DOCUMENT RESUME

ED 137 290 SP 010 933

AUTHOR Armento, Beverly Jeanne

TITLE Selected Correlates of Effective Teacher Behavior During Concept Instruction: Their Design, Utility and Limitations.

PUB DATE Apr 77

NOTE 37p.; Paper presented at Annual Meeting of the American Educational Research Association (New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.

DESCRIPTORS Academic Achievement; Classroom Communication; Classroom Environment; *Concept Teaching; Educational Objectives; *Effective Teaching; *Elementary School Teachers; *Formative Evaluation; *Interaction Process Analysis; Learning Characteristics; Performance Factors; *Teacher Behavior; Teaching Methods; Verbal Communication

ABSTRACT

This document describes the construction, implementation, and implication of selected high inference measures applied in a study of teacher effectiveness in the third, fourth, and fifth grades. Selected independent variables served as hypotheses regarding which behaviors are likely to occur during concept instruction and which are likely to be relevant to student concept learning. Two basic assumptions guided the selection of relevant behaviors: (1) Teacher behavior should be examined in terms of intent. Intent may be derived from instructional objectives. (2) Relevant process variables should be derived from existing theoretical or empirical bases that provide support for expecting certain relationships between instructional behavior and student outcomes. For this investigation, a record of classroom communication between teacher and students was made on audio-tape recordings. Analysis of classroom interaction between teacher and pupils included evaluation of how accurate and complete was the teacher's knowledge of the subject and how effective was the teacher in conveying concepts to the pupils. Teaching techniques were analyzed in the light of resulting student understanding and achievement. (JD)

SELECTED CORRELATES OF EFFECTIVE TEACHER BEHAVIOR DURING CONCEPT INSTRUCTION: THEIR DESIGN, UTILITY AND LIMITATIONS

BEVERLY JEANNE ARMENTO
Assistant Professor of Education
Georgia State University
Urban Life Building, Room 6C0
University Plaza
Atlanta, Georgia 30303

*A paper presented at the Annual Meeting of the American Educational Research Association, April, 1977, New York City.

SELECTED CORRELATES OF EFFECTIVE TEACHER BEHAVIOR DURING CONCEPT INSTRUCTION: THEIR DESIGN, UTILITY AND LIMITATIONS

Researchers and teachers have long been interested in questions surrounding the nature of effective teaching. The traditional teacher effectiveness research paradigm sought correlates between teacher personality, experiential, and/or aptitude variables and criterion variables of student ratings or student achievement. The results of such studies have yielded little useful knowledge (Gage, 1963).

Recently, a more productive approach has been sought through research on the nature of instructional environments. One way of defining a learning environment is in terms of the behavioral characteristics of its participants. The reasoning underlying this emphasis is that the dominant features of an environment depend upon the typical characteristics of its members and that certain environments tend to reinforce or to extinguish specific behaviors. It is assumed that instructional environments differ in the particular behaviors they reinforce and thus tend to produce differential effects in terms of the nature, quantity, and quality of student outcomes.

It may indeed be the case that optimal learning environments differ according to the nature of the anticipated student outcome(s). That is, sub-environments which are highly dissimilar may exist, even within one classroom. These varying settings may serve to reinforce differing learning outcomes. It is conceivable, for example, that a sub-environment which reinforces divergent, creative thought processes may not promote the learning of specific facts and generalizations.

One of the tasks of investigators who attempt to identify optimal learning environments is the determination of relevant aspects of sub-environments which are likely to reinforce specific behavioral outcomes. In this connection, the need for a taxonomy of situations and learner outcomes is apparent. One might ask:

(a) What are the basic categories of school goals (such as concept and generalization learning, the development of divergent thinking skills, the development of problem creation and solution skills)?

(b) Are there identifiable sub-environments which optimally promote such learner outcomes?

Needless to say, such a taxonomy does not exist. We continue to view school learning in terms of subject matter goals rather than in terms of general skills, abilities, and attitudes which are supported by the various academic disciplines. Most process-product research efforts are conducted in the context of describing teaching as it occurs within subject matter parameters with little attention being given to the types of learning or student achievement being promoted or even to the teacher's intent, as reflected in course or instructional objectives.

Gage (1963) and Rosenshine and Furst (1971) have urged teacher behavior investigators to conduct studies such as those being suggested here, in which specifically defined aspects of teacher behavior are examined. Such micro-effectiveness studies could examine teacher and student behaviors in terms of instructional intent, as reflected through statements of objectives. It may be useful to concentrate some process-product research efforts around categories of student achievement in order to investigate possible treatment by outcome relationships.

One category of student outcome is that of concept learning. The study reported here is one in a projected series of investigations aimed at identifying and validating the characteristics of classroom sub-environments which promote optimum levels of concept learning. One long-term goal of this research is to test the generalizability of optimum concept learning sub-environments across types of learners and across subject matter. The purpose of this paper is to describe and critique the conceptualization and utility of the independent teacher process variables and the instruments employed to measure these aspects of teacher behavior.

Identifying Relevant Teacher Variables

Given the lack of a theory or conceptual model of teaching from which to select behaviors, the task of identifying relevant teacher and student process variables is the first critical step in the effort to collect valid and reliable data on the teaching act. Selected independent variables serve as hypotheses regarding which behaviors are likely to occur during concept instruction and which are likely to be relevant to student concept learning. Two basic assumptions guided the selection of relevant behaviors.

(1) Teacher behavior should be examined in terms of intent. Intent may be derived from instructional objectives.

(2) Relevant process variables should be derived from existing theoretical or empirical bases which provide support for expecting certain relationships between instructional behavior and student outcomes.

The variables in this study were derived from previous process-product investigations and from experimental studies of concept learning. The focus was on the verbal cognitive aspect of the teacher's task rather than on all possible dimensions.

The generation of particular teacher process variables was facilitated by asking the question: What are the characterisitcs of a concept instructional event which relate logically to clear, effective instruction? This question was answered as follows. In preparation for a concept instructional sequence, a teacher must respond to at least three pragmatic concerns.

(1) What particular knowledge is needed to achieve the instructional objectives? This question refers to the substantive aspect of instruction.

(2) What terminology ought to be employed to transmit meaningful ideas most

effectively to learners? This question refers to the <u>semantic</u> aspect of instruction.

(3) What particular logical, procedural moves ought to be made during the lesson to meet the instructional objectives most effectively? This question refers to the <u>strategic</u> aspect of instruction.

Substantive, semantic, and strategic components of instruction served as major categories for the generation of teacher antecedent and process variables. Following is a brief explanation of each instructional component along with the names of the variables employed in each category.

<u>Substantive variables</u>. The substantive aspect of an instructional event refers to the body of knowledge explicitly made available to students during the lesson. In a primarily discussion-mode lesson, much knowledge is made available through teacher discourse or explanation. Scriven (1959) suggests the three criteria of accuracy, adequacy, and relevance for satisfactory explanations. These three criteria can aid in the identification of relevant antecedent as well as process variables. Three antecedent questions are:

(1) How accurate is the teacher's knowledge (of the relevant subject)?

(2) How adequate or complete is the teacher's knowledge of the subject? and

(3) Is the knowledge which the teacher is able to generate relevant to the knowledge demanded by the specific instructional objectives?

Five variables related to the substantive aspect of instruction were employed in this study:

<u>Accuracy</u> (1) The concept definitions given or developed by the teacher are accurate.

(2) The concept examples given or accepted by the teacher accurately represent the concept.

<u>Adequacy</u> (1) The teacher explicitly states the necessary knowledge components

(as implied by the instructional objectives).

(2) The teacher explicitly states the necessary concept labels or names (as implied by the instructional objectives).

Relevance (1) The teacher's verbal behavior is appropriate to the achievement of the instructional objectives for the lesson.

Substantiation for the accuracy, adequacy, and relevance variables can be found both in previous process-product investigations and in concept learning studies. Positive relationships have been found between the variable "opportunity to learn the criterion material" and student performance. The "opportunity to learn" variable is similar to the adequacy and relevance variables employed in this study. Rosenshine (1972), Shutes (1969), and Husen (1967) found significant positive correlations between measures of opportunity to learn and student achievement. In both the Rosenshine and Shutes studies, actual tape-scripts of lessons were assessed to determine the extent of content coverage. In Husen's study, teachers rated whether their students had the opportunity to learn the type of problem(s) represented by the test items.

An important aspect of concept instruction is that the concept examples illustrate the critical dimensions of the concept. Experimental investigations on concept learning support the principle that as the critical properties of the concept become more obvious, ease of concept attainment increases (Clark, 1971). Inaccurate concept examples should, then, inhibit the efficient learning of concepts. In addition, experimental concept learning studies have shown that associating the critical properties and instances of concepts with the concept name or label increases the ease of subsequent concept attainment. This aspect of concept instruction is reflected in the adequacy of concept label coverage variable.

Semantic variables. The semantic aspect of an instruction event refers to the teacher's ability to convey meaning through appropriate choices of terminology.

For this study, the semantic component was expanded to include syntactics, which deals with the rules governing word order.

As children progressively become able to perform formal operational thought processes, verbal language becomes increasingly more important as the medium of instruction. For children in grades three through five, verbal language itself is a major component of instruction. Semantic variables ought to play a critical role in the assessment of effective communication in instruction.

Three semantic variables were examined in this study.

(1) The teacher employs a balance of concrete and abstract terminology.

(2) The teacher speaks in complete, rather than incomplete, choppy sentences.

(3) The teacher uses pronouns which clearly refer to their antecedents.

Certain semantic abilities of the teacher could be identified as antecedent predictor variables to be examined in future studies. Abilities which logically relate to verbal performance during concept instruction might include such measures as verbal fluency and divergent production of classes.

Strategic variables. Instructional strategy refers to the total set of movements, or operations, performed by the teacher to achieve the instructional objectives. A strategy is comprised of smaller elements or purposive moves. A purposive move refers to an activity aimed at progressing the lesson from one substantive point to another. An utterance is a verbal expression performed by one person at a given time. An utterance may contain a single purposive move, or may contain several purposive moves. The specific purposive moves identified for this study were derived from instructional variables found to relate to concept attainment and from results of process-product investigations. The purposive strategic moves and the anticipated direction of their relationship with student achievement were:

(1) The teacher gives a concept definition. (positive)

(2) The teacher asks students to give a concept definition. (positive)

(3) The teacher gives a positive or negative concept example. (positive)

(4) The teacher asks students to give positive or negative concept examples. (positive)

(5) The teacher reviews and summarizes the main ideas in the lesson. (positive)

(6) The teacher asks a low order question. A low order question prompts students to engage in recall or translation as cognitive processes. (positive)

(7) The teacher asks a high order question. A high order questions requests students to engage in cognitive processes of comparison/contrast, analysis application, or evaluation. (null)

(8) The teacher changes or shifts the topic of the lesson. (null)

  (a) The teacher signals a shift in the topic. (positive)

  (b) The teacher employs a summary-signal-shift pattern. (positive)

  (c) The teacher shifts the topic while asking a low order question. (negative)

  (d) The teacher shifts the topic while asking a high order question. (negative)

(9) The teacher asks a pair of questions in a series, not allowing time for student response. (negative)

(10) The teacher answers his/her own or a student's question by explaining. (positive)

(11) The teacher repeats his/her own question, following a student's response. (negative)

(12) The teacher rephrases his/her own question, following a student's response. (negative)

(13) The teacher tells students to stop irrelevant behavior; or, the teacher engages in irrelevant behavior. (negative)

(14) Other behavior. Teacher's utterances which contained <u>none</u> of the above purposive moves were coded in this category.

Two additional strategic variables were added on the strength of results of several process-product studies.

(1) The teacher expresses enthusiasm and interest in the content of the lesson.

(2) The teacher displays an on-task approach toward the classroom atmosphere and its interactions.

<u>Designing the Observation Instruments</u>

Two major phases are apparent in the process of measuring classroom behavior: (1) securing a record of a sample of the behaviors to be measured; and (2) quantifying the record (Medley and Mitzel, 1963).

For this investigation, a record of classroom communication between teacher and students was made on audio-tape recordings. Twenty-two teachers of students in grades 3, 4, and 5 were instructed to conduct two concept lessons of forty-five minutes each, on the economic concept specialization. Teachers were provided with background knowledge on the concept and with a set of instructional objectives for the two lessons. Fifteen children were randomly identified, from within the intact class to which the teacher was assigned; this group became the teacher's instructional class. Two full days prior to instruction, students were pre-tested on a criterion-referenced measure matched to the instructional objectives; this test was again administered following the second lesson. The class mean residual gain score was the statistical unit of analysis representing teacher effectiveness of concept instruction.

Despite the fact that the identification of relevant classroom behaviors has been based on theory or on empirical evidence, there is no assurance that the quantification system developed to describe the behaviors will be valid and reliable. If an investigator is to develop a new quantification system, a number of major decisions must be made. Each decision point serves as a source of invalidity. The actual quantification system, at best, should be viewed as hypotheses that certain definitions for variables and certain ways of recording behaviors are related to student outcome measures. The investigator must make decisions about the recording procedure, item content, coding format, and the unit of analysis for each observational system developed (Borich, 1977). These characteristics of observational instruments will be discussed below in the context of the three instruments developed for this study: the Observational System for Concept Instruction (OSCI), the Tally Form and the Rating Form.

Recording Procedure. Sign or category procedures are used to record the frequency of the behavior under consideration. If an event is recorded only once, regardless of its actual frequency of occurrence, the recording instrument is called a sign system. If an event is recorded each time it occurs, the recording instrument is called a category system. A rating system is usually viewed as a modified sign system, wherein the observer makes on estimate of the frequency of an event, usually at the end of an observational session.

Two category systems and a modified category rating system were developed for this study. The Observational System for Concept Instruction (OSCI) was designed to record the sequence and frequency of the teacher (and student) strategic variables (see Figure 1). The actual recording of behavior procedure employed with OSCI played no role in the quantification process; the quantification occurred after the data were gathered and frequencies were summed across the categories.

The Tally Form (see Figure 2) was developed for the quantification of the adequacy of the teacher's substantive presentation variables. From the set of instructional objectives, the investigator derived fifteen knowledge components, or generalizations, which encompassed the essential information required to know the meaning of the concept and to fulfill the tasks implied by the objectives (see Figure 3). These generalizations defined the adequacy of knowledge coverage variable. A list of nine essential concept labels was also derived from the instructional objectives to define the adequacy of concept label coverage variable.

The trained observer listened to the audio-tape recorded lessons for verbal indications of the teacher's explicit inclusion of each of the knowledge components and of each of the concept labels. The essence of the meaning of each of the knowledge generalizations was sufficient; it was not necessary for the teacher's terminology to be exactly that of the listed generalizations. A tally was made on the adequacy Tally Form by the observer for each time the teacher actually stated each of the knowledge components and each of the concept labels. Thus, the frequency of occurrence of each generalization could be computed for each lesson separately or for the combined lessons.

A modified category-rating system was also developed (see Figure 4) to quantify the accuracy and relevance variables, the three semantic variables, and the enthusiasm and class control variables. A seven step graphic scale was employed with the anchors specifying the quality of performance, ranging from (1) very poor performance to (7) very good, outstanding performance on each particular variable.

The observers were trained to listen to the audio-tape recorded lessons for specific examples of teacher behavior which would be indicative of the characteristic underlying each variable. For example, two variables were concerned with the accuracy of concept-specific substantive behavior: accuracy of concept definitions and accuracy of concept examples. As the observer heard the teacher provide either a concept definition or an example, the observer made an assessment of the degree of accuracy of the teacher's statement. The observer made ratings along the continuum for each variable to indicate judgments made during the audio-tape critique. At the end of the lesson, a summary rating, or an average of the ratings made for specific behavioral instances, was made.

The rating scale was constructed in this manner in an attempt to increase the level of objectivity of the ratings by requiring the observer to focus on specific units of behavior. However, the attem[illegible]tual frequency and quality of behavior proved problematic. The obse[illegible]ed with the doubly-difficult task of (1) judging whether or not a particular kind of behavior occurred and determining what kind of event it was (a qualitative judgment); and of (2) assessing the degree to which a particular quality was present in the behavioral sample (a quantitative judgment). This task was especially difficult with the semantic variables where it was frequently difficult to find discrete, easily identifiable examples of behavior which related to the variables as defined. The observer's task was considerably less difficult with the accuracy and relevance variables; the specific units of behavior -- the concept definition, concept example, and teacher utterances -- were overt, discrete, and easily distinguishable. However, the difficulty of specifying behaviors representing the different levels of quality implied in each variable contributed to observer bias.

Item content. Item content specifies the level of inference demanded from the data and from the observers. Rosenshine distinguishes between low and high inference responses. Low inference responses or variables tap the directly observable, specific, explicit phenomena of the environment. High inference responses or variables ask the observer to make a wholistic, global judgment about the meaning of what is occurring. Low inference measures are commonly thought to maximize the objectivity of the data; that is, the more molecular the variable, the more objective the measurement can be. The recording of a sequence of interaction can be best captured by employing a low inference system. High inference variables are usually used to assess general teacher characteristics not easily measured by discrete behaviors.

At one time, rating forms were practically defined as requiring high inference judgments whereas category systems implied the use of low inference behaviors. This is no longer the case. Items requiring a low level of inference can be found on rating forms while items on sign or category systems may demand a high or moderate level of inference (Rosenshine, 1973). The critical dimension regarding the level of inference demanded deals with how directly observable the relevant behavior is. How much judgment must an observer undertake in order to code a particular behavior?

Assessing the three instruments employed in this study in terms of item content, one finds a range from high inference (enthusiasm) to moderate inference (adequacy of knowledge coverage variable) to low inference (off-task talk) variables.

Many of the OSCI items demanded at least a moderate level of inference from the observer. The observer had to be knowledgeable enough regarding the instructional content to be able to discern a positive or negative concept example or a concept definition when provided by the teacher. In very few cases would the teacher signal that a particular statement served the purpose of a definition, an example, or a review. Rather, the observer's task was to infer the intent of the teacher's verbal behavior as specified by the categories of the observational system. With OSCI, the observer must be engaged in a content as well as a process analysis simultaneously. Several category decisions are a function of previously occurring content and processes; thus, considerable information must be stored by the coder who successfully employs the OSCI. On the average, 2½-3 hours were necessary for the coding of each 45 minute lesson. This indicates at least a moderate level of difficulty and certainly more than low inference judgments.

The adequacy of knowledge coverage variable on the Tally Form also demanded at least a moderate level of inference from the observer. While the observer was looking for discrete, overt, concrete behaviors - the appropriate teacher utterances were not consistently easily observable. The observer's task was to assess the meaning of the teacher's explicit statements containing a knowledge component and to match that meaning to one (or none) of the generalizations stated on the Tally Form. The observer's task on the adequacy of concept label coverage variable, by contrast, was low inference. The teacher either did or did not say the name of the concept.

The variables on the Ratin[illegible]n demanded a moderate to high level of inference from the observers. Th[illegible] and relevance variables involved comparing the teacher's behavior with a standard body of knowledge. This meant that the observer had to be a subject matter expert in order to accurately assess these variables. The enthusiasm and on-task variables demanded high inference judgments; attempting to identify specific examples of behaviors illustrative of these two variables to count during the lesson proved to be difficult. Revision of the Rating Form should include renaming the enthusiasm and on-task variables possibly as paired rating scales (stimulating vs. dull; alert vs. apathetic; businesslike, task-oriented vs. laissez faire) to be assessed once at the end of an observation or as ratings which are made every five minutes (or so) during the observation.

Coding format. A single coding format records a behavior on one dimension (Borich, 1977) while, with a multiple coding format, a behavior is coded according to any number of dimensions (Rosenshine, 1973). The OSCI has a type of multiple coding format: behaviors are subdivided into (a) type of speaker -- teacher or student, (b) type of communication -- question asking, or information giving and (c) relationship of communication to task-on-task vs. off-task talk. Behaviors are coded only once, however, but are recorded as they occur sequentially.

Unit of analysis. The unit of teacher behavior which is coded on OSCI is the purposive move. A purposive move refers to an activity performed by the teacher which has the apparent function or effect of progressing the lesson from one substantive or process point to another. Each purposive move is a statement or question which expresses a more or less complete idea and which serves a single function, as defined by the categories of the observational system. Every change of purposive move or speaker necessitates a new coding entry. (See Figures 5 and 6 for examples of coded teacher and student verbal behavior using the OSCI.) A purposive move should be distinguished from an utterance; and utterance is a verbal expression performed by one person at a given time. An utterance may contain one, severa[illegible] ive moves. OSCI is able to record the frequency and sequence of purposive moves, but not the duration of each move. To the extent that time spent on particular purposive moves influences student learning, the absence of a duration-weighting mechanism on the OSCI is a source of distortion.

Various units of behavior were necessary for the variables on the Rating Form. The observers attempted to assess definitions given or accepted, examples given or accepted, and teacher utterances for the accuracy and relevance variables.

Reliability of Observation Instruments.

The accuracy of any observational system is partially a function of (a) the consistency of observations among those judging the behavior and (b) the test-retest reliability or the stability of teacher behavior measured across changes in pupils, content, and/or time. Following is a discussion of these two aspects of reliability as they relate to the three instruments employed in this study.

Rater consistency. The investigator and a trained assistant were the observers for this study. The training program consisted of four parts: (1) gaining familiarity with the substantive aspect of the concept specialization; (2) learning the definitions and distinguishing characteristics of the teacher (and student) process variables; (3) practicing coding and rating the process behaviors using pilot test data; and (4) establishing inter-coder agreements. Three eight-hour training sessions were held before the criterion reliability level of 0.80 of observer agreement was achieved for each of the three instruments.

Coder agreement data were gathered by having the two observers independently critique the same audio-taped lesson using the OSCI. A second lesson was critiqued independently by each observer using the Rating and Tally Forms. Coefficients of observer agreement on OSCI were calculated by using the formula proposed by Scott (Flanders, 1965). Scott's coefficient, pi, (π) is determined by the formula

$$\pi = \frac{Po-Pe}{1-Pe}$$

where Po is the proportion of agreement and Pe is the proportion of agreement expected by chance, which is found by squaring the proportion of tallies in each category and summing these over all categories.

Levels of agreement between observer one and observer two at the end of the training period were $\pi = 0.92$ for each of two independently coded audio-taped lessons of forty-five minutes each. At a mid-point in the data coding period, a second coefficient of observer agreement was calculated, utilizing one of the originally coded audio-tapes. Scott's coefficient of agreement was $\pi = 0.90$.

Consistency checks over time were also computed, comparing each observer's degree of agreement with self on the OSCI and the Tally Form. For observer one, $\pi = 0.83$; for observer two, $\pi = 0.86$. Rating Forms were marked almost identically by the two observers on all consistency checks.

Observers were blind to the criterion variable data during the entire coding period. The classroom process data were contained on twenty-two audio-tapes. The tapes were stratified along grade levels represented and then randomly divided into two sets. Each of the observers was assigned a set of eleven tapes to which the OSCI was applied. The observers then exchanged sets of tapes and applied the Tally and Rating Forms to this new set. This procedure was employed to achieve independence of observations between the OSCI and the other two measures.

With experienced classroom observers and with sufficient training, acceptable levels of rater agreement can usually be achieved. That was the case for this study. However, in planning a replication study, reliability can be improved upon by (a) increasing the number of observers, and (b) excluding the investigator from the observer pool. In addition, independence of observations can be increased by having observers apply only one observational system to each classroom sample.

Teacher stability. Is teacher behavior reasonably stable across content, time, and pupils? Which aspects of behavior might be likely to be stable and which types of behavior might be expected to vary across various changes in setting? These remain unanswered questions. If teacher behavior varies widely across conditions, a separate index of teacher skill would have to be constructed for each situation in order to assess teacher effectiveness. Shavelson and Dempsey (1976) report equivocal findings in their review of the generalizability and stability of measures of teacher behavior. Lack of standardization of measures contributes to an inability to draw comparisons across studies. However, in general, it appears that the global, high inference ratings on teacher behavior appear to be more stable than the low inference, counted measures. Rosenshine (1970) reports moderate consistency in teaching behavior when the same material is taught to different pupils. This generalization was summarized from a limited number of studies, however. While it appears that teacher behavior may be moderately consistent over brief periods of time, behavior is less stable over time and across changes in content. Borich (1977) suggests that we may not be tapping the kinds of behaviors which are relatively stable over time and/or presently used instruments may be confounding the data.

For this study, teachers conducted two lessons of forty-five minutes each on consecutive days to the same group of students. One set of instructional objectives was used to guide the teacher's construction of both lessons. The correlations for the low inference, counted variables recorded on OSCI and the Tally Form are shown on Table 1.

Table 1 here

One might expect some degree of consistency of strategic and substantive behaviors given common content and students. However, given the result [illegible] studies in which low inference variables showed little stability even across two lessons, the correlations shown on Table 1 are surprising. A number of teacher behaviors measured by OSCI remained fairly stable: the giving of concept definitions (0.49), the giving of positive concept examples (0.48), signalling a change in the topic (0.45), explaining (0.67), asking a low order question (0.70), asking a high order question (0.48), signalling and changing the topic simultaneously (0.43), and off-task behavior (0.95). Low frequencies of behavior on several of the variables may have contributed to low stability coefficients. The means for the adequacy of content coverage variable represented the average number of knowledge components provided by the teachers in each lesson. Apparently, teachers were very consistent in their provision of knowledge (0.54) and of concept labels (0.69).

One wonders about the generalizability of these concept-related behaviors across time, type of student, type of concept and across the teaching of concepts from other disciplines. It is hoped that similar investigations can be conducted to examine these relevant variables.

The stability coefficients for the rated teacher variables are shown on Table 2.

Table 2 here

Not surprisingly, the coefficients are consistently high. The question is whether this degree of stability is an attribute of teacher behavior or an artifact of the measurement procedure. The dimensions of teacher behavior assessed by the variables on the Rating Form are not intended to be mutually exclusive. It is conceivable that any given teacher could perform consistently well or poorly on each of the variables. However, several problems with the Rating Form variables are apparent; these conditions probably influenced the observers to make subjective and impressionistic assessments.

The rater's task was to make assessments of behavioral indications of each variable. However, relevant information available to the rater varied from one variable to another and from one teacher to another. This variability may have encouraged the rater to be influenced by other, unknown characteristics of the teachers. Also, for most variables, the definitions proved to be inadequate for the range of behaviors encountered.

As shown in Table 3, high intercorrelations are evident for all of the rated variables.

Table 3 here

This makes one cautious about calling each variable by a separate name. One wonders about the intercorrelations of rated variables in other studies which have reported high stability measures for rated behaviors.

For future investigations, confidence in the rated measures can be enhanced by (a) specifically defining each variable, providing example behaviors at points along the scale; (b) increasing the number of independent ratings; and (c) designing high and low inference independently derived measures of the same characteristic. This last point will enable the examination of the degree of correspondence between sets of logically related behaviors.

Validity of Observation Instruments

Most teacher behavior investigations aim at establishing relationships between measures of behavior and a criterion measure. When the measures are collected at about the same time, the effort is one of establishing concurrent validity. Predictive validity implies the ability of the behaviors to relate to achievement over time. Tables 4 and 5 present the teacher behavior correlates of student achievement on concept tasks administered immediately following instruction.

Tables 4 and 5 about here

(See Armento, Beverly, "Teacher Behaviors Related to Student Achievement on a Social Science Concept Test," A paper presented at AERA, 1976 for a discussion of these data).

The validity issue surrounding teacher behavior studies is that of construct validity or the ability of observation systems to measure the teacher or student behaviors they purport to measure. Borich (1977) proposes that observational systems should be able to demonstrate convergent and discriminant validity. That is, a particular behavior measured on one instrument should correlate significantly with a similar or same behavior measured on another instrument. In addition, that correlation should be "higher than either that between dissimilar behaviors on the same instrument or that between dissimilar behaviors measured by different observation coding instruments" (Borich, 1977, p. 20).

In the study being reported, three instruments were employed. No obvious attempt was made to define the relevant variables along different types of scales; thus, minimal data exist to examine the convergent and divergent validity of the measures. However, an example of this procedure can be illustrated.

Methods

		A		B	
		T. behaviors		T. behaviors	
		Gives definition	Off-task Behavior	Adequacy of Content Coverage	On-task rating
		1	2	1	2
A	1	(.49)			
	2	-.06	(.95)		
B	1	.43	-.43	(.54)	
	2	.32	-.71	.75	(.96)

In the above illustration, Method A is OSCI and Method B represents both the Tally and Rating Forms. The teacher strategic behavior, gives concept definition, can be viewed as similar to the adequacy of content coverage. While the on-task rating should be strongly inversely related to the actual counting of off-task behavior, one would not expect the on-task behavior to diverge from A_1 and B_1 variables. Rather, the on-task rating should be measuring behavior contained in each of the A_1 and B_1 variables, and thus can be expected to be a positive correlate of same.

The premises underlying convergent and discriminant validity are: (1) the correlation between the same behavior measured by the same method (reliability) should be higher than (2) the correlation between the same behavior measured by two different methods -- which in turn, should be higher than (3) the correlation between two different behaviors measured by the same method -- which in turn, should be higher than (4) the correlation between two different behaviors measured by two different methods, (Borich, 1977). By using the premises, one can see that relatively good convergent and discriminant validity is indicated for the behaviors, giving definitions, covering content, and off-task behavior. The on-task behavior variable behaves as expected.

A second illustration can be examined:

		Methods			
		A		B	
		T. Behaviors		T. Behaviors	
		Gives example	Off-task Behavior	Adequacy of concept label coverage	On-task rating
		1	2	1	2
A	1	(.48)			
A	2	(-.42)	(.95)		
B	1	.47	-.30	(.69)	
B	2	.61	-.71	.61	(.96)

Again, the convergence of the gives example and adequacy of concept label coverage variables supports the notion that these measures are assessing similar behaviors. Both convergent and discriminant validity are relatively good for three behaviors, with the on-task rating converging with the specific on-task behaviors.

The intercorrelations for the behaviors measured by OSCI and the Tally Form appear to be internally consistent; that is, related and unrelated items correlate as predicted. This cannot be said for the variables assessed by the Rating Form, where all behaviors converge, and thus probably do not measure identifiable aspects of teacher behavior.

<u>Summary</u>. Future examinations of concept instruction should include a broader range of high and low inference measures which are designed to assess the same or similar dimensions of teacher behavior. This provision will enable a more thorough assessment of the construct validity of the instruments. In addition, the semantic variables, in particular, need to be reconceptualized and a more reliable measure developed for their assessment. Each of the variables presently measured by the Rating Form is in need of refinement and redefinition.

The behaviors measured by OSCI and the Tally Form appear to be more accurate assessments of the variables as defined. Several behaviors demonstrating at least a moderate degree of stability also related significantly to student achievement: the adequacy of content coverage and the adequacy of concept label coverage as measured by the Tally form; and the teacher gives concept definitions and gives positive concept examples.

Substantive, semantic, and strategic teacher behaviors can be revised on the basis of this study. However, these basic categories of behavior continue to be viable; changes are apparently needed in the type of measurement employed with a few of the variables.

It is hoped that replication and extension studies will be conducted with the revised instruments to test the generalizability of the more promising substantive and strategic behaviors across changes in pupils, time, and type and content of concept instruction.

References

Borich, Gary D., The Appraisal of Teaching: Concepts and Process, Addison-Wesley, Reading, Massachusetts, 1977, 396 pp.

Brophy, Jere E. and C.M. Evertson, Process-Product Correlations in the Texas Teacher Effectiveness Study: Final Report, Austin, Texas, 1974.

Clark, Cecil D., "Teaching Concepts in the Classroom: A Set of Teaching Prescriptions Derived from Experimental Research," Journal of Educational Psychology -- Monograph, 62: 253-278, June, 1971.

Flanders, Ned A., Teacher Influence, Pupil Attitudes, and Achievement, U.S. Department of Health, Education, and Welfare, Office of Education, 1965, 126pp.

Gage, N.L., "Paradigms for Research on Teaching," pp. 94-141, N.L. Gage, ed., Handbook of Research on Teaching, Chicago: Rand McNally, 1963.

Husen, Torsten, editor, International Study of Achievement in Mathematics: A Comparison of Twelve Countries, Vol. 2, John Wiley and Sons, Inc., New York, 1967, 168 pp.

McDonald, Frederick J., "Beginning Teacher Evaluation Study, Phase II," Educational Testing Service, Princeton, New Jersey, 1976, 21 pp.

Medley, D.M. and Mitzel, H.E., "Measuring Classroom Behavior by Systematic Observation," in Handbook of Research on Teaching, pp. 287-326, edited by Nate L. Gage, Rand McNally, Chicago, 1963.

Rosenshine, Barak, "Objectively Measured Behavioral Predictors of Effectiveness in Explaining," Technical Report Number 4, Stanford Research and Development Center in Teaching, Stanford, California, 1972, 57 pp. (ERIC ED 028-147).

Rosenshine, Barak, "The Stability of Teacher Effects Upon Student Achievement," Review of Educational Research 40: 647-662, 1970.

Rosenshine, Barak and Norma Furst, "The Use of Direct Observation to Study Teaching," pp. 122-183, in Travers, R.M.W., Second Handbook of Research on Teaching, Rand McNally Co., Chicago, 1973.

Scriven, Michael, "Truisms as the Grounds for Historical Explanations," in Theories of History, pp. 443-475, edited by Patrick Gardiner, the Free Press, New York, 1959.

Shavelson, Richard and Dempsey-Atwood, Nancy, "Generalizability of Measures of Teaching Behavior," in Review of Educational Research 46: 553-611, Fall, 1976.

Shutes, R.E., "Verbal Behavior and Instructional Effectiveness," unpublished doctoral dissertation, Stanford University, 1969.

Figure 1
Observational System for Concept Instruction (OSCI)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Teacher Behavior	Defining (·) (?)																					D
	Examples (·) (?) (p,n)																					E
	Signaling																					S
	Reviewing																					Rv
	Answering, Explaining																					A
	Lower Order Question																					L
	Higher Order Question																					H
	Repeating																					Rp
	Rephrasing																					Rh
	Off-task																					OT
	Other, Digression																					O
Student Behavior	Answer (C,L,T,I)																					A
	Ex/Def (c,i)																					ED
	Misunderstands																					M
	Challenges																					C
	Off-task																					OT
	Other, Digression																					O

Symbols

Teacher Behavior:

· represents the teacher's giving of either a concept definition or a concept example.

? represents the teacher's asking for either a concept definition or a concept example.

p represents a positive concept example.

n represents a negative concept example.

Student Behavior:

In response to a teacher's low or high order question, (C,L,T,I) represent the following:

C is a correct response.
L is a logical response.
T is a true . . . but response.
I is an incorrect response.

c represents a correct concept example or definition.

i represents an incorrect concept example or definition.

Figure 2

Tally Form (Short Form for Data Collection)

Knowledge Components

Specialization is the concentration, focusing on a small aspect of some whole . . . . . . ____________

Specializing accentuates and creates differences . . . . ____________

Division of labor implies role differentiation . . . . ____________

Specialization occurs in at least three forms: technological, occupational, and geographical . . . .
t ____________
o ____________
g ____________

Aim of specialization is efficiency . . . . ____________

Specialization allows people and regions to use to best advantage their differences in skill, knowledge, interest, and resources . .
p ____________
r ____________

Specialization necessitates interdependence . ____________

Specialization necessitates trade . . . . ____________

Specialization implies certain problems:

- need for interdependence . . . . ____________
- possible loss of efficiency in one area . ____________
- low transfer of specialized skills . . ____________
- possible boredom . . . . ____________

Concept Labels

Specialization . . . . ____________

Division of labor . . . . ____________

Technological specialization . . . . ____________
Occupational specialization . . . . ____________
Geographical specialization . . . . ____________

Interdependence . . . . ____________
Dependence . . . . ____________

Trade . . . . ____________
Exchange . . . . ____________

Figure 3
Generalizations Defining the Basic Knowledge
Implied by the Instructional Objectives

1. Specialization is the concentration or focusing upon some small aspect of a defined whole.

2. The process of specializing creates and accentuates differences.

3. Division of labor implies role differentiation.

4. Specialization occurs as the level of technology is differentiated to replace human resources.

5. Specialization occurs as human roles are differentiated in occupational endeavors.

6. Specialization occurs as geographical regions serve differentiated functions.

7. The major aim of any of the three forms of specialization is increased efficiency, or the production of more from fewer resources.

8. Specialization allows people to use to best advantage their differences in skill, knowledge, interest, and resources.

9. Specialization allows people to use regional differences in natural resources to best advantage.

10. Specialization necessitates interdependence.

11. Specialization necessitates exchange or trade.

12. Specialization implies the need for interdependence; this can be a problem.

13. When one specializes in one aspect of production, it is likely that one will lose efficiency in other areas of production.

14. There is often a low degree of transfer of specialized skills and capabilities from one aspect of production to another.

15. Certain repetitive specialized tasks often bring the problem of boredom.

e 4

Rating Form asurement of High Inference
rocess Variables

ating Code:

1 = ver performance on this variable
2 = poor erf nce on this variable
3 = slight average on this variable
4 = avera mance on this variable
5 = sligh average on this variable
6 = good e on this variable
7 = very standing performance on this variable

1. Accuracy of C Specific Teacher Behavior

a. Definitic or developed by the teacher are accurate.

1 2 5 6 7

b. Examples g accepted by the teacher correctly represent ncept.

1 2 5 6 7

Sub-total--Accuracy ________

2. Relevance of Teac Behavior to Instructional Objectives

Teacher es are appropriate to the achievement of the onal objectives for the lesson.

1 4 5 6 7

Sub-total--Relevance ________

3. Teacher Language

a. The teacher uses a balance of abstract-concrete words.

1 2 3 4 5 6 7

3. b. "Empty utterances are minimal. Complete sentences are utilized.

1	2	3	4	5	6	7

c. "Fl[illegible]ng words" are minimal. The referent for [illegible] is seldom in doubt, pronouns clearly [illegible]er [illegible]r antecedents.

[illegible]	2	3	4	5	6	7

Sub-total--Semantics ___________

4. The teacher expresses enthusiasm and interest in the content of the lesson.

[illegible]	2	3	4	5	6	7

Sub-total--Enthusiasm ___________

5. The teacher displays an on-task approach.

[illegible]	2	3	4	5	6	7

Sub-total--On-task ___________

TOTAL ___________

Figure 5
An Example of Coded Teacher Strategic Behavior
Using the Observational System for Concept Instruction

			...	2	3	4	5	[illegible]	8	9
Defining	(?)					•				
Examples	(?)	(p,n)					+/p	+/p...		
Signaling					⊙					
Reviewing				•						
Answering										
Lower ?									•	
Higher										
Repeating										
Rephrasing										
Off-task										
Other										

In this example of coded teacher verbal behavior, the teacher reviews the major ideas already presented in lesson (2), signals a change in the topic (3), changes the topic (represented by the circle in #3), gives a concept definition (4), gives three positive concept examples (5,6,7), and then asks a low order question (8). All of the above purposive moves occurred in one utterance.

Figure 6

An Example of Coded Teacher-Student Process Behavior Using the Observational System for Concept Instruction

		..	2	3	4	5	6	7	8	9
Teacher	Defining (·) (?)									
	Examples (·) (?) (p,n)		?/p			·/p				
	Signaling									
	Reviewing									
	Answering									
	Lower ?									
	Higher ?									
	Repeating									
	Rephrasing									
	Off-task									
	Other									
Student	Answer (C, L, T, I)									
	Exam./Def. (c,i)			c/c	c/c					
	Misunderstands						•			
	Challenges									
	Off-task									
	Other									

The following teacher-student interaction is represented by the coding shown in Figure 6.

T: Can you name an example of occupational specialization that we haven't mentioned yet? (In move 2 the teacher asks for a positive concept example.)

S: My father works on the assembly line at RCA and specializes in installing tubes in color T.V. sets. (In 3, the student gives a correct concept example).

S: My big sister is studying to be a nurse. That's a specialized kind of job. (In 4, a student gives a correct concept example.)

T: Yes, you're both correct. Another example might be occupational therapists. (In 5, the teacher gives a positive concept example.)

S: I don't understand. What's that? (In 6, the student expresses misunderstanding.)

Table [illegible]

Stability of Teacher Low Inference Behaviors across Two Social Science Lessons Taught on Consecutive Days (N=22)

Teacher Low Inference Process Variables	First Lesson $\bar{X}$	First Lesson SD	Second Lesson $\bar{X}$	Second Lesson SD	Stability Coefficient
Gives Concept Definition	2.[illegible]	2.61	2.09	3.07	0.49*
Asks for Concept Definition	[illegible]	1.44	3.00	3.[illegible]	0.33
Gives Positive Concept Example	6.[illegible]4	7.51	6.41	6.[illegible]	0.48*
Gives Negative Concept Example	0.23	0.53	0.23	0.[illegible]	-0.19
Asks for Positive Concept Example	[illegible].18	1.44	0.91	1.[illegible]	0.24
Asks for Negative Concept Example	- Does Not Occur -				
Signals a Topic Change	[illegible].64	4.48	3.64	[illegible]	0.45*
Reviews, Summarizes Main Idea	4.82	3.14	5.00	[illegible]	0.14
Explains, Answers	8.50	7.09	8.50	7.4[illegible]	0.67**
Asks Low Order Question	45.00	21.56	36.91	18.77	0.70**
Asks High Order Question	7.09	6.44	7.27	7.[illegible]	[illegible].48*
Repeats Question After Student Response	2.59	2.81	2.18	2.36	0.25
Rephrases Question After Student Response	4.50	3.43	3.27	[illegible].07	0.35
Signals and Changes the Topic Simultaneously	[illegible].64	3.90	3.32	[illegible].19	0.43*
Changes Topic With a Low Order Question	[illegible].27	7.12	4.27	4.37	0.39
Changes Topic With a High Order Question	1.[illegible]	2.63	0.91	1.11	-0.01
Asks Pairs of Questions	[illegible]	9.62	7.05	6.24	0.33
Tells Students to Stop Irrelevant Behavior	[illegible]	16.86	9.14	[illegible]3.34	0.95**
Other	1.[illegible]	1.61	1.32	1.39	-0.03
Adequacy of Content Coverage	19.[illegible]	10.56	17.55	[illegible].08	0.54**
Adequacy of Concept Label Coverage	11.[illegible]	7.31	10.64	5.[illegible]	0.69**

* $p \leq .05$
** $p \leq .01$

Table 2

Stability of Teacher High Inference Behaviors Across Two Social Science Concept Lessons Taught on Consecutive Days (n=22)

Teacher High Inference Process Variables	Stability Coefficient
Accuracy of definitions	0.79**
Accuracy of examples	0.83**
Relevance of behavior to objectives	0.91**
Balance of concrete/abstract terminology	0.9[illegible]**
Use of complete sentences	0.8[illegible]**
Proper use of pronouns	0.85**
Displays enthusiasm	0.88**
Establishes control over learning situation	0.96**

* $p \leq .05$
** $p \leq .01$

Table 3
Zero Order Correlation Matrix for Rated Teacher Process Variables

	2	3	4	5	6	7
1. Defines Concept accurately	.94	.87	.78	.76	.70	.89
2. Provides accurate concept examples		.95	.84	.81	.77	.90
3. Expresses behavior relevant to objectives			.92	.85	.83	.91
4. Achieves a balance between concrete and abstract terminology				.84	.75	.84
5. Uses complete sentences and clear pronoun referents					.83	.86
6. Displays interest and enthusiasm over the content of the lesson						.87
7. Displays primarily on task, low noise behavior						

TABLE 4. PEARSON PRODUCT MOMENT CORRELATION COEFFICIENTS FOR LOW INFERENCE TEACHER PROCESS VARIABLES AND CLASS RESIDUAL MEAN GAIN SCORES (n = 22)

Teacher Process Variables	Correlation	Level of Significance
Gives concept definition	.426	.02*
Asks for concept definition	.190	
Gives positive concept example	.497	.009**
Gives negative concept example	.225	
Asks for positive concept example	.177	
Asks for negative concept example	Does not occur	
Signals a topic change	-.008	
Reviews, summarizes main ideas	.376	.04*
Asks lower order questions	-.098	
Asks higher order questions	-.047	
Repeats question after student response	-.309	
Rephrases question after student response	-.257	
Signals and changes the topic simultaneously	-.171	
Uses review-signal-shift pattern	.122	
Changes topic with a low order question	-.301	
Changes topic with a high order question	.162	
Asks pairs of questions	-.013	
Tells students to stop irrelevant behavior	-.050	
Other, including substantive digressions	-.043	
Adequacy of content coverage	.456	.01**
Adequacy of concept label coverage	.528	.006**

*p ≤ .05

**p ≤ .01

TABLE 5. PEARSON PRODUCT MOMENT CORRELATION COEFFICIENTS FOR HIGH INFERENCE TEACHER PROCESS VARIABLES AND CLASS RESIDUAL MEAN GAIN SCORES

Teacher Process Variables	Correlation	Level of Significance
Accuracy of concept definitions	.326	
Accuracy of concept examples	.376	.04*
Relevance of behavior to objectives	.370	.04*
Balance between concrete and abstract terminology	.381	.04*
Uses complete sentences and correct pronouns	.274	
Expresses interest and enthusiasm over content of lesson	.478	.01**
On-task, low noise behavior	.279	

*$p \leq .05$

**$p \leq .01$