data. To the degree that a category system reflects the instructional goals of the teacher, that system possesses value and validity for instructional decision making by the teacher.

Chapter II

THE SOCIAL SCIENCE OBSERVATION RECORD: A DISCUSSION MODEL

An Introduction

The Social Science Observation Record (SSOR) is a systematic observation system designed to abstract and describe class discussion. The system was conceived and constructed as a theoretical model for planning and guiding classroom discussions, specifically those discussions directed toward value clarification as an aspect of subject-matter instruction. Use of the model increases precision in planning and provides means by which teachers can guide class discussion systematically. The concepts and ideas presented in this chapter focus on these applications of this theoretical model.

For the classroom teacher to use the Social Science Observation Record as a model for discussion, he need not be a trained analyst. Once he is familiar with the components of this theoretical system, he can apply his knowledge as he plans, organizes, and leads class discussions. Application increases the number of instructional moves he makes and, consequentially, the range of behaviors his students practice. He becomes more alert to what students are saying during class discussions and, consequentially, responds to and guides their behaviors more accurately and purposefully. He operationalizes instructional theories associated with student understanding and value clarification on the basis of specific and observable student verbal behaviors and, accordingly, explores new dimensions of student behavior toward which he plans and teaches. Having acquired facility in applying the SSOR as a model for classroom discussion, he can better comprehend how a trained coder collects, organizes, and interprets data relevant to him as a classroom teacher.

The Social Science Observation Record as originally conceptualized was, and continues to provide those who understand it with, a model for class discussion. In the process of learning to apply the model, it is necessary, first, to define class discussion as conceptualized within this framework. Accordingly, when the generic term class discussion is used in connection with the SSOR, the following is meant:

> ...At least sixty-five percent (65%) of the behavior is recorded as student, teacher, or as both student and teacher talk.
>
> ...At least ten percent (10%) of the behavior occurring is student talk. The teacher need not talk for a class discussion to occur.

Class discussion as a generic form of activity may assume different forms within these relevant attributes. The following are adequate for subsequent discussions of the SSOR:

> Open discussion. At least sixty percent (60%) of the behavior occurring is coded as student verbal behavior. No more than forty percent (40%) of the behavior coded is teacher verbal behavior.
>
> Guided discussion. At least forty percent (40%) of the behavior occurring is coded as student verbal behavior. No more than sixty percent (60%) of the behavior coded is teacher verbal behavior.
>
> Directed discussion. At least twenty percent (20%) of the behavior occurring is coded as student verbal behavior. No more than eighty percent (80%) of the behavior coded is teacher verbal behavior.
>
> Lecture discussion. At least ten percent (10%) of the behavior occurring is coded as student verbal behavior. At least sixty percent (60%) of the behavior occurring is coded as teacher verbal behavior.

By employing these definitions in conjunction with an understanding of the SSOR, teachers, teacher educators, and analysts can convey more clearly and precisely their intended meaning when the term class discussion is their referent. When the term class discussion is used in the sections which follow, the first definition is meant; when the intent is to address attention to a specific form of class discussion the adjectives open, guided, directed, and lecture will so signify.

The Social Science Observation Record is framed by categories of behavior associated with class discussion and organized into four realms of inquiry. Definitions and explanations of the realms of their seventeen categories follow:

Realms of the Social Science Observation Record

The four realms of the SSOR -- Subject Centered, Teacher Centered, Man Centered, and Non-Verbal -- are also referred to by Roman numerals.

Realm I -- Subject-Centered

Realm II -- Teacher-Centered

Realm III -- Man-Centered

Realm IV -- Non-Verbal

In order to examine how these four realms are used as components within the system, each will be defined in light of its function as an aspect of class discussion.

Realm I -- Subject Centered

The Subject-Centered Realm emphasizes student understanding of the content of instruction. More specifically it focusses upon the formal content of subject matter learning as such learning occurs according to the rules of disciplined knowing, a characteristic of acquiring a body of knowledge that can be associated

BEST COPY AVAILABLE

with academic disciplines. Student demonstration of this acquisition represents the minimal objectives any teacher may logically and ethically specify for his students.

Three major student behaviors may be interpreted as indications of understanding of concepts. First, students learn and recall content in the form of data, ideas, opinions, and beliefs. Although recall is insufficient in and of itself for inferring that understanding has occurred; nevertheless, it is recall which enables students to acquire new understandings. Secondly, students assign meaning to the data, ideas, opinions, and beliefs retrieved through recall, thus personalizing the knowledge they have gained. Students' assignment of personal meaning to recall is an indication they have acquired some degree of understanding. This assignment [illegible] may be, and often is, a conventional understanding in light of the student's [illegible] and society. This is not necessarily an understanding which a scholar knowledgeable with regard to subject-matter content would find acceptable. A third step helps overcome this eventuality. During the third phase, students identify and define terms and concepts which may be used as instruments for imposing learned opinions on knowledge as it is acquired and recalled. Thus, defining (and particularly the defining of interpretive concepts) is an essential process in seeking understanding. These three dimensions of understanding--recall, personalization, and defining--are incorporated into the SSOR and assigned to the Subject-Centered Realm.

Identifying concrete examples of student verbal behaviors will indicate the nature and extent of acquisitions within the Subject-Centered Realm. When students are developing understanding, they refer to the focus of study.

They compare and contrast ideas, data, and relationships. They identify and define new words germane to the new knowledge they are gathering or interpreting or to the discussion in which they are engaged. They organize, structure, and elaborate on ideas which enable them to express and share their interpretations and perceptions of the content being learned. They clarify and elaborate what they know, believe, and think. Each of these behaviors are examples of verbal behaviors assigned to the Subject-Centered Realm of the SSOR.

When the goal of instruction is one of helping students to learn and understand content, the Subject-Centered Realm becomes a major focus of instruction. When data collected using the Social Science Observation Record cluster in the Subject-Centered Realm in patterns describing students who are gathering and interpreting data according to defined concepts, one may infer that students are engaged in acquiring knowledge, imposing personal meaning on the knowledge learned, and deriving interpretations from the instrumental use of concepts. Using teacher-centered behaviors, the teacher structures learning contexts and stimulates student behaviors which influence how students proceed in the Subject-Centered Realm.

Realm II -- Teacher-Centered

The Teacher-Centered Realm is associated with those verbal behaviors by which teachers organize and manage discussion activities. When the teacher opts to use class discussion as an instructional approach, he frequently initiates the activity, provides directions, reacts to student verbal and non-verbal behaviors, responds to ideas and questions of students, and culminates the discussion activity. In effect, the teacher organizes, manages, guides, and ultimately controls the verbal performance of students.

All verbal behaviors exhibited by the teacher during a class discussion are coded as instances of the Teacher-Centered Realm.

As the teacher organizes, manages, guides, and controls class discussions he responds to his students and to the instructional situation itself in a variety of predictable ways. He presents instructional situations within which he expects students to follow directions or respond to questions. He summarizes what members of a discussion group have accomplished, points out the strengths of their performance, indicates fallacies or lack of clarity in statements, and notes limitations in understandings. He reacts to student behavior in a variety of ways: (1) by criticizing ideas as wrong, inappropriate, or possessing weaknesses; (2) by commending ideas as useful, correct, or worthy of investigation; (3) by indicating that expressions are not clear; and (4) by suggesting that ideas and beliefs may be inconsistent with one another. Each of these behaviors are catagorized in Realm II. All are ways of initiating and responding to student verbal and non-verbal behaviors.

As the classroom teacher develops his understanding of the Social Science Observation Record, he plans and monitors various aspects of his own verbal performance. By way of illustration, the teacher considers how he elicits student behaviors; how much time he spends organizing, managing, guiding, and controlling the discussions; how students react to his verbal performance; how and to what extent he structures verbal behaviors; and what kinds of student behaviors he values. If provided with reliable data collected by a trained coder, he can assess his behavior as teacher on the basis of his impact on student performance. When the teacher acts as leader of class discussion, he performs most of the leadership functions incorporated in Realm II. Students also use this realm. Students ask questions. Students report that they do not understand ideas expressed by other students or by the teacher and request clarification. They criticize and reinforce the ideas of other students. When students engage in any of these behaviors, they are coded as behaviors in the Teacher-Centered Realm.

The primary function of Realm II is to facilitate student verbal behavior in Realms I and III. If the instructional objectives being pursued are related to student understanding, the teacher uses Realm II in eliciting, guiding, and reacting to student behaviors in the Subject-Centered Realm. If the teacher's instructional objectives are related to the clarification of student values, he uses teacher-centered behaviors to evoke student behaviors in the Man-Centered Realm.

Realm III -- Man-Centered

The Man-Centered Realm stresses verbal behaviors in which students engage in order to clarify their values and feelings. As with understanding, value clarification is one of the major goals of formal education.

If one views schooling as those experiences designed to prepare students to cope with and modify the world in which they live, then they must learn disciplined ways of engaging in conflict resolution, achieving rational consensus, and pursuing their personal welfare as members of a civil society. Students are encouraged to perceive the following:

> Logical knowledge and rules of knowing, thinking, and acting may be made relevant to their experiences.
>
> Knowledge gained experimentally and accepted on the basis of probability may be used to shape their public and private worlds.
>
> Knowledge of the natural and historical contexts within which

events occur may be made relevant to the decisions men must make.

Structuring of feelings, beliefs, and commitments in the arts may be made relevant to the ways men communicate with each other.

Exercise of freedoms through which men speak freely, join groups, explore social alternatives, and yield their assent rationally are legitimate.

To achieve such ends, subject-matter fields are explored as a basis for practicing and refining skills of disciplined, practical judgment. (Raup et al., 1943). Further, teachers plan and guide activities, including class discussions, which stress the clarification of student values in light of student understanding of subject matter. The Man-Centered Realm addresses itself to the value clarification dimensions of class discussion.

In practice, value clarification assumes different guises and is reflected in different patterns of students' language choices. Students react verbally, revealing their value judgments and personal feelings relative to a situation. Students examine those understandings and concepts most vital in their lives and through which they make sense of their public and private behavior choices. Students identify or invent personal and social policy alternatives and explore the benefits and costs likely to be incurred. Students posit an ideal state of affairs or resolve a policy conflict by developing a consensual basis for arriving at agreement. Hence, value clarification becomes the focus of the Man-Centered Realm and is an integral aspect of subject-matter learning.

The Man-Centered Realm provides a conceptual focus which is applied to planning, leading, and analyzing class discussions designed to help students clarify their values. When the teacher begins to structure situations and to raise questions designed to elicit student behaviors in Realm III, he is more likely to secure student responses if his questions are based on student understandings developed in Realm I and if he considers the Non-Verbal Realm as an aspect of his conceptual model of class discussion.

Realm IV -- Non-Verbal

The Non-Verbal Realm provides for the collection of data identifying periods of silence or confusion which accompany class discussions. While the SSOR is basically a verbal system, these two dimensions of non-verbal behavior are important categories in the system.

Silence enables students to organize their thoughts. If the teacher makes an instructional move designed to elicit student behaviors in either the Subject-Centered or the Man-Centered Realms, the best evidence that he wants students to respond and that he places instructional value on his questions is his willingness to wait while students collect and frame their thoughts for presentation to the group. If the teacher respects the ideas and opinions of his students, one of the most effective ways of communicating his respect is to remain silent after each student finishes speaking until he is relatively certain that the student has presented his thoughts to his own satisfaction. This pause or waiting period is an example of silence as it is recorded in the SSOR.

The Non-Verbal Realm also includes a gross form of behavior designated as confusion. Instances in which students laugh, several students talk at once, the teacher or students talk so quietly that they cannot be understood, or student dialogue degenerates into a shouting match are all instances of confusion.

Because these two categories of behaviors -- silence and confusion -- are deliberately built into the model, SSOR data may be used to estimate their function in class discussion. Although the Social Science Observation Record is accurately catalogued as a verbal system, the Non-Verbal Realm is a viable component.

To recapitulate, the SSOR encompasses four realms. These realms stress student understanding of content, teacher-control behaviors, student value clarification, and teacher or student non-verbal behaviors. The preceding discussion of the four realms is summarized in Figure 1. The first column designates each realm by Roman numeral; the second designates each realm by

Realm Number	Realm Label	Employed by	Major Focus
I	Subject centered	Students only	Understanding
II	Teacher centered	Teacher or students	Facilitation
III	Man centered	Students only	Value clarification
IV	Non-Verbal	Teacher or students	Inaudible responses

Figure 1. The four realms of the SSOR.

its name; the third indicates whether each realm is used by students, teacher, or by both the teacher and students; the fourth reviews at least one major function for each realm.

The realms of the SSOR are used interactively. The Subject-Centered Realm establishes, reviews, or diagnoses student understandings and the adequacy of their concepts. This understanding becomes either the basis for or the focus of man-centered inquiry. In the latter instance, students engaging in value clarification first demonstrate that they understand the subject of inquiry and then engage in value clarification. When a discussion proceeds in this manner, the flow is from the Subject-Centered Realm to the Man-Centered Realm.

Class discussions may also flow from the Man-Centered Realm to the Subject-Centered Realm. If during the course of a value clarification discussion students recognize a need for more data, more accurate interpretations, or definitions of new words, they may move to the Subject-Centered Realm until these needs are fulfilled. Then they may return to the Man-Centered Realm.

The teacher is not limited to observing student movement between Realms I and III. Preferably, he takes an active part. For example, he plans and transacts strategies to stimulate students' man-centered statements. He elicits preferences, ideals, policies, and feelings. He then uses instances of student behaviors evoked in the Man-Centered Realm as the focus of subject-centered behavior. He asks students to collect and interpret data according to analytical concepts to gain insight into how they as students value, decide, share feelings, and identify and apply ideals. Subject-centered learnings so derived then become the focus for helping students consider the consequences of their value-based behaviors and of inviting students to consider how they might wish to modify the means they use for sharing feelings, making decisions, implementing ideals, and affirming personal and group preferences.

A dynamic interaction between these two realms is significant. Subject-centered statements add validity, depth, and complexity to man-centered inquiries. At the same time, man-centered inquiries become relevant points of departure for subject-centered inquiries.

Ways in which Realms I, II, and III may be used systematically are presented schematically in Figure 2. Arrow A depicts the flow of student verbal behavior from the Subject-Centered to the Man-Centered Realm without teacher mediation. Arrow D depicts the flow of student verbal behavior from the Man-Centered Realm to the Subject-Centered Realm without teacher mediation. Arrows B and C show the flow from the Subject-Centered Realm to the Man-Centered Realm and from the Man-Centered Realm to the Subject-Centered Realm, respectively, with teacher mediation.

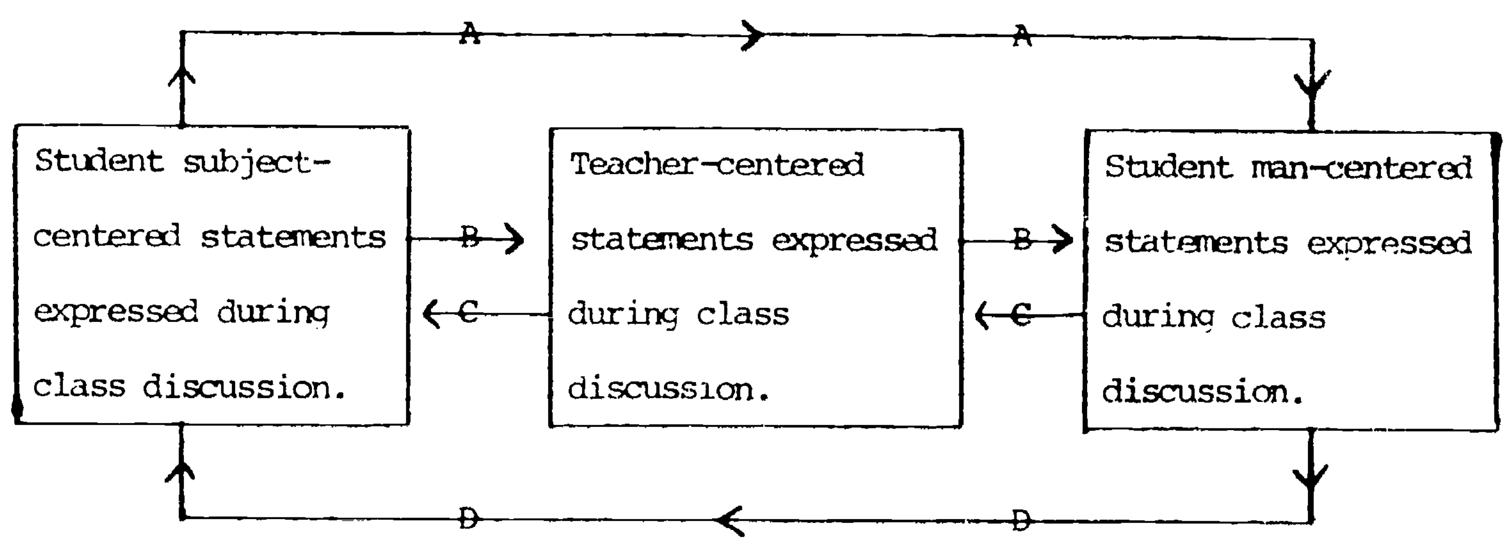

Figure 2. The flow of verbal behavior in the SSOR.

Figure 2 stressed the flow of discussion on the basis of the verbal realms and ignores the Non-Verbal Realm. Figure 3 incorporates the Non-Verbal Realm and identifies six paths along which discussion may move between Realms I and III.

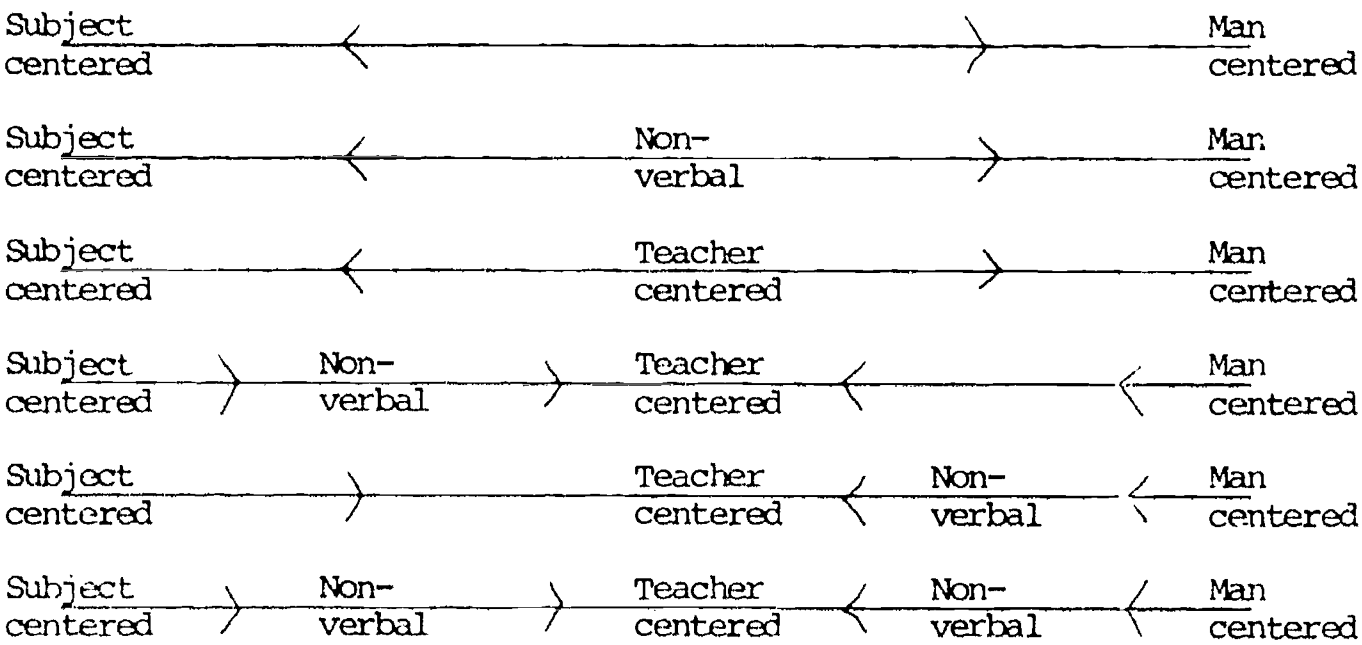

Figure 3. The flow of verbal and non-verbal behaviors in the SSOR.

The four realms of the SSOR enable one to think systematically about class discussions. To think with precision it is necessary to couple an understanding of the realms with a knowledge of the seventeen categories embraced by the system as a model for discussion germane both to the development of student understandings and to the clarification of student values. These categories are now presented and discussed.

Categories of the SSOR: A Functional Introduction

The Social Science Observation Record recognizes seventeen categories of behavior. Within this descriptive system no category is "good," no category is "bad." Each category may, according to a teacher's objectives, operate functionally as an aspect of student and teacher behaviors. The seventeen categories are surveyed rapidly here in order to provide information relative to their location with the four realms and to stress that each may function during class discussions.

The Subject-Centered Realm contains five categories of statements topical, empirical, interpretive, defining, and clarifying. These are listed and at least one function for each identified in Figure 4.

CATEGORY OF STATEMENT - NUMERICAL DESIGNATION	CATEGORY OF STATEMENT - NAME	CATEGORY OF STATEMENT - FUNCTION
1	Topical	Maintaining focus
2	Empirical	Stating facts
3	Interpretive	Assigning meaning
4	Defining	Avoiding semantical confusion
5	Clarifying	Elaborating ideas

Figure 4. The subject-centered categories of the SSOR.

Coding, using these five categories, is limited to identification of students' behaviors. Thus, teachers conceptualize, plan, guide, and analyze discussions with precision where the goal of instruction is one of developing student understanding.

The Teacher-Centered Realm also contains five categories of statements -- informing, commentary, dissonant, interrogative, and confirming. These also are listed and at least one function for each identified in Figure 5.

CATEGORY OF STATEMENT - NUMERICAL DESIGNATION	CATEGORY OF STATEMENT - NAME	CATEGORY OF STATEMENT - FUNCTION
6	Infirming	Criticizing
7	Commentary	Consolidating and structuring
8	Dissonant	Requesting clarification
9	Interrogative	Eliciting responses
10	Confirming	Reinforcement

Figure 5. The teacher-centered categories of the SSOR.

Both the teacher and students use these five categories of behavior. All teacher talk must be assigned to one of these five categories. Thus, the teacher uses these categories to organize, manage, conduct, react, and direct class discussion.

The Man-Centered Realm also contains five categories of behaviors -- preferential, consequential, criterial, imperative, and emotive. One function for each of these categories is identified in Figure 6.

CATEGORY OF STATEMENT - NUMERICAL DESIGNATION	CATEGORY OF STATEMENT - NAME	CATEGORY OF STATEMENT - FUNCTION
11	Preferential	Assigning value ratings
12	Consequential	Anticipating effects
13	Criterial	Identifying grounds
14	Imperative	Considering decisions
15	Emotive	Expressing feelings

Figure 6. The man-centered categories of the SSOR.

BEST COPY AVAILABLE

Only students' behaviors are so coded, and teachers plan value clarification activities and discussions on the basis of these categories.

The Non-Verbal Realm consists of two categories -- silence and confusion. Figure 7 lists at least one function for each category.

CATEGORY OF STATEMENT - NUMERICAL DESIGNATION	CATEGORY OF STATEMENT - NAME	CATEGORY OF STATEMENT - FUNCTION
16	Silence	Wait time
17	Confusion	Adjustment time

Figure 7. The non-verbal categories of the SSOR.

The seventeen categories of the SSOR are summarized according to realm, numerical designation, categorical label, and at least one function as shown in Figure 8. These seventeen categories enable the teacher to plan and lead class discussions designed to eventuate in student understanding, value clarification, or both with precision and purposefulness. To secure these benefits the teacher needs to master definitions for each of the seventeen categories. If there is a possiblity that the services of a trained analyst will be available, the teacher should begin to practice thinking of each category not only by its name but also by its numerical designation.

The Seventeen Categories Defined

Each of the Categories constituting the frame of the SSOR are defined next. The number which accompanies the label for each category is the nominal number by which the category is coded by analysts.

REALM	CATEGORY OF STATEMENT	FUNCTION
I. Subject centered	1. Topical	Maintaining focus
	2. Empirical	Stating facts
	3. Interpretive	Assigning meaning
	4. Defining	Avoiding semantical confusion
	5. Clarifying	Elaborating ideas
II. Teacher centered	6. Infirming	Criticizing
	7. Commentary	Consolidating and structuring
	8. Dissonant	Requesting clarification
	9. Interrogative	Eliciting responses
	10. Confirming	Reinforcement
III. Man-centered	11. Preferential	Assigning value ratings
	12. Consequential	Anticipating effects
	13. Criterial	Identifying grounds
	14. Imperative	Considering decisions
	15. Emotive	Expressing feelings
IV. Non-verbal	16. Silence	Wait time
	17. Confusion	Adjustment time

Figure 8. The seventeen categories of the SSOR. The functions as given are meant to be illustrative but not inclusive. (Source: J. Doyle Casteel and Robert J. Stahl, c. 1972.)

Category 1: Topical statements

Students express topical statements to identify and maintain the focus of discussion. Topical statements give direction to student statements in other categories. The focus of discussion may be a theme (fear), a concept (permutations), an issue (law versus morality), a problem (ecological imbalance), or a question (How can we best explain the American Revolution?). Often the topic of discussion is presented by the teacher as he initiates class discussion. When students restate or attempt to restate the topic of study voluntarily or in response to teacher directions or questions, statements are coded as topical statements.

Category 2: Empirical statements

Empirical statements provide an objective thrust to class discussions. When students specify data, numbers, names, and events, statements are coded as empirical statements. When students share knowledge or what they believe to be true, these statements are also coded as empirical statements. When a student reports what he has read, heard, observed, viewed or remembered as descriptive data, the assumption is that he accepts and expects the group to accept his statements as factually accurate. If either the teacher or students disagree, the statement heretofore purported to be an empirical statement becomes an interpretation. If the student chooses to repeat the idea as one he still believes to be accurate, the statement becomes his interpretation and is thus coded as an instance of category 3, or an interpretative statement. Empirical statements occur as students acquire new knowledge, review knowledge, or seek to pool knowledge germane to a discussion.

Category 3: Interpretive statements

Interpretive statements occur when students seek to assign meaning to data, experiences, and behaviors. When students claim that data or ideas are germane to the topic of discussion, their statements are coded as empirical. When students claim that data or ideas are not relevant to the topic of study, these are coded as interpretive statements. When students compare or differentiate between objects, events, situations, policies, etc., statements are coded as interpretive. When students generalize the meaning of more than one piece of data (what was read, said, seen, etc.), these are also coded as interpretive. Students' interpretive statements are frequently prefaced by such phrases as "I think that," "In my opinion," "I guess that," and "It might mean." Although these are not always valid cues that an interpretive statement is about to be made, in most instances students do proceed to assign meaning.

Category 4: Defining statements

Defining statements enable groups engaged in class discussion to develop the attributes of concepts logically and to use new words. The use of this category tends to provide for semantical clarity and to enable students to develop understandings on the basis of analytical concepts. Defining statements are read from or remembered from an acceptable source (textbooks, reading, dictionary). Defining statements are expressed extensionally through listing examples. Defining statements are expressed operationally in behavioral terms. Defining statements for concepts are expressed in terms of relevant attributes. Defining statements may be (but seldom are) expressed as ideal types by students.

Category 5: Clarifying statements

Clarifying statements are used when students elaborate on or use other verbal maneuvers to improve the communication of what they have said or are about to say. Students clarify by restating rambling remarks more concisely ("What I'm trying to say is..."); and students clarify by restating ideas in more detail. Clarifying statements frequently occur as phrases containing such words as "because," "personally," "it's like this," "in this situation." In effect the student states that he has a "personal" basis but does not report it as his criterion.

Category 6: Infirming statements

Infirming statements are statements rejecting or criticizing statements made by either the teacher or other students. When an idea is labeled as wrong, inaccurate, or needing improvement to be acceptable, an infirming statement delivers the message. Comments intended to ridicule a student or such reactions as "Shut up," "Calm it down," "You're too noisy," "You aren't following directions," and "You've missed the point" are also examples of infirming statements. Where the intent is to identify flaws in an otherwise sound idea, infirming statements are often preceded by a complimentary introduction ("I can see value in your idea,") and then translated into infirming statements ("but I believe you've left this information out."). As the realms and categories of the SSOR are defined, both teachers and students make infirming statements.

Category 7: Commentary statements

Commentary statements are made by teachers in order to initiate, summarize, or structure. Commentary statements are coded when the teacher responds to student requests for information, opinions, values or directions. If the teacher identifies the focus of discussion, the statements are coded as "commentary." If the teacher provides a definition or describes a criterion for purposes of helping students to interpret the meaning of data or to make a decision, his statements as he presents the definition or criterion are commentary statements. All instances of narrative, teacher talk not clear instances of infirming, dissonant, or confirming statements are commentary statements. Rhetorical teacher questions are also coded as commentary statements. Although students use this category to review directions, this is almost exclusively a teacher category when the teacher leads a class discussion.

Category 8: Dissonant statements

Dissonant statements have three important purposes: to express a need for clarification, to inform members of the group that one is intellectually confused, or to point out that a participant (or participants) is making statements which are inconsistent with one another. Typically, teachers expletives ("What?", "Huh?", "Say that again!") fall into this category. Students typically use this category by employing short statements such as "Is this what you want me to do?" or more likely "I don't understand." As the realms and categories of the SSOR are defined, both teachers and students use dissonant statements.

Category 9: Interrogative statements

Teachers use interrogative statements any time they raise a question. Teachers may use interrogative statements to elicit any of the ten categories of statements reserved exclusively for student verbal behavior in the Subject-Centered or the Man-Centered Realms. If each of these student categories is thought of as a move, t' teacher may use interrogative statements to initiate ten student verbal moves. As the realms and categories of the SSOR are defined, both students and teachers make interrogative statements.

Category 10: Confirming statements

The teacher or students use confirming statements to reinforce the behavior of the teacher and of students. Teachers use confirming statements in expressing acceptance of a student's idea or behavior; in acknowledging agreement with the ideas of students; and in encouraging students to continue a line of thought or valuation using such expressions as "I see," "Great," "Keep going," "Don't stop now," and sometimes "Wow." Students use similar expressions to confirm one another, to confirm the teacher's statements, and (unless the teacher is careful) to confirm their questions and encourage the teacher to answer his own questions, the questions intended for students. As the realms and categories of the SSOR are defined, either the teacher or students may use confirming statements.

Category 11: Preferential statements

Students use preferential statements to assign value ratings to ideas, policies, situations and other objects of valuation. Preferential statements frequently occur as students judge objects of valuation using such words as "good," "bad," "best," "worst." Students also use preferential statements to convey their likes and dislikes. Preferential statements occur when students select the preferable option from a limited number of alternatives or rank order objects of valuation from the most to the least preferable. Preferential statements also occur as students classify objects of valuation, grouping those they consider most preferable, least preferable, or both.

Category 12: Consequential statements

Students use consequential statements to express what they perceive to be known or anticipated effects. When students list the effects of a situation (poverty), they use consequential statements. When students list the benefits and costs of a public policy (open housing), they use consequential statements. When students identify the results likely to eventuate from their personal feelings or preferences, they express consequential statements. When students list the results likely to occur if a given interpretation is accepted, they use consequential statements. Student expressions designed to identify the effects, results, benefits, or costs of an idea, a condition, a decision, or a feeling are all coded as consequential statements.

Category 13: Criterial statements

Students use criterial statements to identify the grounds or norms implicit in, or deliberately being used to guide, their thinking, their valuing, or their actions. Criterial statements also serve as frames of reference, enabling students to state preferences, determine consequences, or decide between policy alternatives. Criterial statements may be conditional, in which case phrases such as the following are incorporated: "If this is true--," "When one is faced with this type of situation--," "Let's say one wants to do this--," "Suppose you felt in this way--," and "Imagine you were in his place--." Each phrase sets a condition for what has been or what is about to be said. Criterial statements may be normative: "All men are equal," "To use authority is bad; to be permissive good," and "All human life is valuable."

Category 14: Imperative statements

Imperative statements focus on decision making and decision reporting.

Students use imperative statements to identify alternative policies which might be used to change a situation; to state what they believe ought to be true or ought not to be true, thus describing ideal conditions; and to identify actions which should be taken or should not be taken. Further, when students as individuals or as members of a group formulate and report actions they intend to take, they use imperative statements.

Category 15: Emotive statements

Emotive statements function to express strong personal feelings. When students report how they felt, how they currently feel, or how they would feel toward an idea, a situation, a policy, or a consequence, they use emotive statements. Emotive statements are concerned with the feelings of students or with physical sensations students experience when these are expressed in a manner which denotes a sense of feeling, smell, or taste.

Category 16: Silence

Silence is used when the students or the teacher wait for or take time to formulate a verbal or non-verbal response to a question or idea. When students are computing an answer to a problem in mathematics or studying some picture projected on a screen, a brief period of silence may be expected. Silence also denotes the beginning and ending of coding in the SSOR.

Category 17: Confusion

This category is used to record verbal and non-verbal noise which accompanies, interferes with, disrupts, or makes impossible discussion. Confusion is also used to record student or teacher verbal statements stated so softly that the coder (and by inference members of the group) cannot determine what is being said.

The seventeen categories of the SSOR enable the teacher to plan and lead discussions with categorical specificity. The teacher plans subject-centered learning activities using categories 1 through 5 to develop student understanding. The teacher uses categories 6 through 10 to help plan his instructional behavior as teacher. The teacher plans man-centered learning activities using categories 11 through 15 to stimulate student value clarification. These seventeen categories and a short definition for each category are reviewed in Figure 9.

Summary

The Social Science Observation Record contains four realms and seventeen categories. The four realms are Subject Centered, Teacher Centered, Man Centered, and Non-Verbal. Realm I stresses student understanding and includes five categories of students statements: topical, empirical, interpretive, defining and clarifying. Realm II stresses teacher transactional behaviors and includes five categories which both the teacher and students employ: infirming, commentary, dissonant, interrogative, and confirming. Realm III stresses student value clarification and embraces five affective categories of student statements: preferential, consequential, criterial, imperative, and emotive. Realm IV comprises two non-verbal categories: silence and confusion. The classroom teacher who understands the SSOR may use his knowledge of these four realms and seventeen categories to plan and lead class discussions systematically. For these purposes, he does not require the assistance of a trained analyst. Before he can use the services of a trained analyst, when such is available, he needs to understand how data are collected and organized for feedback and analysis. This information is discussed in Chapter III.

REALM	CATEGORY OF STATEMENT	DEFINITIONS
I SUBJECT-CENTERED	1. Topical	Student statements identifying the theme, the unit, the concept, the issue, or the problem that is the focus of group discussion.
	2. Empirical	Student statements providing verifiable data from memory, observation, reading, or oral presentation.
	3. Interpretive	Student statements assigning meaning to data or experience and expressed in the form of notions, opinions, comparisons, relationships, and connections.
	4. Defining	Student statements as to the meaning of a word or concept by reference to an accepted source, by context, by examples, by operant criteria, or by ideal type.
	5. Clarifying	Student statements rewording, rephrasing, elaborating on, or expanding on other statements by ways of explanations.
II TEACHER-CENTERED	6. Infirming	Teacher or student statements of rejection, criticism, closure, or dissatisfaction expressed in the form of sarcastic, doctrinaire, or negative remarks.
	7. Commentary	Teacher or student statements reviewing or summarizing the directions of a group; or, teacher statements summarizing, consolidating structuring providing new information, new directions, or responding to student requests for information.
	8. Dissonant	Teacher or student statements indicating that what is being said is not understood, is causing confusion, or lacks either internal or external consistency.
	9. Interrogative	Teacher or student questions expressed during group interaction.
	10. Confirming	Teacher or student statements expressing acceptance, satisfaction, encouragement, or praise.
III MAN-CENTERED	11. Preferential	Student statements assigning a value rating or classification to an idea, person, group, object, etc.
	12. Consequential	Student statements identifying the known or anticipated effects of an action, idea, object, feeling, etc.
	13. Criterial	Student statements identifying the basis for a decision, a judgement, an action, an interpretation, etc.; or, developing a table of specifications for use in decision-making.
	14. Imperative	Student statements of what should or should not be; of what ought or ought not to be done; or expressing a decision achieved by the group.
	15. Emotive	Student statements indicating personal feelings; or, efforts to express empathy with regard to the personal feelings of others.
IV. NON-VERBAL	16. Silence	Period indicating quiet, absence of verbal interaction, reading, thinking, non-verbal activities, or work.
	17. Confusion	Verbal or non-verbal interference or commotion making it difficult for members of the group to communicate.

Figure 9. Short definition for the seventeen SSOR categories. (Source: J. Doyle Casteel and Robert J. Stahl, c. 1972.)

Chapter III

THE SSOR AS AN OBSERVATION INSTRUMENT

An Introduction

The Social Science Observation Record provides classroom teachers and instructional theorists with a model of classroom discussion. Because the Social Science Observation Record can also be employed as a systematic observation system, trained coders can collect data enabling teachers and theorists to analyze class discussions as they occur and thus to compare what happens in practice with what was or might have been originally anticipated. Prior discussions of the Social Science Observation Record have stressed that classroom teachers need not have the services of a trained coder to think about and share ideas about class discussions in terms of the Social Science Observation Record. This chapter details how a reliable coder and trained observer collects, organizes, and interprets data which can be used by classroom teachers.

When the reader has comprehended this chapter he should be able (if provided with data by a trained coder) to organize and interpret data using the SSOR as his frame of reference. Given these skills and a knowledge of his objectives the teacher can proceed to analyze what occurred, and, should he desire, he can plan strategies for modifying his behaviors, the behaviors of his students, or both. To achieve the greatest benefit from this chapter the reader is advised to follow three disciplines:

....Think about the categories by numerical designation rather than by label. This is necessary because data are collected and manipulated using numbers as referents for the seventeen categories of the SSOR.

....Master each section of this chapter as it appears before proceeding to the next section. The sections are cumulative and as the reader moves from section to section a knowledge of previous sections of the presentation is presumed.

...Study carefully the figures as these are referred to in the text of the presentation. These figures stress tools and concepts necessary for the organization and interpretation of SSOR data.

If these three disciplines are followed, the reader should acquire understandings and skills by which he can organize, interpret, and analyze SSOR data.

Collection of SSOR Data

A coder trained to use the SSOR records what occurs during a class discussion on a three-second-interval basis. He assigns the appropriate number which designates the category of behavior he "hears" occurring during each interval. He can be pictured as mechanically listening to a discussion in segments of three-seconds duration and encoding each segment. He codes the events in each interval, assigning to it a number representing one of the seventeen categories he has trained himself to recognize. Because he records on paper as he mentally codes, what occurs can be retrieved, organized and interpreted. Because he records what he hears every three seconds, an adequate sample of data concerning teacher and student behaviors is collected for the reconstruction and analysis of class discussion.

For each minute that an analyst codes behavior he records at least 20 times. At the end of five minutes he has recorded at least 100 times, at the end of ten minutes 200 times. If he "hears" two different behaviors occurring within the same three second interval, he codes both in the order in which he hears them. This is done to preserve as complete a record as possible of what is occurring. A trained coder codes about 20 times for each minute of observation. During periods of rapid interchange among students and the teacher he will code more frequently.

How a trained coder collects data and the way in which discussions may be reconstructed using his data can be clarified by reconstructing the data presented in Figure 10. As the analyst begins to observe, code, and record the discussion, he records a 16, the designation for silence, indicating that he did not, as an analyst, hear what occurred earlier in the discussion. As he listens to the class discussion, he codes teacher or student behaviors for each three-second interval. In the example the teacher begins by structuring a question using commentary and interrogative statements (categories 7 and 9). Students respond to the teacher's question by expressing empirical statements (category 2). The teacher responds by using a second interrogative statement (category 9) which results in a shift in student behavior to the expression of interpretive statements (category 3). In response to and following the two instances of interpretive statements, a student conveys his feelings through an emotive statement (category 15). When the teacher (or a student) reacts with a question (category 9), a student expresses a preference (category 11) which leads to some confusion, perhaps laughter or hostility (category 17). The teacher reinforces the preferential statement using a confirming statement (category 10). This leads students to express a criterial statement (category 13) and a consequential statement (category 12). When the discussion is concluded or the coder decides to terminate his coding (in effect, to stop "hearing"), he indicates this conventionally by coding silence (category 16).

Ten minutes of coded data are presented in Figures 11 and 12. The data in Figure 11 were collected during a social studies discussion conducted in an in-service workshop in St. Paul, Minnesota. The data in Figure 12

Category coded by trained coder	Explanation of events occurring in the class discussion during each three second interval
16	By convention, interaction analysis data are initiated by coding silence.
7	The teacher employs a commentary statement to initiate the discussion.
7	The teacher begins to structure a question to which he wants students to respond.
7	The teacher continues to structure the context of his question.
7	The teacher finishes structuring the context within which he wishes students to respond.
9	The teacher uses a question intended to elicit student statements.
2	A student expresses an empirical statement.
2	The same student or a second student expresses what he believes to be a factual statement.
9	The teacher asks a second question.
3	A student expresses an interpretive statement.
3	The same student or a second student expresses an interpretive statement.
15	A student conveys personal feelings in the form of an emotive statement.
9	The teacher (or a student) asks a question.
11	A student expresses a value rating in the form of a preferential statement.
17	A period of noise recorded as confusion occurs.
10	The teacher encourages the student to continue by confirming prior behavior.
13	A student states the basis for his reasoning by expressing a criterial statement.
12	The same student or another student anticipates an effect using a consequential statement.
...	The discussion continues.
16	By convention, coders indicate the end of a discussion or terminate their collection of data by using the number that designates silence.

Figure 10. The reconstruction of SSOR data.

BEST COPY AVAILABLE

Code No. ____________

SSOR DATA COLLECTION FORM

(Casteel and Stahl)

16

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
7	9	9	16	2	7	2	16	7	7	16						
7	9	16	16	10	7	10	16	2	7							
7	2	2	16	7	7	7	16	2	7							
7	3	2	16	9	7	9	16	7	7							
7	10	9	16	2	7	9	16	7	7							
7	7	2	16	9	16	9	16	7	7							
7	7	2	16	7	16	16	16	9	8							
9	7	10	16	2	7	16	16	2	7							
9	7	7	16	9	9	16	7	9	8							
9	7	9	16	2	9	16	9	9	7							
4	9	7	16	7	9	16	2	4	10							
10	2	7	16	7	3	16	9	10	7							
9	7	7	16	7	7	16	9	7	16							
9	7	9	16	7	16	16	10	7	16							
4	7	9	16	7	16	16	7	7	16							
9	7	9	16	7	7	16	7	7	16							
4	7	7	16	7	7	16	2	7	10							
10	16	16	16	7	9	16	10	7	9							
7	9	7	16	9	9	16	2	7	9							
9	9	16	9	2	9	16	10	7	9							

OBSERVED Math teacher DATE 5/16/73 TOPIC Mean, Median, Mode

OBSERVER trained coder PLACE Gainesville, Fla TIME 10 min

OTHER micro teaching session no. 1, 7th grade student

Figure 11. SSOR data for a mathematics teacher.

BEST COPY AVAILABLE

Code No. ________

SSOR DATA COLLECTION FORM

(Casteel and Stahl)

16

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
7	4	3	7	7	12	14	12	10	6	16						
7	9	9	9	6	7	14	9	7	7							
7	5	9	16	5	7	14	10	7	6							
7	5	3	3	5	7	5	9	9	15							
7	5	9	3	5	9	3	9	5	9							
7	3	2	3	5	9	10	8	5	15							
7	3	3	13	5	9	3	7	3	3							
7	4	3	12	11	7	5	7	3	2							
7	4	4	13	6	9	13	7	5	10							
7	3	4	3	9	9	12	7	5	3							
7	5	2	3	11	14	5	7	5	3							
9	4	3	11	11	3	12	7	3	5							
7	4	3	3	13	3	13	7	6	14							
7	4	11	3	13	13	13	10	3	5							
9	4	9	3	5	12	3	3	3	5							
9	4	9	3	5	3	3	3	7	6							
16	4	9	13	3	3	3	3	7	5							
16	3	9	13	15	3	14	3	7	5							
4	3	9	3	13	5	14	3	7	7							
4	3	7	10	3	5	3	13	7	7							

OBSERVED Social Studies teacher DATE 3/13/73 TOPIC social issue

OBSERVER trained coder PLACE St. Paul, Minn. TIME 10 min

OTHER micro simulation, peer group teaching

Figure 12: SSOR data for a social studies teacher.

were collected during a discussion in a mathematics classroom at Howard Bishop Middle School, Gainesville, Florida. Examination of the data reveals that the social studies teacher elicited a large number of man-centered student behaviors (categories 11 through 15) occurring in conjunction with definitions and interpretations (categories 4 and 3 respectively). An analysis of the data for the mathematics teacher differs in that a greater number of student subject-centered behaviors (categories 1 through 5) were observed with special emphasis on empirical data (category 2). The classroom teacher can look at such data-collection sheets and determine: (1) if the coded behaviors are the behaviors he expected, (2) if the behaviors were the results of his own behaviors, and (3) if so, to what degree he affected the behaviors.

If interest is limited to a general picture of what occurred, the data sheet alone is useful. If one instead wishes to analyze a class discussion, the data sheet is not sufficient. For purposes of analysis, the data must be transformed into some kind of interpretive display. This transformation is accomplished by transferring raw data into a matrix. This transformation is initiated by pairing.

Pairing and Tallying Data

When SSOR data are collected, any category as coded may be followed by any of the categories in the system. For example, an empirical statement, category 2, may be followed by category 1, or 2, or 3, or 4, or 5, or 6, or 7, or 8, or 9, or 10, or 11, or 12, or 13, or 14, or 15, or 16, or 17. An interrogative statement, category 9, may be followed by category 1, or 2, or 3, or 4, or 5, or 6, or 7, or 8, or 9, or 10, or 11, or 12, or 13, or 14, or 15, or 16, or 17. This possibility exists for all seventeen categories. Because each of the categories may be followed by itself or any other category, 289 sequences of categories are possible. When the seventeen categories of the SSOR are used to construct a matrix of ordered pairs, the 289 sequenced pairs (17 X 17 = 289) are referred to as "cells of inquiry" (see Figure 13). SSOR data as collected (see Figures 11 and 12) are converted to this matrix for interpretation and analysis.

The first task in converting data to the matrix is called "pairing." Data representing forty-five seconds contain fifteen pairs (the 45 seconds observed divided by the 3-second interval for coding equals 15 pairs). Data for forty-five seconds are paired here.

```
                 16
                  7 ] First pair
Second pair  [    7
                  7 ] Third pair
Fourth pair  [    7
                  7 ] Fifth pair
             [    2
                  2 ]
             [    9
                  3 ]
             [    3
                 15 ]
             [    9
                 11 ]
             [   10
                 10 ] Fifteenth pair
                 16
```

The first pair is the 16-7 pair; the second is the 7-7 pair. The second number in the 16-7 pair becomes the first 7 in the 7-7 pair. The fifth pair is a 7-2 pair and the sixth pair is a 2-2 pair. Except for the number with which coding begins and the number with which coding ends, each number used to record data is part of two pairs. This procedure is followed until all data collected have been tabulated into pairs and transfered into the matrix. Pairing "stores" the interaction between categories of the SSOR and serves as a basis for tallying data and entering it into the SSOR matrix.

SOCIAL SCIENCE OBSERVATION RECORD (SSOR) MATRIX

J. Doyle Casteel and Robert J. Stahl (c. 1973)
College of Education, University of Florida

	1. Topical	2. Empirical	3. Interpretive	4. Defining	5. Clarifying	6. Infirming	7. Commentary	8. Dissonant	9. Interrogative	10. Confirming	11. Preferential	12. Consequential	13. Criterial	14. Imperative	15. Emotive	16. Silence	17. Confusion
1. Topical																	
2. Empirical																	
3. Interpretive																	
4. Defining																	
5. Clarifying																	
6. Infirming																	
7. Commentary																	
8. Dissonant																	
9. Interrogative																	
10. Confirming																	
11. Preferential																	
12. Consequential																	
13. Criterial																	
14. Imperative																	
15. Emotive																	
16. Silence																	
17. Confusion																	

Figure 13: The 289 cells of the SSOR.

Data are first tallied into the SSOR scatter diagram (Figure 14). After data have been collected and paired, each pair can be assigned to one of the 289 cells of the SSOR matrix. For the forty-five seconds paired above, the first pair is 16-7 and belongs in the 16-7 cell. To determine where to place a tally to convert this pair into the SSOR matrix, read down the left-hand side of the scatter diagram to the number 16; this locates the first number in the 16-7 pair. Next read left to right and locate column 7; this locates the second number in the pair. Where the two numbers intersect the 16-7 pair is tallied. This intersection of categories 16 and 7 is designated by the letter "a" in Figure 14. When referring to matrix data, this is read as category 16 followed by category 7. The second, third, and fourth pairs are 7-7 pairs. To tally these pairs read down the left of the scatter diagram to category 7 and then left to right to column 7. Where the two 7's intersect is the appropriate cell in which to tally the three 7-7 pairs. In Figure 14, this cell is designated by the letter "b." The fifth pair is a 7-2 pair. To tally read down the left of the scatter diagram to the number 7 and then to the right to column 2. In Figure 14, this cell is designated by the letter "c." This procedure is continued until all data are paired and tallied in the scatter diagram. The number of tallies for each cell are then counted and the total transferred to the equivalent cell in the SSOR matrix (see Figure 15).

As pairing was initially discussed, an observer codes an event and records a numerical designation for the behavior which occurred. The subsequent behavior coded can designate any category in the SSOR including itself. Frequently the same behavior will last for an extended period of time. In the SSOR system, the seventeen intersections or cells in the matrix which record and describe these extended behaviors are called "extended state cells" (see Figure 16).

Code No. ________

SSOR SCATTER DIAGRAM

J. Doyle Casteel and Robert J. Stahl (c. 1972)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
1																		0
2		II							I									3
3			I												I			2
4																		0
5																		0
6																		0
7		I (c)					III (b)											4
8																		0
9			I								I							2
10																I		1
11										I								1
12																		0
13																		0
14																		0
15									I									1
16							I (a)											1
17																		0
Total	0	3	2	0	0	0	4	0	2	1	1	0	0	0	1	1	0	15

Observed: ______________ Date: ___/___/ 73 ______ Topic: ______________

Figure 14: SSOR scatter diagram with cells for three pairs designated by the letters a, b, and c.

SOCIAL SCIENCE OBSERVATION RECORD (SSOR) MATRIX

Code________________

J. Doyle Casteel and Robert J. Stahl (c. 1973)
College of Education, University of Florida

	1. Topical	2. Empirical	3. Interpretive	4. Defining	5. Clarifying	6. Infirming	7. Commentary	8. Dissonant	9. Interrogative	10. Confirming	11. Preferential	12. Consequential	13. Criterial	14. Imperative	15. Emotive	16. Silence	17. Confusion	Total
1. Topical																		0
2. Empirical		2							1									3
3. Interpretive			1												1			2
4. Defining																		0
5. Clarifying																		0
6. Infirming																		0
7. Commentary		1					3											4
8. Dissonant																		0
9. Interrogative			1								1							2
10. Confirming																1		1
11. Preferential										1								1
12. Consequential																		0
13. Criterial																		0
14. Imperative																		0
15. Emotive									1									1
16. Silence							1											1
17. Confusion																		0
Totals (No.)	0	3	2	0	0	0	4	0	2	1	1	0	0	0	1	1	0	15
Totals (%)	0	20.0	13.4	0	0	0	26.7	0	13.4	6.7	6.7	0	0	0	6.7	6.7	0	100.5

REALM TOTALS / % I 5/33.4% II 7/46.8% III 2/13.4% IV 1/6.7%

Figure 15: Data transferred into the SSOR matrix.

SOCIAL SCIENCE OBSERVATION RECORD (SSOR) MATRIX

J. Doyle Casteel and Robert J. Stahl (c. 1973)
College of Education, University of Florida

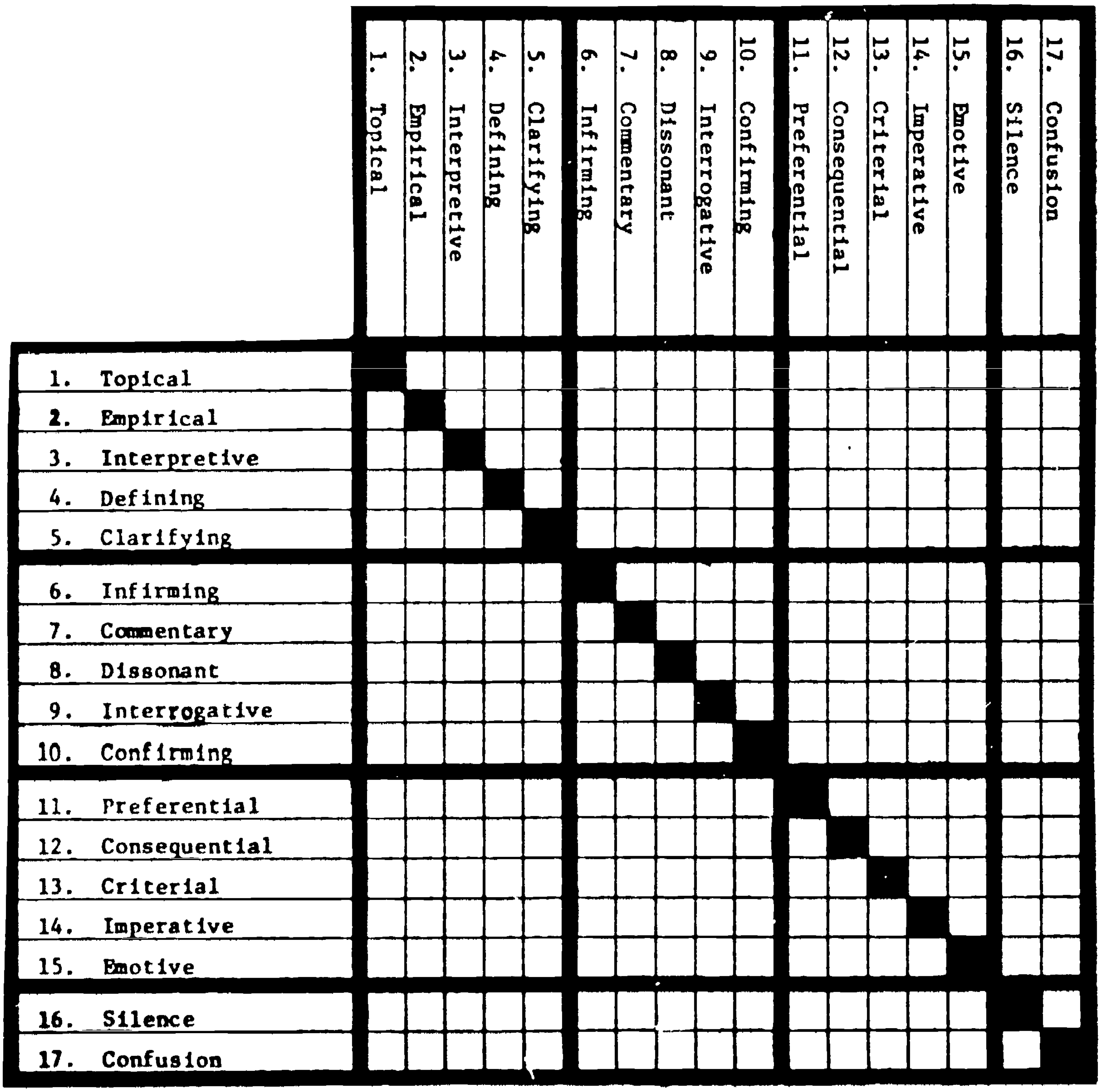

Figure 16: The seventeen extended state cells of the SSOR matrix

For instance, a topical statement (category 1) followed by another topical statement (category 1) yields a 1-1 pair, a category 2 followed by a second category 2 yields a 2-2 pair, and so forth to the 17-17 pair. This pairing sequence holds true for all seventeen categories and forms the seventeen "extended state cells" in the SSOR matrix. Extended state cells record those pairs of data where the first number and the last number of the pairs are the same (1-1, 2-2,.........17-17). These pairs or cells indicate that the particular behavior observed lasted an "extended" or six-second period of time and was recorded as such.

One of the major purposes for transferring the raw data into a matrix is to enhance the descriptive and analytical potential of the SSOR matrix. In addition to setting up the data for simple tabulation of frequencies and percentages for each category and realm, the matrix system allows for an examination of the sequence of behaviors within the total interaction.

The matrix format generates data which would not be available using frequency counts alone. Because data are organized into a matrix and because of the characteristics of the SSOR system, the resulting matrix data can be interpreted in a number of ways. Some of these are described briefly below.

(a) Total Data Count. Assuming the trained coder codes and records approximately every three seconds and every time there is a change in category, there are twenty codings per minute. More frequent codings than this indicate a series of rapid behavior changes from one category to another.

(b) Total Number of Cells Used. With 289 cells existing within the SSOR matrix, the use of a number of cells indicates that a variety of verbal patterns occurred. When a variety of different cells are used, the frequency of use for each cell is often limited. An increase in the total number of cells used may suggest a flexible teacher, a random discussion, or a discussion in which a variety of different verbal behaviors occurred. The authors have not, as yet, explored these three or other possibilities for interpreting data in reference to total cell count.

(c) Extended State Cells. The SSOR matrix not only indicates which of the seventeen extended state cells were coded, but reports the frequency totals and percentages of these behaviors as well.

(d) Category Usage. While seventeen categories exist in the system, rarely are all seventeen observed during a class discussion. The matrix identifies not only those categories which are used; it also indicates those not used. Both ways of analyzing data provide valuable insights as to what occurred. Category data are also provided in terms of the frequency of occurrence for each category as well as in terms of the percentage for each category of behavior observed.

(e) Realm Usage. The four realms of the SSOR are identified and separated in the matrix by heavy lines making the examination of realm data more convenient. The data in Realms I, II and III are sub-divided into five parts (categories) each while the fourth realm has two sub-divisions (categories). Realm data allow for interpretation as to whether student or teacher verbal behavior predominated, and as to whether student behavior stressed understanding, value clarification, or both.

(f) Submatrices. The 289 cells in the SSOR matrix lend themselves to logical groupings or combinations of cells called submatrices. Each of the twelve submatrices has unique attributes. An understanding of each in terms of the entire matrix adds still another interpretive dimension.

These aspects of the SSOR matrix may be used to analyze SSOR data. Should one want to analyze and interpret collected SSOR data in order to describe classroom discussion behavior he will need to understand the kind of data provided by the matrix.

An explanation of a sample SSOR matrix may serve to enhance the reader's understanding of how the data in matrix form can be interpreted. Figure 17 contains data for approximately fifteen minutes of discussion collected under micro-teaching conditions. A pre-service social studies teacher was told to lead a discussion on the importance of sexual equality. Her students were five seventh graders studying world geography. The teacher was fulfilling the intern-

SOCIAL SCIENCE OBSERVATION RECORD (SSOR) MATRIX

J. Doyle Casteel and Robert J. Stahl (c.1973)
College of Education, University of Florida

Code__________

BEST COPY AVAILABLE

	1. Topical	2. Empirical	3. Interpretive	4. Defining	5. Clarifying	6. Infirming	7. Commentary	8. Dissonant	9. Interrogative	10. Confirming	11. Preferential	12. Consequential	13. Criterial	14. Imperative	15. Emotive	16. Silence	17. Confusion	Total
1. Topical																		0
2. Empirical		11	5		3		2	2	5	4	1		3	2		1	3	42
3. Interpretive		4	2		1		4	1	1	2		1	1	1				18
4. Defining				3	1						1							5
5. Clarifying		5		1	2	1	1	1	1		1	1	1		1			16
6. Infirming					1		3		1									5
7. Commentary		5	2		1	1	74	1	11	3	2				1		2	103
8. Dissonant		2	1		1					1								5
9. Interrogative		10	4				4		17				1	1	1	2	1	42
10. Confirming		2			2	2	9		3									18
11. Preferential			1							3	1		1	1				7
12. Consequential			1		1		2			1			1					6
13. Criterial		1			1						1	2		5				10
14. Imperative		2	1		1		1		1	2		1	1	2				12
15. Emotive							1					1	1					3
16. Silence							2		1									3
17. Confusion			1		1	1			1	2								6
Totals (No.)	0	42	18	5	16	5	103	5	42	18	7	6	10	12	3	3	6	301
Totals (%)	0	14	6	2	5	2	34	2	14	6	2	2	3	4	1	1	2	100

REALM TOTALS/% I 81 / 27 % II 173 / 58 % III / 12 % IV /3 %

Total Count 301 Cells Reached (289) 103 Categories Used (17) 16

Extended State Cells (17) 8

Figure 17: Fifteen minutes of SSOR data for a social studies teacher.

ship assignment necessary for certification. The data in Figure 17 are interpreted below.

In the example, there were 301 recordings of which 119 were student verbal behaviors in either the Subject-Centered or the Man-Centered Realms; 173 were teacher-centered behaviors. Thirty-nine percent of the behavior coded was student verbal behavior in Realms I and III of the SSOR (27% + 12% respectively). Consequently, this discussion is called a directed discussion. Had student verbal behavior been one percent more, the discussion would have been a guided discussion.

The teacher helped her students to identify and pool empirical data (category 2), to interpret the meaning of the data (category 3), and to clarify and elaborate on their ideas and opinions (category 5). She probably could have helped them spend more time in defining (category 4) the term sexual inequality. Her students did not emit any instances of topical statements (category 1), thus denoting that at no time during the discussion did they themselves identify the focus of their discussion.

The teacher helped students to clarify their values with regard to sexual inequality. While students used all five of the categories in the Man-Centered Realm, they tended to emphasize criterial statements (category 13) and imperative statements (category 14). Students stated the basis for their reasoning and identified alternative policies more often than they stated their preferences and feelings.

The teacher employed and made regular use of three categories of statements in the Teacher-Centered Realm. Her behavior suggests that she frequently needed to provide directions, give new information, consolidate past events, and set the stage for her questions. All the above-mentioned teacher behaviors are coded as category 7. She frequently asked questions (category 9) (some which took longer than three seconds--cell 9-9) and indicated her acceptance of student behaviors (category 10). The behaviors recorded in Realms I and III strongly suggest that her behaviors in Realm II were facilitative. The teacher relied heavily on commentary and interrogative statements to guide the discussion. Both criticism (category 6) and reinforcement (category 10) were used, but neither appears to have been employed excessively.

In the Non-Verbal Realm, silence (category 16) was seldom recorded. When silence did occur it did not last more than three seconds, leaving the 16-16 cell empty. In no instance did student verbal behavior in either the Subject-Centered or Man-Centered Realms follow silence. Of the three occurrences coded, the teacher each time initiated the verbal behaviors with which silence was broken.

Of the 289 cells in the matrix, 103 cells (36%) were reached. This indicates that the teacher and her students used a number of language patterns, incorporating sixteen of the seventeen categories during the discussion.

In terms of looking at specific cells or verbal patterns, the 13-14 pair occurred five times while the 14-13 pair only occurred once. While the teacher reacted consistently to student imperative statements, she did not respond to student criterial statements. The 5-2 cell count indicates that five times during the discussion students sought to clarify previous statements by referring to empiric data to assist them. The one tally in the 11-11 cell and the absence of tallies in the 15-15 cell indicate that students did not objectify their preferences and feelings for any extended period of time.

The teacher reinforced student behaviors in both the Subject-Centered Realm (the 2-10 and 3-10 cells) and in the Man-Centered Realm (the 11-10 and 14-10 cells). It may be inferred that the teacher wanted students to express themselves in both the Subject-Centered and Man-Centered Realms.

The teacher used commentary statements (category 7) to structure her questions (category 9). Evidence supporting this statement is found in the 7-9 cell. Eleven different times she used commentary statements to set up her questions. Less than half of her questions, however, were longer than three seconds in length and tended to be followed by student subject-centered behaviors. The teacher seldom answered her own questions (the 9-7 cell).

If the teacher's goal was to lead a directed discussion in which students developed understanding and clarified their values with regard to sexual inequality, the data in the matrix suggest that she was successful. If the teacher wished to lead a guided discussion, she might have met her goal had she responded less frequently to student interpretive, preferential, and imperative statements and waited for students to determine ways of expressing and elaborating their ideas and values, or had she not responded to her own questions. Of these two it would appear that she may have wanted to increase use of silence inasmuch as she already used the 9-16 cell but apparently did not wait long enough for students to speak after brief periods of silence.

The SSOR matrix is usually interpreted in the light of known teacher objectives. The foregoing discussion does demonstrate how a teacher can discover intuitive purposes guiding his behaviors by obtaining SSOR data, tallying his data in a scatter diagram, building a matrix, and proceeding to interpret his behavior. Whether he carries out these tasks by himself or these are done by a trained observer, he possesses a mirror by which he can groom himself as a discussion leader.

Submatrices of the SSOR Matrix

If the classroom teacher or observer wants to increase his powers of analysis and interpretation further, he needs to understand different sections of the SSOR matrix called submatrices. Submatrices are small groups of related cells which identify special kinds of interrelations useful for matrix analysis.

The SSOR matrix contains twelve logical submatrices. These are designated by capital letters from Submatrix A through Submatrix L. The location of these twelve submatrices within the SSOR matrix are shown in Figure 18.

Submatrix A. This submatrix contains twenty-five cells or verbal patterns of student subject-centered statements that occur when a discussion remains in the Subject-Centered Realm for more than one three-second interval. The use of these cells indicates that students are interacting with one another, and as the use of this submatrix increases, discussions tend to be either guided discussions or open discussions. Extensive use of this submatrix and of a number of subject-centered cells suggests that students can operate independently in the Subject-Centered Realm. For the teacher whose objectives include student understanding, Submatrix A constitutes a major focus for planning, guiding, and eventually analyzing his class discussions.

Submatrix B. This submatrix provides twenty-five cells for describing how the teacher reacts to student subject-centered behavior. If one wants to know what subject-centered behaviors were infirmed or confirmed, he finds the necessary data in Submatrix B. To determine what categories of student subject-centered behaviors led the teacher to make commentary statements or to use interrogative statements, one examines Submatrix B. When teacher-centered behaviors are heavily concentrated in Submatrix B, the following inferences can be drawn: the

SOCIAL SCIENCE OBSERVATION RECORD (SSOR) MATRIX

J. Doyle Casteel and Robert J. Stahl (c. 1973)
College of Education, University of Florida

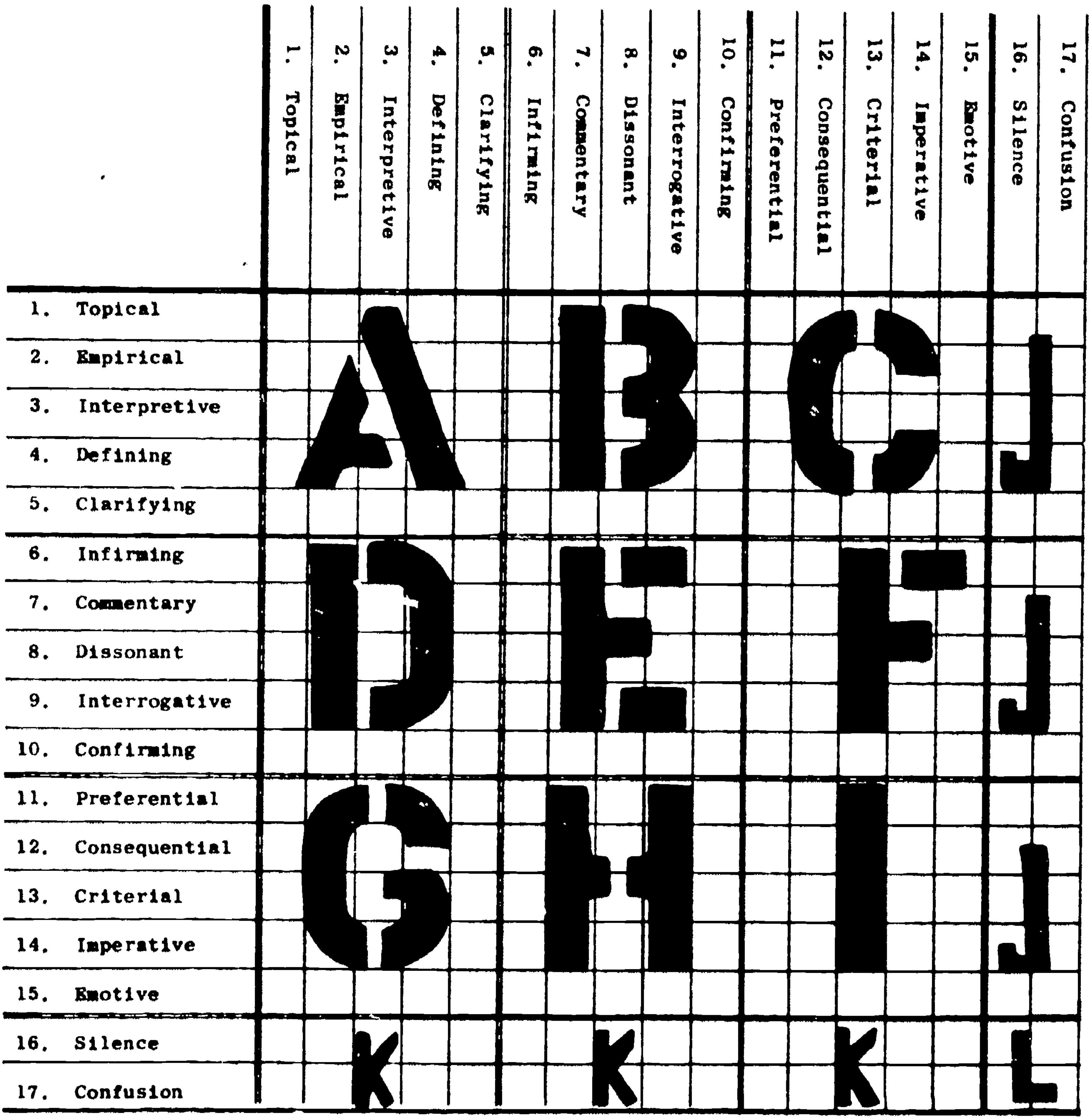

Figure 18: The 12 submatrices within the SSOR matrix.

discussion tends to be a marginally direct or lecture discussion in which students are reviewing or drilling; the teacher is not structuring a context for his questions and is probably asking low level knowledge questions; and the teacher is asking a number of questions followed by short subject-centered student statements during each minute of discussion.

Submatrix C. This submatrix contains twenty-five cells for recording student man-centered statements which follow but are contiguous with subject-centered statements. Extensive use of this submatrix indicates that students know how to move from the Subject-Centered to the Man-Centered Realm and perceive that such behaviors are a legitimate aspect of inquiry which occur in conjunction with the development of cognitive understanding. If a teacher encounters students who perceive value clarification as inappropriate classroom behavior or if he is attempting to increase his students' abilities and willingness to affirm and share values, increments in this submatrix would suggest progress toward these goals.

Submatrix D. This submatrix contains twenty-five cells for recording student subject-centered statements to teacher-centered behaviors. How students respond to criticisms, to reinforcement, to comments, to requests for clarification, and to questions by expressing themselves in subject-centered categories can be detected by analyzing this submatrix. If the majority of student subject-centered behavior collects in this submatrix, the discussion is probably a marginally direct or lecture-demonstration discussion. If student behaviors cluster here and teacher verbal behaviors are clustered in Submatrix B, one infers that students are reciting or reviewing, or that the teacher is raising lower level cognitive questions, and that student responses are congruent with the questions posed by the teacher. One could speculate with some reasonableness that students are engaging in designating and verbalizing without due regard for whether or not they understand what they are saying.

Submatrix E. This submatrix contains twenty-five cells or verbal patterns for recording teacher-centered statements following other teacher-centered statements. All teacher-centered statements lasting at least six seconds in duration are recorded in this submatrix. If the teacher criticizes student verbal behaviors and proceeds to follow his criticism with other kinds of statements, the teacher's statements, with the exception of his first statement, will be recorded in Submatrix E. The teacher wanting to examine the manner in which he structures his questions would expect to investigate his use of the 7-7, 7-9 cells. Submatrix E provides the teacher with data as to the kinds of verbal behaviors he engages in during the class discussion. It provides information relative to the kinds of statements that he uses to guide and manage the class discussion as well as data about the frequency of these statements. Students can also use teacher-centered categories. Since students' use of the teacher-centered categories tends to be quite minimal, student use of these categories is seldom reflected in Submatrix E.

Submatrix F. This submatrix contains twenty-five cells recording how students react to teacher-centered behaviors when they respond with man-centered statements. When the teacher deliberately uses his influence to elicit man-centered behaviors, this is indicated by codings in Submatrix F. Student man-centered statements which coincidentally follow teacher talk are also recorded here. If the only man-centered statements coded are located in Submatrix F, the data may be interpreted to mean that students did not express or attempt to clarify their values except at the direct request of the teacher, and these statements were extremely brief. If Submatrix F is used about as often as Submatrix H, it suggests that the teacher quickly responds to student man-centered behavior. If Submatrix F data are accompanied by data in Submatrix I, this is interpreted as indicating that the teacher used his influence to elicit man-centered statements and that students continued to engage in value clarification behaviors as a result of the teacher's influence.

Submatrix G. Submatrix G contains twenty-five cells for recording data when students voluntarily move from the Man-Centered to the Subject-Centered Realm. This can be interpreted in different ways. If Submatrix G is used almost as often as Submatrices C and F and if Submatrix I contains little or no data, students are probably responding briefly to teacher requests for value clarification statements and returning immediately to the Subject-Centered Realm. If Submatrix G is used in conjunction with Submatrix I, students are probably returning to the Subject-Centered Realm to collect, share, and interpret more data; to clarify their man-centered statements; to define a troublesome concept; or to reestablish the focus of discussion. The same data may indicate that students are analyzing how they engaged in value clarification.

Submatrix H. Submatrix H contains twenty-five cells of verbal patterns describing how the teacher responds to student behaviors in the Man-Centered Realm. If this submatrix is used in conjunction with Submatrix F and if little use is made of Submatrix I, the data are interpreted as meaning that the teacher, who wants to secure value clarification statements from his students, is reacting too quickly to student responses in the Man-Centered Realm.

Submatrix I. Like Submatrices A and E, Submatrix I data indicate in-depth realm usage. When man-centered student statements tend to gather in Submatrix I, this suggests that students are allowed to express their values, their ideals, their grounds, and their feelings for long periods of time without teacher intervention. Behaviors in Submatrix I are interpreted as meaning that value clarification of a prolonged nature is occurring. For the teacher whose objectives include value clarification, Submatrix I becomes a major focus for the analysis of class discussions he plans for and teaches toward.

Submatrix J. This submatrix contains thirty cells. When silence or confusion follow subject-centered, teacher-centered, or man-centered statements, these behaviors appear in Submatrix J. Submatrix J is further divided into three five-celled segments. The first segment of J (J_1) describes the silence occurring after student behavior in the Subject-Centered Realm. The second segment (J_2) records silence following teacher-centered behaviors. The third segment of J (J_3) records the silence following student man-centered behaviors. If segment 2 is used but segments 1 and 3 are not, the data suggest that the teacher believes his questions and ideas are worth consideration but those expressed by his students are not of equal value. This description of Submatrix J places deliberate stress on the silence or "wait time" component of segments 1, 2, and 3 of the submatrix.

Submatrix K. Submatrix K also contains thirty cells. This submatrix records what student or teacher behaviors follow silence and confusion. It, too, is sub-divided into three five-celled segments. The first segment (K_1) indicates what, if any, student subject-centered statements follow silence. The second segment (K_2) is used if silence is followed by teacher-centered behavior. The third segment (K_3) is used if silence is followed by student man-centered statements. The three segments of Submatrix K were deliberately stressed here to alert the reader to the importance of who (teacher or student) "breaks" the waiting period.

Submatrix L. This submatrix contains only four cells indicating the interactional possibilities of silence and confusion. Silence can be interpreted as either an indication that students are formulating ideas or that they do not know how to respond to the situation. If silence translates to confusion, whether this be descriptive noise or a student speaking so quietly he cannot be understood, the latter interpretation can be justified. Periods of confusion can be normal moments of excitement or adjustment or they can indicate that the teacher needs to step in and restructure the discussion. If confusion translates to silence, the first interpretation can be argued; if confusion remains stable, the second interpretation becomes more probable. This submatrix is important if one wishes to study extended periods of "wait time" (Rowe, 1973).

Knowledge of the twelve submatrices increases the ability of the teacher to plan, teach, and analyze class discussions. An idea of the data generated by the SSOR matrix may be grasped by studying the matrix summary sheet in Figure 19. The teacher can cast objectives, monitor class discussions, and analyze data using different dimensions of the SSOR.

Summary

This chapter has presented information essential for the collection, organization, and interpretation of SSOR data. Data are collected on a three-second-interval basis and may be used to reconstruct class discussions. Once data have been collected, categories used contiguously are arranged in

Category Totals (No.)																	
Category Totals (%)																	

Realm Totals/% I ____/____ % II ____/____ % III ____/____ % IV ___/___ %

Total Count ________ Cells Reached (289) ________ Categories Used (17) ________
Extended State Cells (17) ________ Ext. State Cell Freq. ________ 7-9 Count ________

Subm. Use (No.): A___ B___ C___ D___ E___ F___ G___ H___ I___ J___ K___ L___ TOTAL___
Subm. Use (%): A___ B___ C___ D___ E___ F___ G___ H___ I___ J___ K___ L___ TOTAL___
Subm. Use (Cells): A___ B___ C___ D___ E___ F___ G___ H___ I___ J___ K___ L___ TOTAL___

Figure 19. Data summary sheet for the SSOR.

ordered pairs and tallied in a scatter diagram. After the data have been tallied, the count for each cell of the 17 x 17 scatter diagram is entered into the equivalent cell of the SSOR matrix. When data have been transferred to the SSOR matrix, category totals and percentages, realm totals and percentages, extended state cell usage and percentages, and cell totals are computed. If the teacher's objectives are available, data may be used to examine the degree to which his objectives are met. (See Figure 20). If the teacher's objectives are not available, his intent and purposes may often be determined by studying the available data.

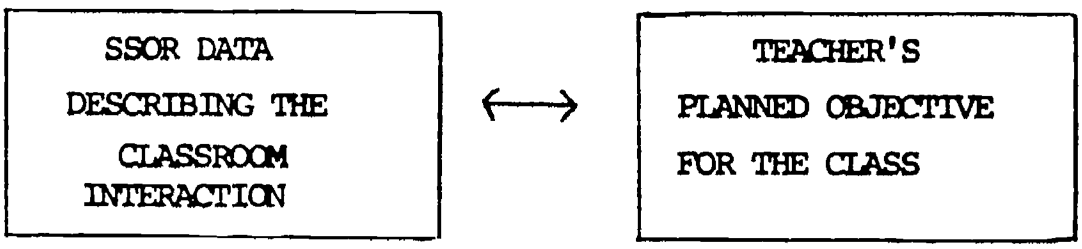

Figure 20: The relationship between SSOR data and teacher planning objectives.

One question which may be on the mind of the classroom teacher is: "Can knowledge of the SSOR help me to shape my behavior in directions I want to go?" Chapter IV will discuss pilot studies relevant to this question conducted by the authors using a sixteen-category earlier version of the SSOR system.

CHAPTER IV

REPORT ON PILOT STUDIES USING THE SSOR

An Introduction

In previous chapters, the Social Science Observation Record has been presented as a theoretical construct. How the SSOR is used as a model of class discussion for purposes of helping students engage in indicative (subject-centered) and value clarification (man-centered) inquiries has been explained. How the SSOR as a category observation system is used to collect, organize, and interpret data has been described. In this chapter, data from pilot studies are reported.

The studies reported here focus on four questions:

1) Observation instruments have been used by student-teaching supervisors to help student teachers modify their instructional behaviors (Amidon and Hough, 1967; Bondi, 1968, etc.). Can the SSOR be used to secure changes in the behavior of interns?

2) Observation instruments have been used to collect data with regard to what does not occur in the classroom during student-teacher interactions (Flanders, 1970; Ober et al, 1971). Can the SSOR be used to analyze the non-occurrence of expected events during interactions between students and teachers?

3) Interval observation systems are inclusive in that a limited number of categories are used as if these were the only dimensions of behavior worth observing (Medley and Mitzel, 1963). If the opinions of instructors of social studies methods courses are accepted as a criterion, do the categories of the SSOR capture data with regard to significant classroom events?

4) Interval observation systems have been used to collect descriptive data and provide feedback to teachers concerning their instructional behavior. This feedback has enabled teachers to teach subsequent lessons in which their intents as teachers (goals) and their behaviors as teachers (performance) are more congruent (Flanders, 1970; Rosenshine and Furst, 1973). Can teachers use SSOR feedback to align what does occur during teaching with what they expect to occur (plans)?

Tentative answers to these questions are to be provided in subsequent sections of this chapter.

BEST COPY AVAILABLE

Modifying Instructional Behavior

The shaping potential of the SSOR was initially explored during the winter quarter of 1972. The SSOR was used as a feedback device to modify the behavior of interns in directions desired by the researchers. This study was conducted:

> To determine if each of the sixteen categories of the SSOR identified at this time were categories used by teachers in social inquiry discussions.
>
> To increase the total number of categories used by each teacher participating in the study.
>
> To increase the total number of extended state cells used by each teacher participating in the study.
>
> To increase the number of cells of the SSOR matrix used by each teacher participating in the study.
>
> To determine how many of the 256 cells contained in the SSOR matrix could be obtained in practice.
>
> To secure the opinions of participating teachers as to the practical classroom effectiveness of feedback based on the SSOR.

Six subjects participated in the study. All participants were undergraduate students at the University of Florida and planned to become social studies teachers. All were engaged in student teaching in the same junior-senior high school (grades 7-12) at the time of the study. All had studied a value clarification handbook called "Verbal Strategies of Valuing" (Casteel, 1968), in preparation for the study. All participants were assured that their performance during the study would not in any way detract from their grades.

Participants were instructed to identify five students from one of their classes with whom they felt secure and who would be willing to participate in two-evening micro-teaching sessions. Participants were encouraged to prepare (for each micro-teaching session) a separate value sheet congruent with Louis

Raths's concept of the value clarification process (Raths et al, 1966). They were informed that each value sheet should be one that students had not seen prior to each micro-teaching session. Each value sheet was to be one related to the content being studied in the class they were currently teaching. Participants were also asked not to discuss the SSOR or their feedback sessions with others until the study had been concluded. To protect the anonymity of each participant, student teachers were assigned letter designations and are referred to as subject A through F.

The six subjects taught two micro-lessons one week apart. Two subjects taught a session for three consecutive nights. Subjects A, E, and F selected senior high school students. Subjects B, C, and D selected junior high school students.

All micro-lessons were taught in the evening at the P. K. Yonge Laboratory School. The room, room arrangement, and observers were the same for all sessions. Except for the first micro-lessons of subjects A and B, video tape replay was provided for each participant and her students.

Each student teacher taught for approximately ten minutes. As each teacher taught, an observer coded what was occurring during each three-second interval according to the categories of the SSOR. Between-observer reliability for the coders was computed according to Scott's formula (Scott, 1955); the coefficient of inter-observer agreement was computed at 0.86.

While each teacher watched the video tape replay of the lesson with students, SSOR data were organized into a matrix for interpretation. After video tape replay, each subject dismissed his students and returned for a discussion of SSOR data. During this conference each subject was helped to identify two or three categories which had not been used or had been used sparingly during the first lesson. Each subject was then encouraged to use new categories and to increase the frequency with which other categories were used during his second lesson. Each subject was helped to identify extended state cells not used during the first lesson and asked to use these cells during his second lesson. Each subject was also assisted in an analysis of Submatrices A and I of the SSOR matrix. SSOR feedback procedures, used following the first micro-teaching session, were repeated for the second session.

Sixteen of the sixteen categories in the SSOR system at the time of the study were coded as having occurred during the twelve lessons. This suggested that the categories included in the system do identify and describe classroom behavior used by teachers in social inquiry discussions.

Five of six subjects increased the total number of categories they used from the first to the second micro-teaching session (Figure 21). The sixth used fifteen categories in his first effort and used one less during his second session.

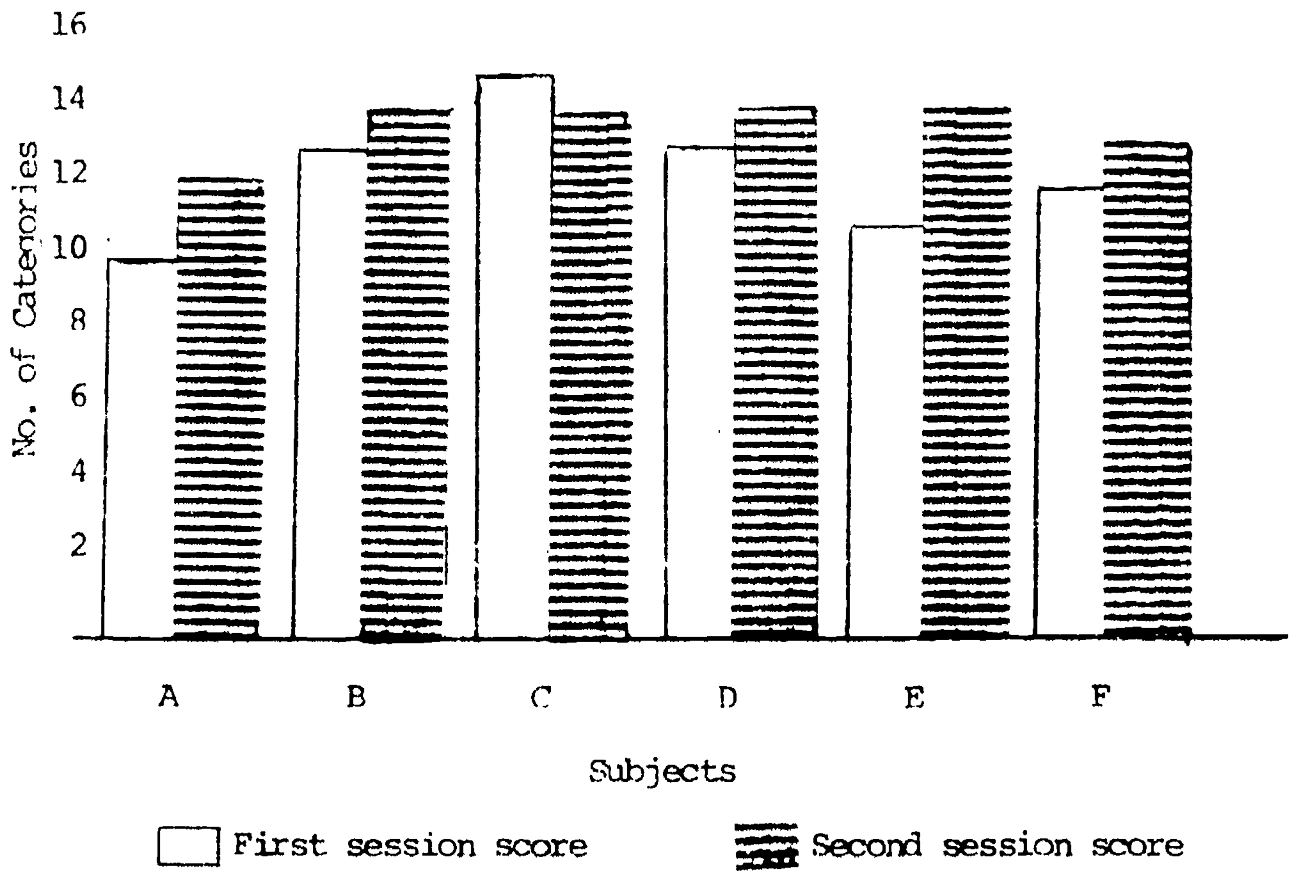

Figure 21. The number of categories used by the interns.

The average number of categories coded for the first lesson was 12.3 categories; for the second lesson the mean was 13.5 categories. There was an average gain of 1.2 categories (9.7%) for all subjects.
The results of this study suggest that categories of the SSOR can be used to describe instructional behaviors occurring during classroom interaction.

Five of six subjects increased the total number of extended state cells coded from the first to the second lesson (Figure 22).

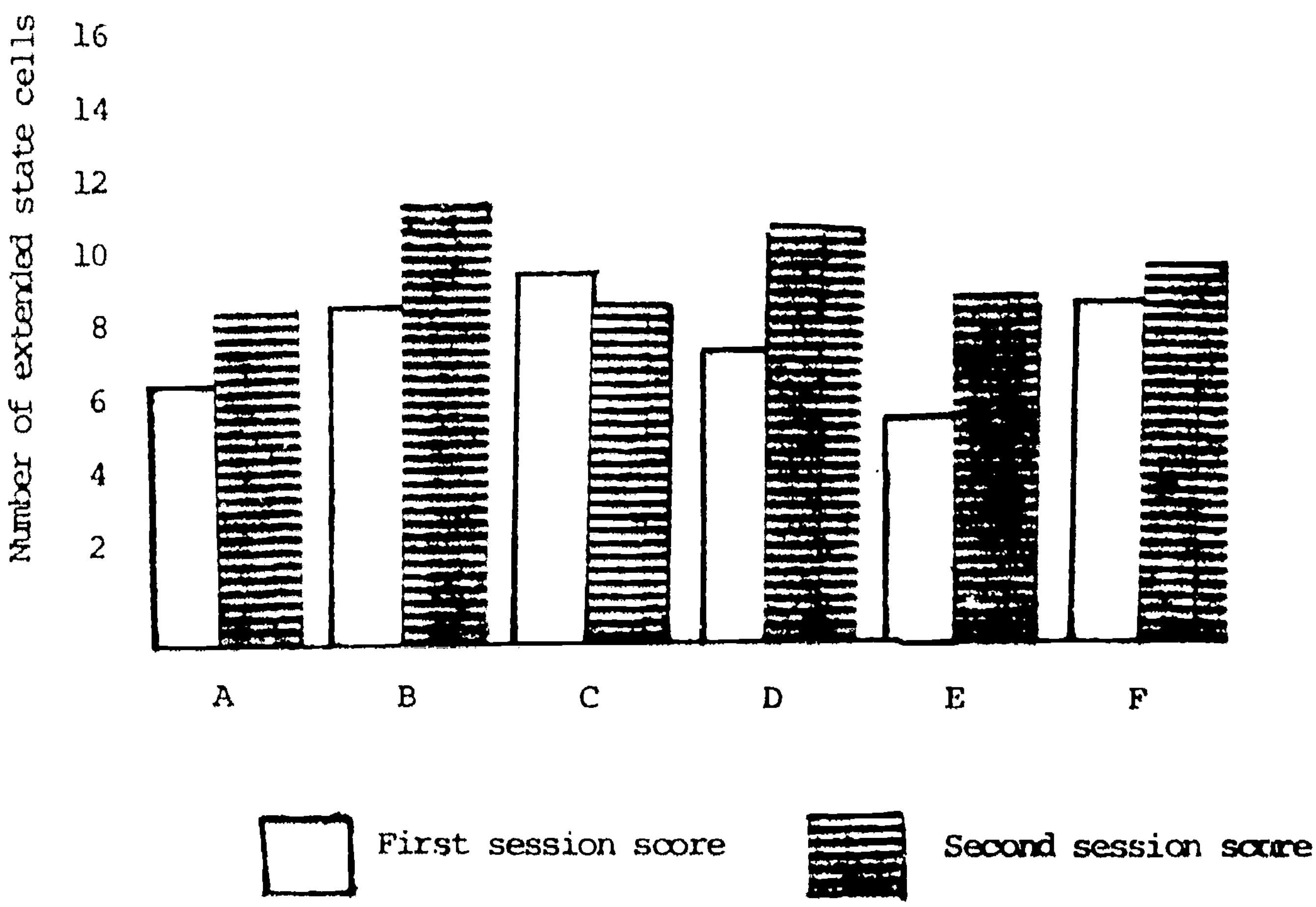

Figure 22. The number of extended state cells used by the interns.

Sixteen extended state cells existed in the SSOR system at the time of this study. For the first teacher the mean number of extended state cells used was 8.17. For the second, the mean number of extended state cells used was 10.0. This is a mean increase of 1.83 for extended state cells (22.4%). An increase in extended state cells may be associated with use of more categories for sustained periods of time.

All six subjects increased the total number of cells used from the first to the second session (Figure 23). The average number of cells coded for the first lesson was 56.8 cells while the second session recorded an average of 68.8 cells per subject. This represents a mean

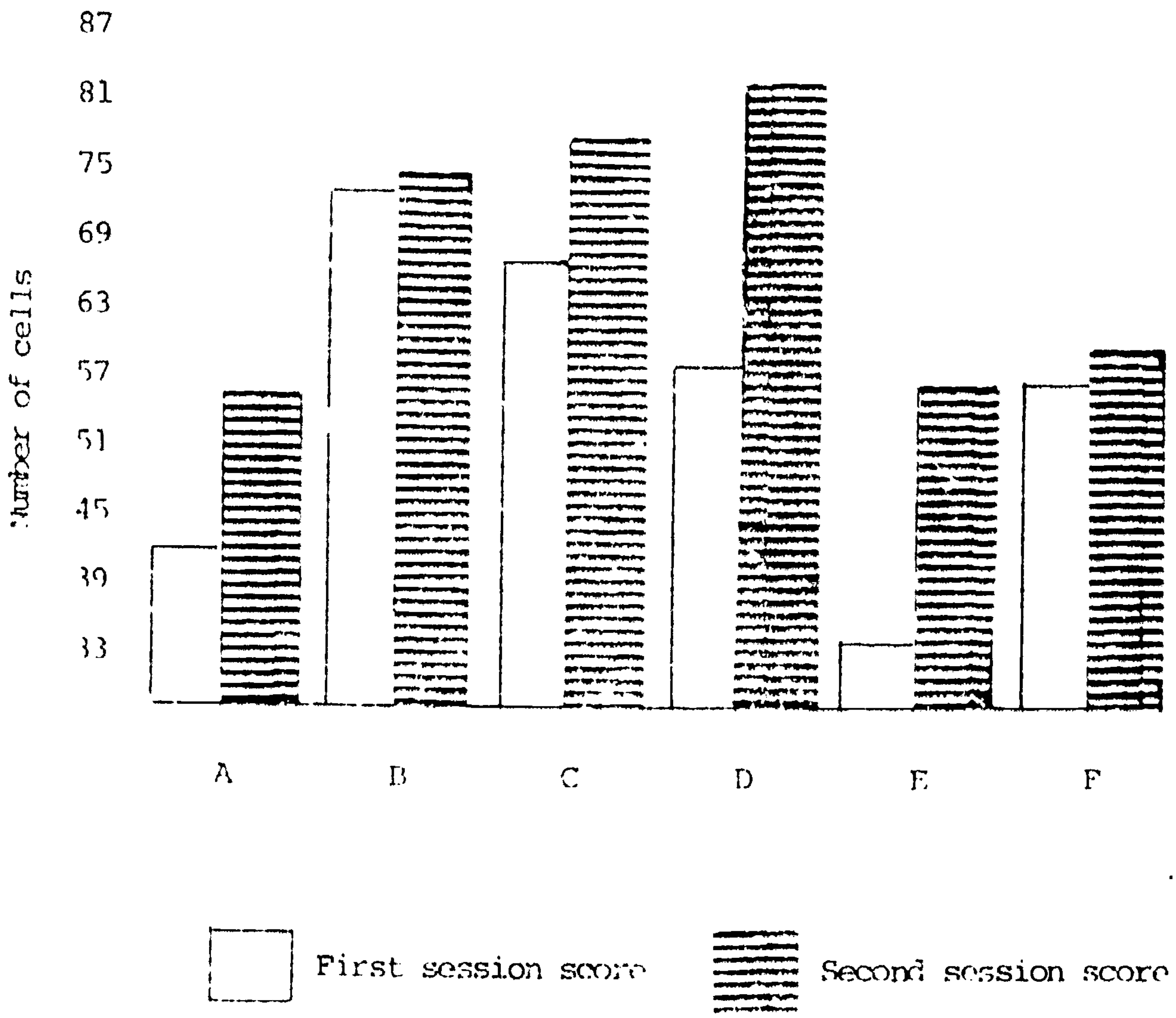

Figure 23. The number of total cells used by the interns.

growth of 12.0 cells (21.1%) per subject for the second session. This growth may be related to an increase in the number of categories subjects attempted to use. It suggests that these categories followed other categories in a number of different ways rather than the continual use of the same combination of categories.

Of the 256 total cells in the 16 X 16 matrix, 171 cells were used by the six interns (see Figure 24). This represents 66.8 percent of the total number of cells. All 256 cells or verbal patterns do exist and can be used during classroom discussion. The data indicate that within the time limits provided during the study many of these cells were reached. Results of this study warranted the tentative conclusion that the SSOR can be used to help teachers change the number, kinds, and length of verbal patterns of behavior occurring during class discussions; further, this change may occur in directions desired by teachers as well as by researchers.

Each subject was asked to write a reaction paper commenting on his experience with micro-teaching and SSOR feedback. All subjects submitted a brief appraisal. Two reported that the experience had been of no value and suggested that the time between the two teaching sessions was too short. Four reported that they had modified class discussion behaviors as a result of participation in the study. They claimed they felt more secure in their role as teacher and attributed their increased confidence to feedback provided in terms of the SSOR.

The results of this study are to be interpreted cautiously. Since no control group was used, the changes found between the first and second teaching sessions could have been the result of some factor other than the feedback procedures described. Nevertheless, some conclusions are drawn with confidence. Student categories of the SSOR do occur during class discussions. Interns were able to secure changes in their discussion behavior. A majority of the interns participating in the study believed that SSOR feedback helped them to change their behaviors in directions they valued.

Analyzing Non-occuring Events

Observation instruments sensitize observers to expected events which

SOCIAL SCIENCE OBSERVATION RECORD (SSOR)

J. Doyle Casteel and Robert J. Stahl (c.1971)

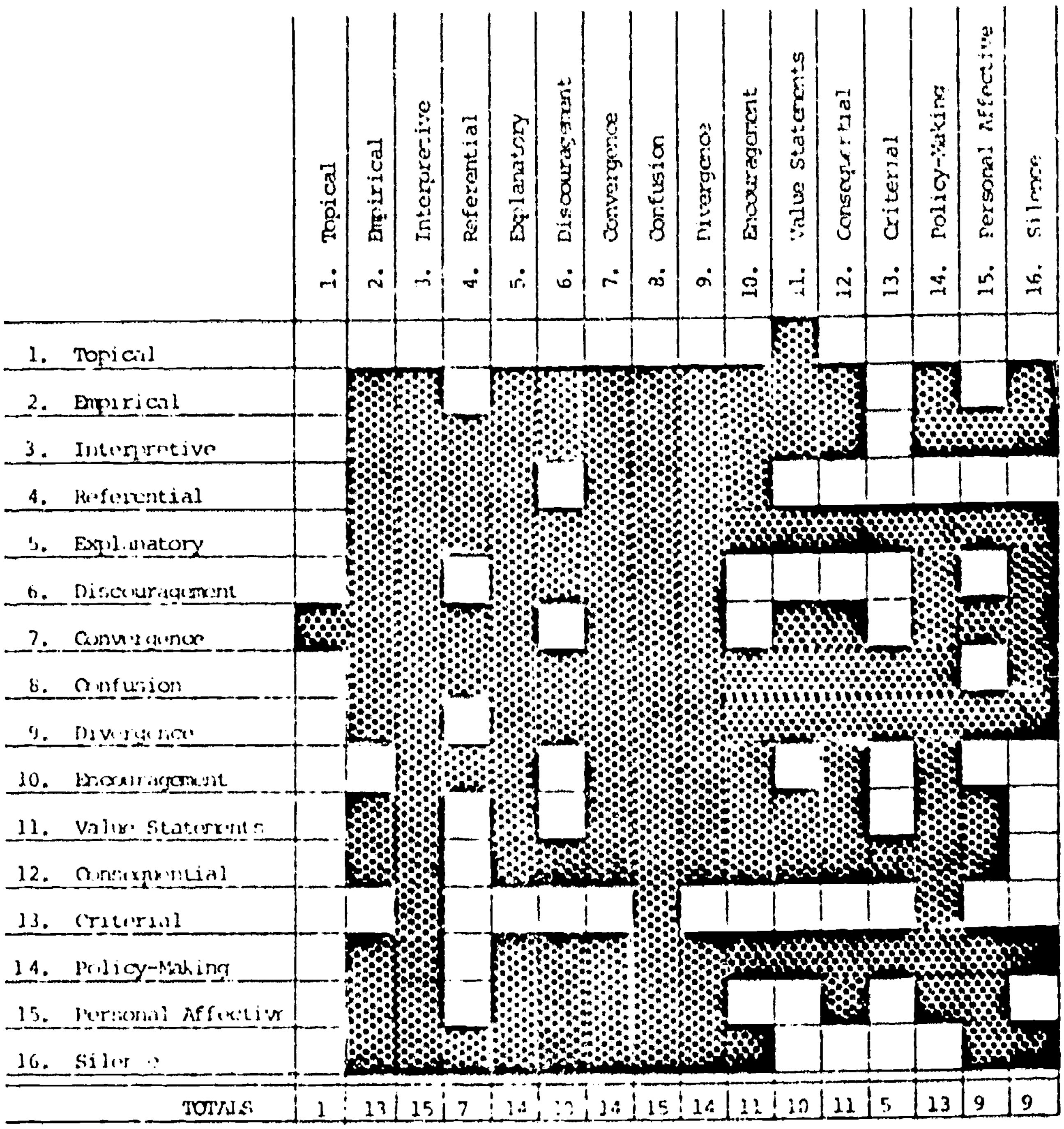

Figure 24: The cells used during the study with the social studies interns.

either do not occur or occur less frequently than anticipated. Data collected in the pilot study reported above were analyzed in this fashion. The analysis that follows illustrates how a teacher may use the SSOR in order to become more sensitive to student behaviors during class discussion.

Only one instance of topical statements (catetory 1) was coded for twelve teaching sessions. The six subjects failed to elicit statements regarding the focus of instruction from their students. Instead, each teacher tended to review and present the focus of instruction for his students. Among the implications of this teaching behavior one finds: (1) the teachers were willing to risk the assumption that students understood and could use the focus of instruction to guide their behavior; (2) the degree to which students in fact used the focus of instruction as a frame of reference within which to share relevant information, assign meaning, and make value judgements is problematic; (3) and the failure to elicit overt statements as to the focus of instruction means that the topic of discussion was not used to analyze relationships among the data reported, the opinions expressed, and values affirmed by students, tending to restrict student inquiries during micro-lessons to lower levels of thinking and valuing. Teachers involved in this study ran needless risk when they chose not to elicit topical statements to maintain purposefulness, and left unopened this doorway to cognitive and affective analysis of relationships.

Five of the six teachers made no effort to elicit from students the basis for the ideas they expressed, the values they affirmed, or the public policies they suggested (category 13, criterial statement). Among the implications of this data, one finds: (1) since students did not establish and state grounds of knowing, thinking, and valuing, they could not engage in rule-governed rational behaviors; (2) since students did not analyze the basis for their ideas and values, they could not objectify the ideals and standards which guided their verbal behavior; and (3) students limited themselves to the persuasive tool of explaining or rephrasing their previous statements (category 5) and identifying possible and probable results (category 12). What was lost was the identification of the grounds or basis for the decisions made so that the consequences and policies could be more appropriately judged as good or bad.

The teachers elicited few defining statements from students. Where students were asked to define terms, teachers accepted casual definitions. No subject sought to help students build a definition in the form of a number of concrete examples, a major approach used by historians. No subject opted to help students develop an operational definition, a common approach valued by social scientists. No subject attempted to help students list criterial attributes for language terms to which they referred. Among the implications of how words and concepts were defined, one finds: (1) the micro-lessons as taught were inconsistent with procedures found in history and the social sciences, fields presumed to provide social studies teachers with procedural rules they are expected to use and teach students to use during social inquiry; (2) the micro-lessons as taught were inconsistent with consensus-seeking procedures in that semantic confusion was risked; and (3) the teachers, all teaching social studies, demonstrated no overt concern to help students develop or clarify social science concepts, although conceptual instruction is considered to be one of the major characteristics of the "new" social studies. In this respect, the twelve micro-lessons taught were more akin to rap sessions than to disciplined inquiries.

The preceding analysis illustrates how the SSOR is used to interpret expected events which either do not occur or occur less frequently than expected during class discussion. Teachers can use such data as the basis for rethinking how they plan and teach.

A Survey of Social Educators

Results from a survey of selected instructors of social studies methods courses are reported in this section. The purpose of the survey was to determine if the sixteen categories of the Social Science Observation Record as they then existed were categories which would be perceived as being significant by those surveyed.

In the spring of 1972, thirty-two questionnaires were mailed to a selected sample of social studies educators. The first page contained category labels for sixteen categories and asked that the respondent assign a value rating to each category. A second page provided the respondent with a short definition for each of the sixteen categories. A copy of the definitions used by respondents is shown in Figure 25. Eighteen respondents returned questionnaires as requested.

Instructions as to how the questionnaire was to be answered were brief:

> Below you will find sixteen categories of verbal statements that can be emitted or elicited during oral social inquiry. Rate the relevance of each category for social inquiry according to the following code:
>
> 1. Very insignificant
> 2. Insignificant
> 3. Somewhat insignificant
> 4. No opinion
> 5. Somewhat significant
> 6. Significant
> 7. Very significant

Since the intent of the survey was to assess the opinion of social studies educators, the phrase "oral social inquiry" was deliberately left undefined. The assumption was that each respondent would rate each category in terms

THE SOCIAL SCIENCE OBSERVATION RECORD (SSOR): SHORT DEFINITIONS

REALM	CATEGORY OF STATEMENT	DEFINITIONS
I. SUBJECT-CENTERED	1. Topical	Student statements identifying the theme, the unit, the concept, the issue or th problem that is, has been, or will be the focus of group discussion.
	2. Empiric	Student statements providing verifiable data from memory, observation, reading, or oral presentation.
	3. Interpretative	Student statements assigning meaning to data or experience and expressed in the form of notions, opinions, comparisons, relationships, and connections.
	4. Referential	Student statements defining a word or concept by reference to an accepted source, by context, by examples, by operant criteria, or by ideal type.
	5. Clarification	Student statements rewording, rephrasing, or expanding on other statements by way of explanations.
II. TEACHER-CENTERED	6. Infirmable	Teacher or student statements of rejection, criticism, closure, or dissatisfaction, expressed in the form of sarcastic, doctrinaire, or negative remarks.
	7. Commentary	Teacher or student statements reviewing the directions of a group, or teacher statements summarizing, clarifying, or providing information, directions, or responding to student requests for information.
	8. Confusion	Teacher or student statements that what is being said is not understood, is causing dissonance, or lacks either internal or external consistency, or verbal or non-verbal behavior making it difficult for members of the group to communicate.
	9. Interrogative	Teacher or student questions expressed during group interaction.
	10. Confirmatory	Teacher or student statements expressing acceptance, satisfaction, encouragement, or praise.
III. MAN-CENTERED	11. Preferential	Student statements assigning a value rating or classification to an idea, person, group, object, etc.
	12. Consequential	Student statements identifying the known or anticipated effects of an action, idea, object, feeling, etc.
	13. Criterial	Student statements identifying the basis for a decision, a judgement, an action, an interpretation, etc., or developing a table of specifications for use in decision-making.
	14. Imperative	Student statements identifying alternatives that the group might consider, or student statements of what should or should not be, of what ought or ought not to be done, or expressing a decision achieved by the group.
	15. Emotive	Student statements conveying personal feelings, or efforts to express empathy with regard to the personal feelings of others.
SI-LENCE	16. Silence	Period indicating quiet, absence of verbal interaction, reading, thinking, non-verbal activities, or work.

Figure 25. The short definitions provided the social educators during the survey. (All teacher talk must be recorded in one of the teacher categories. Student statements may also be recorded in the teacher realm.) (Source: J. Doyle Casteel and Robert J. Stahl, c. 1971.)

of his own particular frame of reference.

The word "significant" was selected as the basis for a continuum, according to which categories were to be ranked. The intent was to choose a value-free word, the rationale being that the identification of negative or dysfunctional behavior has value as well as does the identification of positive and functional behavior. In some instances this result was not realized, particularly with regard to categories 7 (convergence), and 8 (confusion). The comments of some respondents indicated that they used a continuum moving from very bad to very good. Data for these three categories and Realm II are highly suspect.

Data on how categories were rated are presented in Figure 26. The total score (T) was determined by multiplying the frequency of a rating by the assigned value rating. For example, the score for category 1 was compiled in this manner: (3 X 1) + (5 X 10) + (6 X 5) + (7 X 2) = 97. In the Subject-Centered Realm, the interpretive (3) and clarifying (5) categories were rated as the most significant (mean scores 6.50 and 6.67 respectively); opinions as to the significance of empirical behaviors were most varied as suggested by a standard deviation score of 1.200. In the Teacher-Centered Realm, divergent influence (category 9) was rated as most significant (a mean score of 6.33). Confusion (8) was rated as the least significant category (a mean score of 4.17) in the realm (and in the system). Opinions as to the significance of the category of confusion were also more varied than were those for any other category in the system. In the Man-Centered Realm, criterial (13) and policy (14) statements were ranked most highly (a mean of 6.62 and 6.56 respectively); however, preferential (11) and consequential (12) statements were also ranked as

Category Numbers	Frequency of Rating							Total	Mean*	Standard Deviation
	1	2	3	4	5	6	7	T	$\bar{X}$	S.D.
1	0	0	1	0	10	5	2	97	5.39	.916
2	0	1	0	0	4	8	5	105	5.83	1.200
3	0	0	0	1	1	4	12	117	6.50	.857
4	0	0	0	1	8	6	3	101	5.61	.849
5	0	0	0	1	1	1	15	120	6.67	.840
6	1	1	2	3	1	5	5	91	5.01	1.893
7	0	0	1	2	6	6	3	98	5.44	1.096
8	1	3	2	4	4	2	2	75	4.17	1.757
9	0	0	0	1	1	7	9	114	6.33	.840
10	0	0	0	1	5	4	8	109	6.06	1.984
11	0	0	0	1	4	6	7	109	6.06	.983
12	0	0	0	1	1	10	6	111	6.17	.786
13	0	0	0	0	0	7	11	119	6.62	.502
14	0	0	0	1	1	3	13	118	6.56	1.149
15	0	0	0	1	7	3	7	106	5.89	1.023
16	0	1	0	6	4	3	4	92	5.11	1.410

*N = 18.

Figure 26. The sixteen SSOR categories as rated by the social educators surveyed.

significant. Affective statements (15) received the lowest ranking (5.11 mean score). Opinions as to the importance of policy statements varied most as the 1.149 standard deviation indicates. Opinions as to the significance of silence (16) were highly varied with seven respondents reporting that they held no opinion as to the significance of silence or that they did not perceive silence as being a significant aspect of class discussion.

Data for Realms I, II, and III are reported in Figure 27. Realms I and III were rated as significant (mean scores of 6.00 and 6.26 respectively). Realm II was rated as somewhat significant (a mean of 5.38). However, because the labels chosen for these categories resulted in the tendency by the raters to use a good-bad continuum for rating categories 6, 7, and 8, scepticism concerning the data is warranted.

This survey suggests that the ten student categories (categories 1 though 5 and 11 through 15) were perceived as categorizing significant events, whose occurrence or non-occurrence during class discussions is important. This survey led to major changes in Realm II and to the modification of some student category labels in Realms I and III.

	Frequency of Realm Totals																	Totals	Mean*	Standard Deviation
Realms	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35			
I		1							1		3	6	4		1	1	1	540	30	3.125
II	1	2			3			1	1	3	1	2	1	2			1	484	26.89	4.639
III				1						1	1	2	4	4	1	1	3	563	31.28	3.064

*N = 18.

Figure 27. The three SSOR realms as rated by the social educators surveyed.

An Experimental Study

The purpose of this study was to determine if using the SSOR as a feedback system would lead to modification of teacher and subsequent student verbal behaviors in directions that pre-service teachers believed desirable. Four areas of the SSOR which could be used to indicate changes in classroom verbal behavior were identified and used for casting hypotheses for testing. It was postulated that pre-teachers who received SSOR feedback between their first and second micro-teaching sessions would show increases in four areas. In brief, this study was conducted to determine if:

a) those receiving SSOR feedback would increase the number of cells they used;

b) those receiving SSOR feedback would increase the number of categories they used;

c) those receiving SSOR feedback would increase their Realm III totals; and

d) those receiving SSOR feedback would increase their Submatrix I totals.

These four areas were selected as foci for the study for several reasons. Explanation for the choices follows.

The use of cells in an interaction analysis matrix has been associated with teacher flexibility (Bondi, 1968). This variable was considered as a "teacher flexibility factor" by the authors at the time of the study. An increase in the total number of cells used would mean that a teacher was becoming more "flexible" in guiding the classroom discussion than one who remained constant or showed a decrease in cell use. To determine if SSOR feedback increases a teacher's ability to be flexible in the classroom, an increase in total cells used was selected as the appropriate variable in this study.

An increase in the total number of the sixteen categories in the SSOR used would indicate that statements performing a greater number of functions

BEST COPY AVAILABLE

had occurred. An increase in the total number of categories would suggest that the teacher was able to elicit a greater variety of student behaviors as they were functional and relevant to class discourse. An increase in the number of categories employed was selected as a variable for analysis because it was thought that SSOR feedback would alert teachers to more varied kinds of behaviors that they could elicit from students during class discussions.

Realm III of the SSOR yields a frequency count of value clarification statements occurring during class discussion. A growth in Realm III behaviors was expected because of the stress placed on "affective" and "values" education by social studies methods instructors at the University of Florida and because of the strong "humanistic" orientation of the College of Education at the University of Florida. That participants in the study desired to lead "humanistic" discussions was later confirmed in conversation with the students who participated in the study. An increase in the frequency of Realm III behaviors would indicate that high school students had increased their use of man-centered statements as classified by the SSOR categories. Realm III totals were also selected as a variable for analysis because it contained categories representative of the kinds of statements emphasized by contemporary social studies educators as desirable kinds of statements for teachers to elicit from their students (Oliver and Shaver, 1966; Newman and Oliver, 1970; Metcalf, 1971; Hunt and Metcalf, 1968).

The use of Submatrix I yields a frequency count of value clarification behaviors lasting more than three consecutive seconds. An increase in Submatrix I totals would indicate that students not only increased the frequency of their value clarification behaviors but also did so for extended periods of time. A teacher expecting to lead a value clarification inquiry would expect to have a large number of tallies in this submatrix. An increase in Submatrix

I totals was selected as a variable because this would be a more discriminating indicator of student changes in man-centered behavior than would be frequency counts of Realm III behaviors.

There were ten pre-service teachers in each of the experimental and control groups. Each group was assigned in accordance with a Pre-test/Post-test experimental design (Campbell and Stanley, 1963):

$$R\ 0_1 \quad X \quad 0_2$$
$$R\ 0_3 \qquad\quad 0_4.$$

This design has no control for external validity; however, it does control for all sources of internal validity. Standard t-score values and an analysis of covariance were used to determine the degree of difference between the two groups. Hypotheses were cast in terms of attaining a .95 level of significance in the difference between the two groups. In the analysis of the data, differences in group mean scores as well as t values and F ratios are reported.

To secure subjects for this study, volunteers were sought from an undergraduate social studies methods course. The students in the methods class were shown one video tape of a small group discussion led by an intern and one showing five high school students responding freely and openly to the question "What makes a good teacher?" They were asked to observe the teacher's behavior to determine if she displayed behaviors they themselves would like to possess. A question and answer period, focusing on describing the behaviors and the methods students identified as being associated with "good" and effective teaching, followed the first tape. The discussion following the second tape emphasized the discrepancies between the behaviors identified by the methods students as "good" teaching in the first tape and those behaviors seen as being "good" by high school students in the second.

The methods students were then informed that the intern leading the discussion in the first tape had learned a systematic way of planning and leading discussions that they themselves could learn. They were told that there was evidence to support the claim that their participation in the study would help them to behave in ways they wanted to behave in as classroom teachers.

Twenty students volunteered. They were randomly assigned to experimental and control groups of ten students each. All subjects were told that their participation in this study would have no effect on their grade in the course. All were assigned a time and place for the specific activities in which they were expected to participate. These activities were: (1) a familiarization session with video tape equipment and the micro-teaching lab at the P. K. Yonge Laboratory School; (2) a first and second teaching session two weeks apart during which they would lead a discussion with four to six high school students using the same discussion topic for both micro-teaching sessions; and (3) a feedback session lasting approximately two hours. All students were randomly assigned an alphabetical code. At no time during the study were students told that there were experimental and control groups.

Each student was randomly assigned to two half-hour blocks of time exactly two weeks apart. (One of the researchers handled the micro-teaching lab and equipment during the study.) As each methods student came to the micro-teaching lab, he or she was introduced to from four to six high school students who had been randomly selected from the Study of Man Department classes of the school. After a brief introduction, the video tape unit was turned on. The researcher then left the room to prevent his presence from influencing either the methods student or the high school students. After twelve minutes, the researcher returned and turned off the video tape recorder. The tape was immediately rewound, and each methods student watched a video tape replay of his lesson. Once again the researcher left the room to avoid making comments or suggestions which might contaminate the study. Any "Hawthorne effect" due to video tape replay and the micro-teaching experience must be considered to have operated equally on both groups of participants.

The tapes were observed and coded during the afternoons and evenings. In order to insure that differences in the data would not be the result of differences between coders, one researcher coded all twenty subjects for both the first and second sessions. Using Scott's method for computing between-observer reliability, the coder consistently obtained reliability coefficients for intercoder and intracoder aggreement between .73 and .79 for SSOR categories. This coder was not told the names or alphabetical code of students in the control or experimental groups. He identified each data-collection sheet by the topic discussed as he coded each lesson. When students attended the feedback sessions, their data sheets were identified only in terms of the topic they had chosen to discuss. After all second session tapes had been coded, topics were matched to alphabetical codes and names.

Feedback was provided to members of the experimental group in a group session. Members of the experimental group came to the Mead Library on the P. K. Yonge Laboratory School campus on the evening designated for their group. Students were told that the information they were going to receive in reference to their micro-teaching session was to be presented in the form of SSOR data. Students were also told that after they received the data, only questions related to the SSOR would be answered. The researchers then presented an overview of the SSOR. Students received handouts on each element in the system as a copy was projected via a transparency on the overhead and discussed. At no time was a "good" matrix or a "bad" matrix described or suggested. Teaching strategies were not discussed. After this presentation of the system, the SSOR matrix for each subject was distributed. Subjects were told that they were not to discuss their data, the SSOR, or any aspect of the feedback session with anyone. The researchers remained available to provide additional explanation of the SSOR; however they made it clear that they would make no efforts to evaluate the students' matrices. At this point students were told they could leave at any time. Some left immediately while others stayed for as long as two hours. (At the end of the study, members of the control group participated in a similar feedback session.)

A sixteen-category earlier version of the Social Science Observation Record was used in this study (see Figure 28). Besides different names

REALM	CATEGORY OF STATEMENT
I. Subject Centered	1. Topical 2. Empirical 3. Interpretive 4. Referential 5. Explanatory
II. Teacher Centered	6. Discouragement 7. Convergence 8. Confusion 9. Divergence 10. Encouragement
III. Man Centered	11. Value Statements 12. Consequential 13 Criterial 14. Policy making 15. Personal-Affective
IV. Silence	16. Silence

Figure 28. The spring, 1972, version of the SSOR.

for several of the categories, this version of the SSOR differs from the present category system in two respects. First, two categories which emphasized teacher-centered convergent and divergent behaviors (categories 7 and 9 respectively) have since been changed to the present commentary and interrogative categories (7 and 9 respectively). Second, the former category 8, labeled confusion, was split into two categories (confusion and dissonant statements) in the present system.

The results of this study stated in the form of null hypotheses are discussed along with relevant data pertaining to the difference between the group mean scores, the t value, and an F ratio for an analysis of covariance. (See Figures 29 and 30.)

First Null Hypothesis

There would be no difference in the increase in the total number of cells used by those who received the SSOR feedback between the first and second micro-teaching sessions and those who did not receive feedback at this time. In examining the difference in the means between the two groups, the experimental group increased its use of cells by 10.6 cells (21.1% increase) while the control group showed an increase of only 2.1 cells (3.6% increase). The hypothesis was supported by a t value of 1.5442 obtained for the difference between the two groups. A t value of 1.73 was required for significance at the .05 level of confidence for 19 degrees of freedom. (The t value of 1.5442 was found to be significant at the .10 level of confidence.) The hypothesis was also supported by an F ratio of 9.759 obtained for group means difference. An F ratio of 4.45 was required for significance at the .05 level of confidence for 1 and 17 degrees of freedom. No statistically significant difference at the .05 level using the t value and F ratio in terms of an increase in the use of cells was found between the two groups. The null hypothesis was not rejected.

Area in which hypothesis was tested	t (1/19)[b]	F (1/17)[a]
Increase in cell usage	1.5442	0.759
Increase in categories used	1.5132	1.966
Increase in Realm III totals	2.0527[c]	5.116[c]
Increase in Submatrix I totals	2.1166[c]	5.133[c]

a. [redacted] analysis of variance.
b. Student t test.
c. Significant at the .05 level.

Figure 29. Student t scores and F ratios for the experimental and control group difference.

Second Null Hypothesis

There would be no difference in the increase in categories used by those who received the SSOR feedback between the first and second micro-teaching session and those who did not receive feedback at this time. An examination of the difference in the means between the two groups reveals that the experimental group increased its use of categories by 1.6 categories (15.1% increase) while the control group showed a decrease of.2 categories (1.7% decrease). The hypothesis was supported by a t value of 1.4132 obtained for the difference between the two groups. A t value of 1.73 was required for significance at the .05 level of confidence for 19 degrees of freedom. (The t value of 1.4132 was found to be significant at the .10 level of confidence.) The hypothesis was also supported by an F ratio of 1.966 obtained for group means difference. An F ratio of 4.45 was required for significance at the .05 level of confidence for 1 and 17 degrees of freedom. No statistically significant difference at the .05 level of confidence, using either the t value of F ratio in terms of an increase in the total

BEST COPY AVAILABLE

Areas in which hypotheses were tested	Group Mean on Pre-test	Group Mean on Post-test	Mean Difference	Percent Difference
Increase in cell usage:				
Experimental Group	50.2	60.8	+10.6	+21.1%
Control Group	57.5	59.6	+ 2.1	+ 3.6%
Increase in categories used:				
Experimental Group	10.6	12.3	+ 1.6	+15.1%
Control Group	11.9	11.7	- .2	- 1.7%
Increase in Realm III totals:				
Experimental Group	10.5	26.1	+15.6	+148.6%
Control Group	13.6	15.9	+ 2.3	+ 16.9%
Increase in Submatrix I totals:				
Experimental Group	5.0	13.8	+ 8.8	+176.0%
Control Group	5.8	6.7	+ .9	+ 15.5%

Figure 30. Mean score differences between the experimental and the control groups.

number of categories used, was found between the two groups. The null hypothesis was not rejected.

Third Null Hypothesis

There would be no difference in the increase in Realm III totals by those who received the SSOR feedback between the first and second micro-teaching sessions and those who did not receive feedback at this time. An examination of the difference in the means between the two groups revealed that the experimental group increased its Realm III totals by 15.6 tallies (148.6% increase) while the control group showed an increase of only 2.3 tallies (16.9% increase).

The hypothesis was rejected by a significant t value of 2.9527 obtained for the difference between the two groups. A t value of 1.73 was required for significance at the .05 level of confidence for 19 degrees of freedom. The hypothesis was also rejected by a significant F ratio of 5.116 obtained for group means difference. An F ratio of 4.45 was required for significance at the .05 level of confidence for 1 and 17 degrees of freedom. When both the t value and F ratio were used, statistically significant differences were found at the .05 level between the two groups in reference to their increase in Realm III totals. This hypothesis was rejected by both the t value and an F ratio.

Fourth Null Hypothesis

There would be no difference in the increase in Submatrix I totals by those who received the SSOR feedback between the first and second micro-teaching sessions and those who did not receive feedback at this time. An examination of the difference in the means between the two groups revealed that the experimental group increased its use of Submatrix I by 8.8 tallies (176.0% increase). The hypothesis was rejected by a significant t value of 2.1166 obtained for the difference between the two groups. A t value of 1.73 was required for significance at the .05 level of confidence for 19 degrees of freedom. The hypothesis was also rejected by a significant F ratio of 5.144 obtained for group means difference. An F ratio of 4.45 was required for significance at the .05 level of confidence for 1 and 17 degrees of freedom. Statistically significant differences were found at the .05 level using both the t value and F ratio between the two groups in reference to the increase in Submatrix I totals. This hypothesis was rejected by both the t-value and F-ratio scores.

Summary

Results for pilot studies designed to explore the utility and molding power of the Social Science Observation Record have been reported. The SSOR may be used to help teachers change their behavior. The SSOR is useful for describing and interpreting expected classroom events which do and do not occur. The categories of the SSOR are perceived to be relevant to secondary school student behaviors by social studies methods instructors. Feedback in terms of SSOR data when provided to pre-service teachers tends to result in increases in the number of value clarification statements expressed by their students. An increase in these statements is valued by pre-service teachers.

The studies reported here are little more than a beginning. Unanswered questions abound. Are patterns of student behavior associated with different levels of thinking? Are different patterns of student, teacher, or student and teacher behaviors correlated with student learning, student achievement, and student perceptions? When teachers use indirect influence do they secure patterns of student responses which differ significantly from those obtained when they use more direct influence? Do teachers who ask probing questions follow consistent and reliable strategies which can be identified and described by the SSOR; if so, can these behaviors be taught to teachers who do not use probing questions? What of teacher conditional moves, explanatory moves, structuring, and other strategies? Correlational and experimental studies germane to these questions are still to be designed, conducted, and reported.

Chapter V

DEVELOPING INTERCODER AGREEMENT

An Introduction

Initial analysis of data for the experimental study reported in the preceding chapter indicated that the Social Science Observation Record has potential value for pre-service and in-service teacher education. Experimental subjects, given the benefit of SSOR feedback, modified their behavior as teachers and secured consequential change in student behavior. These changes were consistent with the self-reported objectives of both experimental and control group subjects. These changes were also in a direction valued by a selected group of social studies methods instructors who responded to a questionnaire. Futhermore, the change in student verbal behavior was congruent with results reported for a national survey of randomly selected teachers of social studies methods (Tucker, 1972).

At this point, a decision had to be made concerning whether to proceed to engage in correlational and experimental studies or to determine the degree to which SSOR data could be coded reliably. The decision was made that between-observer reliability warranted attention prior to further studies or to the reporting of studies already completed. In order to report what had been done and to plan subsequent studies it was necessary to determine if the realms, categories, and submatrices were adequately defined and stated, thus conveying to others the concepts and rules which had been developed and used for the collection, organization, and interpretation of data reported in the studies.

The historical development of the SSOR influenced this decision. The SSOR was slowly framed over a four-year period. It did not originate as a deliberate effort to develop an observation system. Working as a post doctoral fellow at the University of Washington, Casteel identified four interrogative modes of teacher behavior and organized these into an instructional module for elementary teachers. This module was field tested in two in-service workshops conducted in the spring of 1969 in Seattle (Jarolimek,1969). In the winter quarter of 1970, Casteel used this module with nine University of Florida social studies interns who were doing peer-group teaching designed to increase their ability to lead class discussions. When the interns involved reported that they had accidently learned to recognize ("code") the intent of a leader's questions and could do so repeatedly and accurately, the possibility of developing a feedback system dawned. By the summer quarter of 1970, the interrogative modes had been converted into categories of student statements likely to follow different kinds of questions. At this point, all teacher talk was limited to one category. The categories were organized to form a ten-by-ten matrix. (See Figure 31.) When this format was used in conjunction with peer-group teaching and feedback, members of a methods class coincidentally learned the system to the degree necessary to exercise control over a teacher's relative success or failure in obtaining instructional objectives. During the fall of 1970, Stahl became a partner. Together he and Casteel began to construct a model of class discussion incorporating categories associated with student understanding and value clarification verbal behaviors. In subsequent quarters pre-service social studies teachers were asked to participate in peer-group teaching and were provided feedback in terms of the SSOR. In the spring of 1971, this experience with peer-group teaching was used as a basis from which the sixteen categories were identified. At this time, these categories were organized into four realms of associated behaviors. Submatrices were identified and this recognition began to generate conjecture as to their meaning. Further work with pre-service

THE SSOR MATRIX

	1. Topical	2. Empirical	3. Interpretive	4. Evaluational	5. Teacher Talk	6. Referential	7. Valuational	8. Decision	9. Personal-Affective	10. Silence		TOTALS
1. Topical												
2. Empirical												
3. Interpretive												
4. Evaluational												
5. Teacher Talk												
6. Referential												
7. Valuational												
8. Decision												
9. Personal-Affective												
10. Silence												
TOTALS												

Figure 31. The ten category SSOR matrix · (Source: J. Doyle Casteel, c.1970).

social studies teachers in the fall of 1971 led to a systematization of equipment including data-collection sheets, scatter diagrams, and the SSOR matrix. In the winter of 1972, the pre-post study reported in Chapter IV was conducted, and Casteel and Stahl found that they coded reliably; i.e., both could code the same lesson and collect the same data. In the Spring of 1972, the experimental study reported earlier was conducted and the questionnaire to selected social studies educators was mailed. An analysis of those questionnaires returned led to changes in the Teacher-Centered Realm and to some alteration of category labels. The only change of substance made after this date was the creation of category 17, confusion, as a distinct non-verbal category.

The foregoing chronology explains the problem. For more than two years the SSOR was "teased" into shape as a model of discussion and as a feedback instrument. During this period categories had been identified, modified, rejected, and reconstructed. When the developers checked to determine if they were coding discussion in the same way (reliably), they consistently obtained reliability coefficient scores of 0.73 or better for categories and scores of 0.93 or better for realms using Scott's formula (Scott, 1955). They could not, however, estimate the degree to which the definitions and ground rules of which they were consciously aware and reporting were indeed the definitions and rules they used to code data. A basic need, then, was to make certain that the SSOR as developed and understood by those who developed it could be communicated accurately to others.

There was little point in proceeding immediately to the planning of correlational and experimental studies unless it could be determined that the system could be communicated to and learned by others. The SSOR needed to be tested for and, if necessary, made reliable. Two reliability studies were planned and conducted to meet this objective. Subsequent sections of this chapter will describe what is meant by "between-coder reliability" and its significance, explain how the reliability studies were organized and conducted, and report reliability data obtained from the two studies.

The Meaning of Between-Coder Reliability

Any observation instrument--whether a paper and pencil test or a category system--is intended to make available factual data which otherwise would not be available for analysis or, if available, would not be assigned its warranted significance. Once the task of developing a reliable observation system has been accomplished, both the occurrence and non-occurrence of expected events can be interpreted.

Kaplan (1964) refers to between-coder reliability by the term *intersubjectivity*. By this he means that as observers categorize and record data, they use the same subjective criteria consistently. Public criteria make subjective factors common to everyone.

The concept of intersubjectivity relates to the SSOR. A number of observers observing the same class discussion would code and record the same SSOR data. What is occurring during the discussion would reveal itself to all observers engaged in the subjective act of coding student and teacher behaviors into one of seventeen categories for each three-second interval. If all observers are trained to see and code the same events, their intersubject agreement is evidence that the different observers are using the same concept and ground rules (factors common to all observers) to code data. This commonality then "testifies" to the objectivity of the instrument. The degree to which they agree may be computed and called their coefficient of between-observer reliability.

Between-observer reliability, as Kaplan explains it, assumes more methodological importance than some have assigned it (Medley and Mitzel, 1963). If observers agree reliably, a classroom teacher who has comprehended Chapters

II and III can interpret data collected by a trained coder. Further, the teacher may be confident that the data are objective in the sense that his understanding, as he interprets data, and the understanding of the coder, as he coded the data, are sufficiently comparable to be judged reliable. For the teacher interested in modifying either his verbal behavior (Realm II) or that of students (Realms I and III), between-observer reliability offers assurance that observed changes in either his behavior or that of his students are to be interpreted as "true" changes in events observed rat than as coding errors. For the researcher, between-observer reliability develops confidence that data collected for analytical purposes are sufficiently objective for framing inferences. This allows him to analyze and interpret data for purposes other than those purposes for which they were originally collected. Data about instructional behavior collected through systematic observation are difficult to collect (Medley and Mitzel, 1963). They need not be discarded when original goals have been fulfilled. These reasons help to explain why between-coder agreement is important in discussion about interaction analysis observation systems.

How Between-Coder Reliability Was Developed

Two reliability studies were conducted by the investigators. Each study encompassed eight sessions. The first study extended over a three-week period of time. The second was concentrated into two weeks. The procedures, materials, and sequence for both studies are reconstructed below in the form of a summary for each session.

Session 1

Trainees were presented with a package of materials entitled "The Social Science Observation Record: A Presentation Package." This package provided trainees with a one-page overview of the system and the following elements: function chart, short definitions, data-collection sheets, scatter diagram, SSOR matrices, descriptions of submatrices, and results of the pre-post study reported in Chapter IV. The intent was two-fold: (1) to help trainees understand the system they were learning, and (2) to train them to transfer raw data to the SSOR matrix accurately. At the end of this session, trainees were provided with seventeen flash cards. Each flash card had the number for a category printed on one side and the category label designated by that number on the reverse side. Trainees were instructed to study the flash cards and use the function chart found in the presentation package in order to develop three skills: (1) given a category name, they could report its number without pausing to think; (2) given a category number, they could report its name without pausing to think; and (3) given a categorial function, they could, on request, report immediately either the correct category number or label. The first skill was emphasized because SSOR data are collected by assigning numbers to behaviors occurring during three-second intervals.

Session 2

Trainees were divided into two-person teams. The team members were instructed to drill one another using the flash cards distributed at the end of the first session. Each team member practiced stating category numbers for category labels as these were presented by his partner. When teams had become facile at this task, each trainee was presented with a twenty-five-page package, entitled "Categories of the SSOR: Extended Definition" (Casteel and Stahl, 1972). This package included long definitions for each category with selected examples of student or teacher statements for each category and ground rules. (Ground rules are efforts to anticipate coding problems and to provide all coders with the same rules to follow.) Trainees read these definitions. As questions arose, they were answered by the trainers. At the end of the session trainees were asked to continue practicing with their flash cards, studying the extended definitions, and memorizing the ground rules for coding.

Session 3

During this session trainees reviewed prior learnings and began to code printed transcripts. Trainees divided themselves into pairs and drilled one another using flash cards (see session 2 above). This lasted for approximately twenty minutes and was intended to maintain what had already been memorized.

After drilling, trainees were presented with one or two pages of printed transcripts of a class discussion. The coding rule followed was that each line of print was to be coded once, but, if a change of behavior occurred within a line, the line should be coded twice or more to capture changes as they occurred. For the first transcripts, one trainee read a line orally, and a second trainee coded the line as one (or more) category of the SSOR. Trainees were encouraged to, and in fact did, raise questions when they could not agree as to why a line was coded in a particular way. Explanations were made by referring to appropriate sections of "Categories of the SSOR: Extended Definitions." As much as possible the trainees were referred to the various definitions and the function chart they possessed.

In the next activity, trainees remained in their two-member teams. Each team coded by numbers a page of printed transcript in consultation with each other. (Trainees were not allowed to use any SSOR materials.) At the end of each page the trainer, who served as the criterion person for all subsequent reliability tests, read his codings. If differences existed between the trainees and the trainer, the trainer explained how he had arrived at his coding. Care was taken to explain that the objective of this exercise was for trainees to learn to "read" and eventually "hear" as the trainer read and heard. Debates as to who was correct and who was incorrect were avoided.

After practicing in this fashion for approximately two hours, trainees were handed a transcript (PT01) and asked to code it independently as an initial check on their ability to code discussions. After coding, trainees used a scatter diagram and transferred their data into the SSOR matrix. Total category counts, realm counts, and submatrices counts were completed. When one of the trainers had checked the matrix for accuracy, each trainee could leave. The third session ended.

Session 4

This session began with a review of numerical designations (flash card drill) and proceeded to a brief discussion of those categories found to cause coding difficulty in PT01. Trainees were asked to pair off into teams and were handed a printed transcript which had already been coded by the trainers. Each team was asked to identify the basis that led the trainers to code the transcript as it was coded using "Categories of the SSOR: Extended Definitions" and the function charts as reference. This required approximately one hour, and trainers refused to help teams for fear that any explanation might be perceived as defensive behavior. Next, trainees coded short sections of another printed transcript and periodically checked their coding against the coding of the criterion trainer. This lasted for approximately one hour. At this point trainees individually coded two transcripts designated as PT06 and PT07. Transcript PT06 stressed categories 1 through 10 and was constructed to diagnose trainee ability to code subject-centered behaviors reliably. Transcript PT07 stressed categories 6 through 15 and was designed to determine trainee ability to code man-centered behaviors reliably. Again, the trainees were allowed to leave only after their completed matrices had been checked by the trainers.

Session 5

Trainees were invited to peruse their reliability scores as they arrived. (Anonymity was provided by assigning a letter designation to each trainee known only to the trainee.) A brief discussion to clear up any coding problems followed with emphasis on the need to obey the ground rules. The skill of timing was then introduced. Trainees first listened to an audiotape that beeped every three seconds. Then they used data-collection sheets and wrote a letter of the alphabet each time they heard a beep. After practice in conjunction with the tape, trainees were asked to write a letter of the alphabet every three seconds using their own judgement as to the length of three seconds. Following two to three such practices all trainees were consistently coding between 19 and 22 letters per minute. This was considered an acceptable level of time by the research. (That timing developed so rapidly was a pleasant surprise for the trainers as well as for the trainees. However, practice at timing with and without the tape was reviewed at the beginning of each subsequent session.) This drill involved use of the same data-collection sheets used for coding video tapes later.

At this point, video tapes were introduced. First, short segments varying in length from fifteen seconds to two minutes were played, and the criterion trainer called aloud the category of behavior occurring during each interval. After doing this several times, the video tape was rewound to the original segment coded by the trainer. Trainees were instructed to code the tapes following two additional rules: (1) once you've coded an interval forget it. Even if you believe you've made an error don't attempt to correct it for this will confuse your timing and increase the chances of errors in a number of subsequent intervals, and (2) if more than one category of behavior occurs during an interval, code both in the order in which they occur.

BEST COPY AVAILABLE

After giving these directions, the tape was started, and the trainer indicated that trainees were to begin coding when he called aloud the first category he heard and recorded. The trainer and trainees then coded silently for one minute. After the one-minute coding episode was over, the trainer read aloud his codings by which trainees checked theirs. If discrepancies occurred, the tape was replayed with the trainer explaining the basis for his coding. As with printed transcripts, discussions as to who was right and who was wrong were avoided as much as possible. The stress was on understanding the basis the trainer used for his codings. Five such practice segments occurred during this session. The practice segments continued for all remaining session except the last.

After a break, trainees first watched and then coded a discussion (VT01 for the first group and VT13 for the second group of trainees.) They then transferred their data to the SSOR matrix and totaled category, realm, and submatrix counts prior to leaving. This transfer consistently required thirty to forty-five minutes.

Session 6

Trainees reviewed timing as explained above. Trainees coded several segments of a video tape and checked their coding for congruence with the trainer criterion as described for session 5. Again, the length of three practice segments varied from less than thirty seconds to more than two minutes. The criterion trainer frequently coded the entire segment orally while trainees watched and listened. At the end of this session trainees watched and coded a video taped discussion, built matrices, and left.

Session 7

As trainees arrived they were invited to view their reliability scores for the second video tape test. Timing was practiced. Training with taped segments continued. A video-taped discussion was watched, then coded. After a break, a second video tape was watched and then coded. Trainees concluded this session by transferring the data for the two video tapes into separate matrices. Their completed matrices again served as their ticket to leave.

Session 8

Trainees reviewed timing. Trainees watched a video tape. Trainees then coded this video tape. They transferred their data into the SSOR matrix and took a short break until all had completed their matrix transfer.

At this time trainees were told where their reliability scores were to be posted, asked to critique procedures used, and thanked for their cooperation. This ended the final session.

The reconstruction of reliability training sessions describes the procedures, materials, and sequence used. Hopefully, the description is adequate to help the reader understand how between-observer reliability training sessions are conducted. Scores for individual trainees and results of two studies which examined between-observer reliability are reported next.

Objectives for the Reliability Studies

As described in Chapters II and III, the SSOR has three major dimensions--realms, categories, and submatrices. The experimental study reported in Chapter IV involved hypotheses cast in terms of these dimensions. For this reason, both reliability studies were directed at training coders to develop specific criterion reliability scores for realms, categories, and submatrices.

Castcel served as the criterion coder for both sessions. This means that all reliability figures were computed against his codings rather than using the standard practice of computing between-observer reliability by pairing off the trainees. He coded each of the printed transcripts and video tapes at the same time as the trainees. His codings were used as the criterion against which all trainees were compared. In order to maintain stability as the criterion observer, he consistently checked his reliability by running inter-coder trials with Stahl. He also checked his intra-coder reliability by coding the same printed transcripts and tapes on different days. Scores of 0.80 or better for realms, categories, and submatrices were obtained consistently using Scott's method to compute between-observer reliability. Scott's method was also used to compute the reliability tests for all trainees. Scott's method was selected becuase it appeared to be the conventional formula used by persons such as Flanders (1960), Ober, et. al. (1971), and Bondi (1968). (See Figure 32 for an example of how reliability was computed using Scott's method as used by Ober, et. al. [1971].)

Reliability scores were computed for realms, categories, and submatrices. The typical practice found was to determine between-observer reliability with reference to categories only. Failure to compute other dimensions used to analyze and interpret behaviors has been criticized (Medley and Mitzel, 1963). The SSOR matrix contains four realms, seventeen categories, and twelve submatrices. Reliability scores were calculated for each of these.

Different goals for the two studies were established. For the first study a mean reliability score of 0.60 for the last two video tapes was posited as being adequate. For the second study a mean reliability score of 0.70 for the last two video tapes was posited as being adequate. The standards were borrowed from Ober, et. al. (1971). These authors suggest that a between-observer reliability score of 0.60 indicates sufficient mastery of a system for it to be used by a classroom teacher who wished to collect and use data reflecting his own instruction. They further suggest a reliability score of 0.70 for those who are serious about learning a system. The authors posited that a 0.70 was an acceptable minimal level for collecting SSOR data

SSOR RELIABILITY COMPUTATION FORM

Category Reliability

Cat.[a]	Crit. %[b]	Coder Nos.[c]	Coder %[d]	% Diff.[e]
1	.02	6	.03	.01
2	.09	16	.07	.02
3	.28	73	.32	.04
4	.00	0	.00	.00
5	.15	26	.11	.04
6	.01	0	.00	.01
7	.07	17	.07	.00
8	.01	2	.01	.00
9	.10	25	.11	.01
10	.02	5	.02	.00
11	.02	8	.03	.01
12	.03	6	.03	.00
13	.03	10	.04	.01
14	.03	5	.02	.01
15	.01	4	.02	.01
16	.12	21	.09	.03
17	.02	7	.03	.01
Total	1.01	231	1.00	.21

Po = 1.00 - Sum(%/Diff.) - 1.00 - __.21__ = Po = __.79__

$$Pe = (\text{Highest \% Criterion})^2 + (\text{2nd Highest \% Criterion})^2$$

$$(\underline{.28})^2 + (\underline{.15})^2 = \underline{.078} + \underline{.022} = Pe = .10$$

$$r = \frac{Po - Pe}{1.00 - Pe} = \frac{.79 - .10}{1.00 - .10} = \frac{.69}{.90} = .77 = r$$

a. category numbers.

b. percentage of total tallies recorded by criterion coder for each category.

c. Number of tallies in each category by the trainee coders.

d. percentage of total tallies recorded by trainees for each category.

e. difference between criterion and trainee percentages.

Figure 32 : Example of how reliability is calculated using Scott's method as described by Ober ,et.al.(1971).

for research purposes. The 0.70 score on the second test was selected to determine if trainees could learn the SSOR to the degree necessary for participation as trained coders in the collection of research data. These standards were established primarily so that the trainers could test and evaluate their materials. These standards were not stressed with the trainees. A teacher need not be able to code data reliably to be able to use the SSOR as a frame of reference for examining his instructional behaviors. (See Chapter IV.)

The First Study

The first study was conducted between October 30 and November 16, 1972. Participants in the study were in-service social studies teachers enrolled in the secondary social studies methods block in the College of Education, University of Florida. Students were invited to participate in a reliability study in lieu of three weeks of participation and observation in local schools. Students were informed that their grades would not be affected if they chose to participate in the study. Twelve students volunteered for the study. Each student was randomly assigned a letter of the alphabet for identifying his papers. These letters were assigned to prevent any possible bias on the part of the trainers while tabulating the results of the data. The criterion trainer coded all printed transcripts and video tapes at the same time that trainees coded them. The second trainer handled the matrix data and computed all the reliability scores.

The twelve trainees coded and recorded four printed transcripts for which realm reliability was computed. The mean score for each trainee for realms is shown in Figure 33. For all reliability checks with printed transcripts, all twelve trainees attained the 0.60 standard with all twelve also having a mean score of 0.70 or better. The mean score for all the trainees on each of the four printed transcripts was 0.76 or better. The mean score for the last two

	Realm			Category			Submatrix		
Subject	$\overline{PT}$	$\overline{VT}$	$\overline{Total}$	$\overline{PT}$	$\overline{VT}$	$\overline{Total}$	$\overline{PT}$	$\overline{VT}$	$\overline{Total}$
A	.80	.83	.81	.62	.70	.66	.76	.77	.77
B	.80	.82	.81	.58	.64	.61	.76	.70	.73
C	.87	.80	.83	.46	.59	.53	.75	.74	.75
D	.85	.67	.76	.65	.55	.60	.74	.60	.67
E	.92	.85	.88	.62	.71	.67	.83	.73	.78
F	.82	.71	.77	.58	.54	.56	.74	.71	.72
G	.80	.68	.74	.55	.61	.58	.76	.63	.69
H	.82	.81	.82	.53	.50	.52	.82	.79	.80
I	.74	.80	.77	.28	.51	.40	.67	.72	.69
J	.73	.79	.76	.51	.66	.58	.67	.77	.72
K	.82	.77	.80	.57	.68	.62	.79	.69	.74
L	.74	.78	.76	.54	.64	.57	.71	.74	.72
MEAN	.81	.78	.79	.54	.61	.58	.75	.71	.73

$\overline{PT}$ represents the mean of three printed transcript scores.

$\overline{VT}$ represents the mean of five video tape scores.

$\overline{Total}$ represents the mean of all eight scores.

Figure 33: The mean scores for the first reliability study. (Using Scott's method for computing between-observer agreement.)

tests for each trainee revealed that all twelve trainees passed the 0.60 standard with nine reaching or surpassing the 0.80 level. (See Appendix A for a report of all scores on all tests for each trainee.) No trainee was below 0.60 on any of the tests for realm reliability using printed transcripts.

Four tests for realm reliability using video tapes were also taken. On the first check, seven of the eight trainees attending the session met the 0.60 standard, five had reliability coefficients of 0.70 or better, and the mean for all trainees was 0.76. On the third tape, eleven trainees attained the 0.60 standard. On the final tape, only one trainee fell below the 0.60 minimum. Mean scores for the last three checks tend to improve from 0.76 to 0.78 to 0.80. Mean realm-reliability scores for the last two video tapes were considered. Eleven trainees coded these two tapes. Only one trainee failed to meet the 0.60 standard as a mean score for the last two tapes. Six of eleven trainees had reliability coefficient scores of 0.80 or better.

The mean scores for all trainees for all printed transcripts and video tapes were computed. All twelve trainees met or surpassed the minimum 0.60 score posited by the trainees. All twelve also bettered the 0.70 level with six meeting the 0.80 score when the mean of all scores was computed.

Category-reliability scores were computed for all trainees for all printed transcripts and video-tape tests. The mean score for each trainee for categories is shown in Figure 33. For all reliability checks with printed transcripts, three of the twelve traineees attained the 0.60 standard. The range of mean scores for all trainees on each of the four tests was from 0.44 to 0.60. For the first printed transcipts (PT01) six of twelve trainees attained 0.60 or better. For the second printed transcript (PT06) five of eleven attained 0.60 or better. For the third printed transcript only two of ten attained the 0.60 standard, and the mean score for the ten dropped to 0.44. For the fourth printed transcript four of ten scored 0.60 or better with a mean score for ten trainees of 0.55. Only four of the twelve trainees failed to reach the 0.60 minimum on at least one of the four printed transcript tests while three of the trainees averaged 0.60 or better over all four tests.

Attention to the four video tapes coded reveals that two of eight trainees met the 0.60 standard on the first tape (VT13). The mean for the eight on this tape was 0.56. On the second tape (VT13) six of twelve trainees met or surpassed the 0.60 standard. The mean score for the group was 0.59. Eight of eleven trainees met the 0.60 standard on the third tape (VT14) while ten reached the standard on the fourth tape (VT15). The group means for these two tapes were 0.61 and 0.67 respectively. Eight of the eleven undergraduate volunteers who took the last two video-tape tests attained mean scores for these two tests of 0.60 or better. Three of the eight had a category-reliability mean score of 0.70 or higher on these two tests. It is noteworthy that ten of the eleven trainees did reach the 0.60 standard on the last video-tape test.

The mean category scores for all trainees for their combined printed transcript and video-tape tests were computed. Five of the twelve trainees met the 0.60 standard posited by the trainers. Seven of the twelve failed to meet this minimum standard when their mean scores were computed.

Between-coder reliability scores were also computed for submatrix agreement on all the printed transcript and video-tape tests. The mean for each trainee for submatrix-agreement scores is also shown in Figure 33. For the four tests for submatrix reliability using printed transcripts, each trainee reached the 0.60 level at least once. All met or surpassed the 0.70 level at least once. On the first test (PT01) eleven of the twelve bettered the 0.60 standard with a group mean computed at 0.75. Eleven of eleven coders met the 0.60 level on the second test (PT06). On test three (VT07) seven of ten met the 0.60 criterion. Nine of ten trainees bettered the 0.60 standard posited by the trainers on test four (VT06). These nine also surpassed the 0.70 level. The group mean for the last printed transcript test was figured to be at 0.78. All twelve coders averaged 0.60 or better on all their printed transcript tests with ten attaining a mean score of 0.70 or higher.

When video tapes were made the focus of coding, the tendency on the part of the coders was to code with greater agreement (higher reliability). Six of the twelve trainees were present for all four video-tape tests. On the first video tape (VT13), seven of eight coders attained the 0.60 standard while compiling a group mean score of 0.70. The second test (VT13) found all twelve trainees coding reliability at the 0.60 level, with seven meeting a 0.70 score. Ten of eleven trainees bettered the 0.60 standard on the last two video-tape tests (VT14 and VT15). The group mean for each of these was 0.74 and 0.70 respectively. An examination of the mean score of each trainee for the last two video-tape tests reveals that eleven of the twelve met or surpassed the minimum 0.60 standard, with ten reaching the 0.70 level. All twelve of the trainees averaged 0.60 or better on the four video-tape tests.

The twelve undergraduate coders met or exceeded the criterion, 0.60, established for submatrix reliability. All twelve achieved this task on both printed transcripts and video-tape replay tests.

In summary, of the twelve undergraduate trainees, twelve met the 0.60 standard on their mean realm scores, five met the criterion for categories, and twelve met the criterion for submatrices. The result of this first attempt at training coders to code reliably indicated that, with between eighteen and twenty-four hours of training, a participant could learn to code classroom behaviors at the 0.60 level. This study gave the trainers confidence that the system itself could be communicated accurately to others to the degree that they could code with a high level of agreement with the trainers. However, one of the problems reported by the volunteer coders was the need for more ground rules to be included in the materials entitled "Categories of the SSOR: Extended Definitions." Several trainees reported that the trainers apparently had not specified all the ground rules they were using when they coded data. A second suggestion was that the short practice segments of from fifteen seconds to two minutes be employed more often. A third suggestion was made with regard to conducting future studies. It often took up to forty-five minutes to transfer the transcript or tape data to the matrix; a considerable portion of each of the last four session was spent in building matrices. The trainees pointed out that if a method was available to bypass this time consuming task, more of the training session time could be used for training coders. While the trainers did make adjustments in the materials and procedures in response to the first two student-coder requests, they did not solve the problem concerning time spent transferring data to the matrix.

The Second Study

The second study was conducted between January 15 and 26, 1973. Ten persons volunteered to participate in the study. These ten included two pre-service social studies teachers, two social studies interns, two graduate assistants in the Department of Secondary Education, one graduate student, one science teacher, one social science teacher, and one undergraduate Early Childhood Education student. The training sessions were conducted at P. K. Yonge Laboratory School and were held four nights a week for two consecutive weeks. Again letter codes were given to each student. The two trainers fulfilled the same role as they had in the earlier study. Between-coder reliability tests for both printed transcripts and video tapes are discussed below.

Realm reliability using printed transcripts was tested three times during the second study. Because of various personal problems, only five of the ten trainees coded all three printed transcripts. The mean score for each trainee for realm is shown in Figure 34. For all reliability checks with printed

Subject	Realm			Category			Submatrix		
	$\overline{PT}$	$\overline{VT}$	$\overline{Total}$	$\overline{PT}$	$\overline{VT}$	$\overline{Total}$	$\overline{PT}$	$\overline{VT}$	$\overline{Total}$
A	.81	.79	.80	.71	.68	.69	.80	.67	.71
B	.66	.68	.67	.84	.45	.41	.59	.54	.56
C	.88	.89	.88	.74	.68	.70	.83	.77	.79
D	.74	.80	.78	.69	.64	.66	.72	.66	.68
E	.89	.75	.78	.64	.65	.65	.84	.62	.66
F	.87	.82	.84	.70	.65	.67	.86	.73	.78
G	.81	.70	.75	.71	.61	.65	.80	.64	.71
H	.80	.75	.76	.53	.64	.62	.81	.60	.64
I	.83	.70	.73	.58	.65	.64	.81	.60	.65
J	.93	.90	.91	.91	.85	.85	.90	.81	[illegible]
$\overline{X}$	.82	.78	.79	.65	.65	.65	.80	.66	.70

$\overline{PT}$ represents the mean of three printed transcript scores.

$\overline{VT}$ represents the mean of five video tape scores.

$\overline{Total}$ represents the mean of all eight scores.

Figure 34: The mean scores for the second reliability study. (Using Scott's method for computing between-observer agreement.) (See Appendix for detailed data on these reliability scores.)

transcripts, all ten trainees attained the earlier 0.60 standard with nine of the ten attaining a mean score of 0.70 or better. The mean score for all trainees on each of the three printed transcript tests was 0.71 or better. Six of the seven trainees who participated in the first printed transcript test passed the minimum 0.60 standard with five of the seven meeting or surpassing the 0.70 level. The mean for all trainees was 0.79. Five of six trainees attained the 0.60 level on the second test with four passing the 0.70 level. A mean of 0.71 was computed for the second test. Eight of eight trainees surpassed the 0.70 level with seven of the eight having scores of 0.80 or higher on this last printed transcript test. The mean score for the eight trainees on the third test was 0.87.

A similar pattern was revealed when reliability scores for realms were tabulated when five video tapes were used. Six of the eight trainees taking the first test using video tapes met the 0.70 standard posited by the trainers, with their mean score being 0.74. Five of eight met the 0.70 standard on the second video-tape test. On the third test, eight of nine trainees passed the 0.70 standard with six of the trainees surpassing the 0.80 level. The mean of the nine scores for this test was 0.81. Seven of nine and nine of nine trainees met or passed the 0.70 standard on the fourth and fifth text respectively. On the fianl test, three trainees reached the 0.90 level. The mean of the nine trainees for the last video-tape test was 0.83. All trainees reached the 0.70 standard on realm reliability for video tapes. The mean realm score for their last two video tapes were examined for consistency. Of the ten trainees all ten had reliability mean scores of 0.80 or better.

Computation of the mean scores for all trainees for all printed transcripts and video tapes revealed that nine of the ten trainees surpassed the 0.70 standard with four meeting or surpassing the 0.80 level. Only one of the ten trainees failed to meet the 0.70 level when the mean of all his scores was computed.

Between-observer agreement in reference to categories was computed for three printed transcripts and five video tapes. The mean scores for each trainee for category reliability is shown in Figure 34. Of the ten trainees, five were present for all three printed transcript tests while four trainees coded only one printed transcript. Three of seven met the 0.70 standard for the first printed transcript (PT01), three of six for the second (PT02), and four of seven for the third transcript (PT06). The group mean scores for these three tests were 0.73, 0.59, and 0.63 respectively. The difference in mean scores among the three tests is partially explainable in that one trainee dropped from a 0.63 first-test score to a 0.11 score on the second, and a 0.27 score on the third test. Six of the ten trainees met the 0.70 minimum standard at least once on the three printed transcripts. Five of ten averaged 0.70 or better for transcript tests.

With regard to the five video-tape tests, the results were more positive. Again, half of the trainees were present for all five tape tests. The first video-tape (VT01) test found only one of eight trainees meeting the 0.70 standard. The second tape (VT02) found none of the eight trainees meeting the minimum standard. In both the first two tape tests the mean of the group was 0.55. The mean of the third video-tape test rose to 0.63 with two of nine trainees meeting the 0.70 level. Five of nine trainees reached or bettered the 0.70 standard on tests four (VT04) and five (VT05). The means for these tests were 0.72 and 0.71 respectively. An examination of the last two video tapes that each trainee coded reveals that six of the ten averaged better than 0.70. When the mean score for all video-tape tests was completed, only one trainee met the 0.70 standard with a 0.85 average score. However, eight of the ten surpassed the 0.60 mark.

An examination of the mean of all trainees for all tests taken reveals that only two trainees averaged 0.70 or better. Nine of ten averaged better than 0.60 on all their tests. The group mean for eight tests was 0.65.

Mean submatrix reliability scores are also shown in Figure 34. Only five of ten participants coded all three printed transcripts. Six of the seven trainees coding the first transcript(PT01) bettered the 0.70 standard while compiling a group mean score of 0.78. Four of six and eight of eight met or surpassed the 0.70 level on the second and third tests respectively. Nine trainees bettered the 0.70 level on the second and third tests. All ten trainees bettered the 0.70 standard at least once on the printed transcript tests while eight of the ten trainees averaged 0.80 or higher over all three tests.

When video tapes were first introduced, none of the eight trainees present attained a score of 0.70 or higher. On the second video-tape test (VT02) only one of the eight trainees met the established standard. The group mean for these two tests was 0.55 and 0.56 respectively. A noticeable improvement was seen in the third test when six of nine trainees bettered the 0.70 criterion score. On the final two tests four of nine and six of nine met the minimum standard of 0.70. The group mean for each of the last three video-tape tests was above this standard with respective scores of 0.72, 0.72, and 0.73. While all ten trainees met the 0.70 standard at least once on video-tape tests, only three had attained average scores on video tapes above this level.

The mean scores for all trainees on the submatrix reliability tests were computed. Five of the ten trainees met or surpassed the 0.70 minimum standard set by the trainers with nine of ten bettering the 0.60 score.

In summary, of the ten volunteer participants in this study nine met the 0.70 standard on their mean realm scores, six met the criterion level for categories, and six met the criterion for submatrices. The results of this study to determine if trainees could attain a score of 0.70 indicate that, with between eighteen and twenty-four hours of training, some trainees could meet the suggested criterion level for coding data for research purposes. These results suggest that the SSOR can be learned by others to the degree necessary for research purposes. However, in spite of modifications in the materials and procedures as suggested by participants in the first study, category reliability remained a slow process for trainees. The participants themselves could not identify any reason except the time factor of twenty-four hours as being the cause of this slow growth in between-observer agreement. The trainers did compute one interesting fact that may have contributed to the outcome of the study. While the total time of the training sessions was twenty-four hours, the actual time spent training participants was fourteen hours. More than six hours alone was spent transferring data from the eight tests into matrices. The trainers were unable to remedy this situation. The participants were positive toward the training sessions, the procedures used, and their own efforts. The suggestions they made for further modifications required minor changes in the materials. The alterations have since been completed.

In Conclusion

The conclusion that trainees can learn the SSOR and demonstrate relatively high levels of between-coder agreement is warranted. The reliability of an observer is his score as computed according to the mathematical formula suggested by Scott (1955). Another formula would probably have resulted in somewhat different scores. Trainees who scored below 0.60 in the first study or 0.70 in the second study did not fail. Neither did those who met these criteria pass a test of competence. Their scores represent a level of agreement on a particular test with a criterion coder taking the same test at the same time. The criterion scores served primarily as goals toward which the trainers worked and assessed their efforts, materials, and training procedures. Between-observer agreement does represent one way of determining whether efforts to convey concepts and ideas through speaking and writing are understood in the same sense and may be applied in the same sense and in the same manner by an audience with whom one wished to communicate. Not every person who volunteers for training can be assured that he will meet even the 0.60 level within the time limits specified in these studies.

The Social Science Observation Record (SSOR) can be learned in varying degrees by pre-service and in-service teachers. Classroom teachers do not need to be reliable coders in order to understand and use the SSOR as a descriptive, analytical or modification tool. A teacher can use the SSOR as a model of discussion or as a systematic framework to look at his own behavior without in-depth training. However, the research reported in this monograph and elsewhere indicates that teachers are quite different teachers when they have some knowledge of an observation system than when they have no such knowledge.

Thus, the teacher who has knowledge of the SSOR possesses a frame of reference that he can use to examine and modify his behaviors and those of his students.

APPENDIX

Reliability scores from two reliability studies.

Realm Reliability Scores: 1st Study

Subject	PT01	PT06	PT07	PT06	VT13	VT13	VT14	VT15	Mean PT	Mean VT	Mean All Tests
A	.66	.86	.85	.82	.66	.83	.88	.93	.80	.83	.81
B	.92	.74	.76	.79	.87	.75	.77	.87	.80	.82	.81
C	.83		.85	.94	.90	.62	.80	.87	.87	.80	.83
D	.86	.92	.72	.91	.77	.78	.64	.48	.85	.67	.76
E	.88	1.00	.93	.86	.69	.88	.89	.93	.92	.85	.88
F	.75	.89		.83		.68	.69	.80	.82	.71	.77
G	.66	.95	.71	.88	.51	.70	.74	.77	.80	.68	.74
H	.75	.82		.88		.80	.72	.90	.82	.81	.82
I	.81	.70	.76	.70	.96	.71	.78	.75	.74	.80	.77
J	.64	.83	.71			.81	.85	.72	.73	.79	.76
K	.69	.91	.85			.74	.81	.80	.82	.77	.80
L	.63	.65	.90	.79	.78	.78			.74	.78	.76
Mean	.76	.84	.80	.84	.77	.76	.78	.80	.81	.78	.79

PT refers to Printed Transcripts.
VT refers to Video Tape.

Results of the SSOR Reliability Study: October 30 to November 16, 1972.
Final results represent 18 hours of training.
All figures computed using Scott's method for computing between-observer reliability.

Category Reliability Scores: 1st Study

Subject	PT01	PT06	PT07	PT06	VT13	VT13	VT14	VT15	Mean PT	Mean VT	Mean All Tests
A	.67	.59	.67	.54	.63	.70	.72	.76	.62	.70	.66
B	.73	.61	.42	.57	.57	.66	.62	.69	.58	.64	.61
C	.53		.36	.48	.57	.48	.62	.69	.46	.59	.53
D	.48	.69	.72	.69	.51	.59	.58	.53	.65	.55	.60
E	.62	.60	.59	.67	.71	.71	.64	.78	.62	.71	.67
F	.53	.56		.64		.49	.51	.61	.58	.54	.56
G	.63	.76	.21	.61	.52	.64	.61	.66	.55	.61	.58
H	.51	.53		.55		.49	.40	.60	.53	.50	.52
I	.47	.51	-.05	.20	.42	.38	.62	.61	.28	.51	.40
J	.47	.54	.51			.53	.71	.74	.51	.66	.58
K	.47	.69	.54			.68	.67	.69	.57	.68	.62
L	.65	.49	.42	.58	.53	.74			.54	.64	.57
Mean	.57	.60	.44	.55	.56	.59	.61	.67	.54	.61	.58

PT refers to Printed Transcripts.
VT refers to Video Tapes.

Results of the SSOR Reliability Study: October 30 to November 16, 1972.
Final results represent 18 hours of training.
All figures computed using Scott's method for computing between-observer reliability.

Submatrix Reliability Scores: 1st Study

Subject	PT01	PT06	PT07	PT06	VT13	VT13	VT14	VT15	Mean PT	Mean VT	Mean All Tests
A	.70	.88	.70	.76	.70	.73	.85	.80	.76	.77	.77
B	.83	.74	.70	.77	.74	.76	.65	.63	.76	.70	.73
C	.80		.60	.85	.83	.63	.76	.75	.75	.74	.75
D	.78	.84	.56	.76	.74	.67	.57	.43	.74	.60	.67
E	.91	.86	.77	.77	.63	.65	.87	.77	.83	.73	.78
F	.57	.78		.86		.71	.65	.76	.74	.71	.72
G	.72	.86	.65	.80	.48	.64	.70	.70	.76	.63	.69
H	.81	.84		.81		.85	.75	.76	.82	.79	.80
I	.82	.70	.57	.59	.72	.70	.77	.67	.67	.72	.69
J	.69	.78	.55			.80	.79	.73	.67	.77	.72
K	.69	.90	.79			.63	.76	.69	.79	.69	.74
L	.69	.60	.79	.78	.75	.72			.71	.74	.72
Mean	.75	.80	.67	.78	.70	71	.74	.70	.75	.71	.73

PT refers to Printed Transcripts.
VT refers to Video Tapes.

Results of the SSOR Reliability Study: October 30 to November 16, 1972.
Final results represent 18 hours of training.
All figures computed using Scott's method for computing between-observer reliability.

Realm Reliability Scores: 2nd Study

Subject	PT01	PT02	PT06	VT01	VT02	VT03	VT04	VT05	Mean PT	Mean VT	Mean All Tests
A	.70	.91		.86	.70	.78		.82	.81	.79	.80
B	.93	.32	.74	.61	.45	.90	.65	.79	.66	.68	.67
C	.87	.88	.89	.77	.92	.89	.91	.94	.88	.89	.88
D	.57	.74	.91	.81	.80	.90	.76	.74	.74	.80	.78
E			.89	.84	.65		.82	.70	.89	.75	.78
F	.87	.78	.96	.77	.77	.83	.83	.91	.87	.82	.84
G	.65	.88	.91		.58	.75	.68	.79	.81	.70	.75
H			.80	.72	.78	.57	.92	.77	.80	.75	.76
I			.83	.52		.81	.77		.83	.70	.73
J	.93					.85	.85	.99	.93	.90	.91
Mean	.79	.71	.87	.74	.71	.81	.80	.83	.82	.78	.79

PT refers to Printed Transcripts.
VT refers to Video Tapes.

SSOR Reliability Study: January 15th to 26th, 1973.
Final results represent 14 hours of training.
All figures computed using Scott's method for computing between-observer reliability.

Category Reliability Scores: 2nd Study

Subject	PT01	PT02	PT06	VT01	VT02	VT03	VT04	VT05	Mean PT	Mean VT	Mean All Tests
A	.63	.78		.56	.65	.67		.84	.71	.68	.69
B	.63	.11	.27	.40	.36	.23	.66	.61	.34	.45	.41
C	.65	.81	.75	.59	.67	.66	.77	.69	.74	.68	.70
D	.67	.66	.75	.73	.65	.63	.66	.52	.69	.64	.66
E			.64	.51	.58		.80	.72	.64	.65	.65
F	.87	.46	.76	.66	.53	.69	.67	.72	.70	.65	.67
G	.72	.70	.72		.47	.74	.64	.58	.71	.61	.65
H			.53	.59	.46	.60	.80	.74	.53	.64	.62
I			158	.58		.66	.72		.58	.65	.64
J	.91					.79	.78	.97	.91	.85	.86
Mean	.73	.59	.63	.55	.55	.63	.72	.71	.65	.65	.65

PT refers to Printed Transcript.
VT refers to Video Tapes.

SSOR Reliability Study: January 15th to 26th, 1973.
Final results represent 14 hours of training.
All figures computed using Scott's method for computing between-observer reliability.

Submatrix Reliability Scores: 2nd Study

Subject	PT01	PT02	PT06	VT01	VT02	VT03	VT04	VT05	Mean PT	Mean VT	Mean All Test
A	.72	.87		.57	.69	.73		.70	.80	.67	.71
B	.68	.37	.72	.39	.27	.74	.61	.68	.59	.54	.56
C	.86	.79	.85	.62	.79	.75	.79	.88	.83	.77	.79
D	.73	.62	.80	.57	.53	.82	.69	.71	.72	.66	.68
E			.84	.58	.57		.69	.63	.84	.62	.66
F	.86	.78	.95	.62	.69	.78	.76	.78	.86	.73	.78
G	.73	.87	.80		.53	.69	.63	.71	.80	.64	.71
H			.81	.65	.44	.52	.76	.63	.81	.60	.64
I			.81	.43		.68	.68		.81	.60	.65
J	.90					.75	.86	.81	.90	.81	.83
Mean	.78	.72	.82	.55	.56	.72	.72	.73	.80	.66	.70

PT refers to Printed Transcripts.
VT refers to Video Tapes.

<u>SSOR Reliability Study</u>: January 15th to 26th, 1973.
Final results represent 14 hours of training.
All figures computed using Scott's method for computing between-observer reliability.

REFERENCES

Amidon, E. J. & Flanders, N. A. Interaction analysis as a feedback system. Interaction analysis: Theory, research and application, E. J. Amidon and J. B. Hough (editors). Reading, Massachusetts: Addison-Wesley, 1967, 121-140. (a)

Amidon, E. J. & Flanders, N. A. The role of the teacher in the classroom: A manual for understanding and improving teacher classroom behavior. Minneapolis: Association for Productive Teaching, Inc., 1967. (b)

Amidon, E. J. & Hough, J. B. (editors). Interaction analysis: Theory, research and application. Reading, Massachusetts: Addison-Wesley, 1967.

Aschner, M. J. A system for classifying thought processes in the content of classroom verbal interaction. (Cooperative Research Project, No. 965.) Urbana, Illinois: Institute for Research on Exceptional Children, University of Illinois, 1965.

Aschner, M. J. & Gallagher, J. J. Aschner-Gallagher system for classifying thought processes in the context of classroom verbal interaction. Mirrors for behavior: Volume II, A. Simon and G. Boyer (editors). Philadelphia: Research for Better Schools, Inc., 1967, Aschner-Gallagher, I-60.

Bellack, A. A., Kliebard, H. M., Hyman, F. T., & Smith, F. L. The language of the classroom. New York: Columbia University Press, 1966.

Bondi, J. C., Jr. The effects of interaction analysis feedback on the verbal behavior of student teachers. Gainesville, Florida: University of Florida, College of Education, Unpublished Ed.D. dissertation, 1968.

Brown, B. B. The experimental mind in education. New York: Harper and Rox, 1968.

Brown, B. B., Ober, R. L. Soar, R. S., & Webb, J. W. The Florida taxonomy of cognitive behavior: Directions. Gainesville, Florida: Institute for Development of Human Resources, University of Florida, January, 1967 (mimeographed).

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. Handbook of research on teaching, N. L. Gage (editor). Chicago: Rand McNally, 1963, 171-236.

Casteel, J. D. Verbal strategies of valuing. Gainesville, Florida: University of Florida, College of Education, 1968 (ditto).

BEST COPY AVAILABLE

Casteel, J. D. & Stahl, R. J. Categories of the SSOR: Extended definitions. Gainesville, Florida: University of Florida, College of Education, 1972 (ditto).

Casteel, J. D. & Stahl, R. J. The social science observation record (SSOR): Revised introductory packet: 02. Gainesville, Florida: P. K. Yonge Laboratory School, February, 1973 (mimeograph).

Flanders, N. A. Analyzing teaching behavior. Reading, Massachusetts: Addison-Wesley, 1970.

Flanders, N. A. Teacher influence, pupil attitudes, and achievement. (Cooperative Research Project, No. 397.) U. S. Office of Education, Minneapolis: University of Minnesota, 1960.

Flanders, N. A. Teacher influence, pupil attitudes, and achievement. (OE-25040, Cooperative Research Monograph No. 12.) Washington, D.C.: U. S. Department of HEW, 1965.

Fuller, F. F. FAIR system manual: Fuller affective interaction records. Mirrors for behavior: Volume IX, A. Simon & G. Boyer (editors). Philadelphia: Research for Better Schools, Inc., 1970, 43.1-43.103.

Hunt, M. P. & Metcalf, L. E. Teaching high school social studies. 2nd edition. New York: Harper and Row, 1968.

Jarolimek, J. (editor). Social studies education: The elementary School--focus on inquiry. Social Education, 1969, 33 (5, May), 533-550.

Kaplan, A. The conduct of inquiry. San Fransisco: Chandler Publishing Co., 1964.

Kaplan, L. The Florida taxonomy of affective behavior. Gainesville, Florida: Institute for Development of Human Resources, University of Florida, 1970 (mimeographed).

Massialas, B. G. Evaluating social inquiry in the classroom. The High School Journal, 1969 LIII (2, November), 75-86.

Medley, D. M. & Mitzel, H. E. Measuring classroom behavior by systematic observation. Handbook of research on teaching, N. L. Gage (editor). Chicago: Rand McNally, 1963, 247-328.

Metcalf, L. E. (editor). Values education: Rationale, strategies, and procedures. NCSS 41st Yearbook, Washington, D.C.: National Council for the Social Studies, 1971.

Morine, G., Spaulding, R. S., & Greenberg, S. Discovering new dimensions in the teaching process. Scranton, Pennsylvania: International Textbook Company, 1971.

Newman, F. M. & Oliver, D. W. Clarifying public controversy: An approach to teaching social studies. Boston: Little, Brown and Company, 1970.

Ober, R. L., Bentley, E. L., & Miller, E. Systematic observation of teaching: An interaction analysis—instructional strategy approach. Englewood Cliffs, New Jersey: Prentice-Hall, 1971.

Oliver, D. W. & Shaver, J. P. Teaching public issues in the high school Boston: Houghton-Mifflin, 1966.

Raths, L. E., Harmin, M., & Simon, S. B. Values and teaching: Working with values in the classroom. Columbus, Ohio: Charles E. Merrill, 1966.

Raup, B. R., Axtelle, G. E., Benne, K. D., & Smith, B. O. The improvement of practical intelligence. New York: Harper and Brothers, 1943.

Remmers, H. H. Rating methods in research on teaching. Handbook of research on teaching, N. L. Gage (editor). Chicago: Rnad McNally, 1963, 329-378.

Rosenshine, B. Evaluation of classroom instruction. Review of Educational Research, 1970, 40 (2), 279-300. (a)

Rosenshine, B. Interpretative study of teaching behaviors related to student achievement, final report. (National Center for Educational Research and Development: DHEW/OE--ERIC ED 051 116.) Philadelphia: Temple University, 1970. (b)

Rosenshine, B. & Furst, N. The use of direct observation to study teaching. Second handbook of research on teaching, R. M. Travers (editor). Chicago: Rand McNally, 1973, 122-183.

Rowe, M. B. Relation of wait-time and rewards to the development of language, logic, and fate control: Parts I and II. Journal of Research in Science Teaching. (In press.)

Scott, W. A. Reliability of content analysis: The case of nominal coding. Public Opinion Quarterly, 1955, 19 (3), 321-325.

Simon, A. & Boyer, G. (editors). Mirrors for behavior: Volumes I-VI. Philadelphia: Research for Better Schools, Inc., 1970.

Simon, A. & Boyer, G. (editors). Mirrors for behavior: Volumes VII-XIV. Philadelphia: Research for Better Schools, Inc., 1970.

Smith, B. O. & Meux, M. O. A study of the logic of teaching. Urbana, Illinois: University of Illinois Press, 1970.

Soar, R. S. Optimum teacher-pupil interaction for pupil growth. Educational Leadership Research Supplement, 1968, 26 (3, December), 275-280.

Soar, R. S. Research finding from systematic observation. Journal of Research and Development in Education, 1970, 4 (1, Fall), 116-121.

Tucker, J. L. Teacher education and the "new social studies." Social Education: Research Supplement, 1972, 36 (5, May), 548-554, 560.

Withall, J. The development of a climate index. Journal of Educational Research, 1951, 45 (October), 93-100.