DOCUMENT RESUME

ED 033 081 SP 003 209

By-Claus, Karen E.

Effects of Modeling and Feedback Treatments on the Development of Teachers' Questioning Skills. Technical Report No. 6.

Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Pub Date Jun 69
Note-55p.
EDRS Price MF-$0.25 HC-$2.85
Descriptors-Elementary School Teachers, *Feedback, Models, Preservice Education, *Prompting, Questioning Techniques, *Teaching Models, *Teaching Skills, *Training Techniques, Video Tape Recordings

The effects of cueing procedures in modeling and in feedback treatments on the acquisition of teacher questioning behavior were investigated in order to determine if, as hypothesized, providing cues from a supervisor on the desired behavior during modeling and feedback treatments would increase the frequency of a teacher's use of higher-order questions (those that encourage more complex pupil thinking). A presentation-practice-feedback instructional paradigm was used within a microteaching format: (1) 10-minute pretest teaching session, (2) viewing of videotaped models displaying the criterion behavior, (3) practice in matching the model behaviors in a videotape microteaching session, (4) viewing the playback, (5) posttest teaching session. The 40 preservice elementary teachers were randomly assigned to four experimental conditions in a 2 x 2 fixed-effects design: (a) no cues in either modeling or feedback conditions, (b) cues only in feedback conditions, (c) cues only in model treatments, (d) cues in both training conditions. Transcripts of the teaching session were rated independently, using an eight-category system developed to rate the dependent variable questions. Analysis of variance produced findings supporting part of the hypotheses: it was concluded that in the acquisition of a complex teaching skill, observational learning with cueing is more effective than feedback, with or without cueing, in producing desired behavior change. (JS)

STANFORD CENTER FOR
RESEARCH AND DEVELOPMENT
IN TEACHING

Technical Report No. 6

EFFECTS OF MODELING AND FEEDBACK TREATMENTS ON THE DEVELOPMENT OF TEACHERS' QUESTIONING SKILLS

Karen E. Claus

School of Education
Stanford University
Stanford, California

June 1969

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the Office of Education (Contract No. OE-6-10-078, Project No. 5-0252-0501).

Table of Contents

EFFECTS OF MODELING AND FEEDBACK TREATMENTS ON THE DEVELOPMENT OF TEACHERS' QUESTIONING SKILLS

Karen E. Claus[1]

The Questioning Skill

This study was conducted as part of the Technical Skills of Teaching Project of the Stanford Center for Research and Development in Teaching. The purpose of the research was (a) to define a central teaching skill which stimulates pupil inquiring behavior within the context of student-teacher dialogue and (b) to train beginning teachers to use this skill. Questioning was selected as the teaching skill.

Empirical evidence has shown that teachers tend to use mainly factually oriented questions in their classroom discourse. Creative problem solving, however, implies complex questioning skills. On the assumption that complex or higher order problem solving skills are learned and can be sharpened through practice, the first step to higher order problem solution is the asking of a higher-order question. Since teachers have shown that they do not utilize this type of question, an important teacher training goal should be to train them in its use. The present study utilized an empirical approach in which training methods previously demonstrated to be effective were manipulated experimentally in order to determine the most efficient training combinations.

The Type of Questions Teachers Ask

Investigations by Flanders and Amidon (1961) and Bellack, Kliebard, Hyman and Smith (1966) have shown that over half of formal classroom communication consists of some form of teacher question followed by a student response. The overwhelming majority of these questions are designed to elicit recognition or recall of factual information. Dodl (1965) identified and studied five types of classroom question-answer patterns and found that information-seeking questions were by far the most typical. In earlier studies, analysis of tapescripts of the classroom discussion of gifted children revealed that routine and cognitive-memory questions were more common than questions which elicited convergent, evaluative, or divergent thinking (Aschner, 1963; Aschner, Gallagher, Perry, Afsar, Jenner & Farr, 1965; Gallagher & Aschner, 1963; Gallagher, 1965 a & b). Guszak (1966) analyzed questions asked in second-, fourth-, and sixth-grade reading circles and found that the factual aspects of reading comprehension

---

[1]Research Psychologist, University of California, San Francisco Medical Center. The research reported here was carried out while Dr. Claus was a Research Assistant at the Stanford Center for Research and Development in Teaching.

were emphasized most. These findings are not new. Almost sixty years ago Rommiett Stevens used stenographic reports of classroom discussions across several subject areas to show that well over 50 percent of the total number of questions asked were memory questions (Stevens, 1912, pp. 45-48).

There is evidence that pupil communications, too, mirror this preoccupation with facts. Davis and Tinsley (1967) found that factual questions dominated teacher question types and that the correlation between the cognitive level of questions asked by teachers and those asked by pupils was a positive .90. Hudgins and Ahlbrand (1967) reported that, while 56 percent of teacher overt verbal behavior was concerned with factual content, pupil communications in the same classrooms emphasized factual matters 81 percent of the time.

Bloom described factual knowledge acquisition as the most common educational objective in American education (1956, p. 28). As the descriptive studies cited above illustrate, the classroom discourse of most teachers tends to reinforce this goal. There are several possible explanations. First, teachers may avoid higher-order questions because of the difficulty involved. To ask a question that elicits pupil behaviors such as seeking relationships or making judgments, the teacher must understand the processes involved in this type of thinking. Guszak (1966) found diminishing congruency between teacher questions and pupil responses as grade level increased. He suggested that this might be due to upper grade teachers' lacking the answers to their own questions. It was stated bluntly by Butler: "When we say that a teacher's questions are poor, we actually mean that her knowledge and thinking are poor" (1939, p. 195).

Second, it may be unrewarding for a teacher to ask a question which demands difficult intellectual operations on the part of the pupil. Relationship questions are not easily answered, and teachers may encounter resistance when these questions are used. This is illustrated in a story about John Dewey's visit to a class in which he asked, "What would you find if you dug a hole in the earth?" He received no response so he repeated the question, again obtaining only silence. To this the teacher chided, "You're asking the wrong question." Then she asked, "What is the state of the center of the earth?" In chorus, the reply came back, "Igneous fusion" (cited in Bloom, 1956, p. 29). Taba (1966) found that the most marked discrepancies between what teachers sought and what they elicited from students occurred at the highest cognitive levels. One interpretation of this finding was that teachers did not "push for the limits" but asked only what they knew the students could answer.

Another reason that teachers do not use higher-order questions could be that they simply do not know how. Only recently has progress been made in the analysis of cognitive operations, with development of hierarchical systems of classification such as those of Bloom (1956) and Guilford (1956)

in which memory or "knowledge acquisition" is seen as basic and prerequisite to more complex thinking which could be induced by selective questioning techniques. It is not surprising, given the notorious sluggishness of educational innovation, that this information has not filtered into teacher training programs and classroom usage.

The Importance of Higher-Order Questions

Higher-order questions presumably engender critical or reflective thinking, which is included as a goal in virtually all statements of educational objectives. General treatments of teaching techniques consider the question as a cue to the thinking process (Burton, 1962, Ch. 18; Rivlin, 1961, Ch. 7; Butler, 1939, Ch. 10). Dewey's emphasis on transactional experiences or "learning by doing" stresses problem solving in which the proper question sets the stage by introducing a problem. As the first step in the reflective thinking process a question can be considered as the verbalized shorthand form of a problem.

Because of its powerful inquiry-initiating characteristics, the higher-order question may be a key teaching tool in "heuristic teaching" styles, those constellations of attitudes and techniques by which the teacher actively helps students seek knowledge and understanding on their own terms. As a questioner the teacher is seen as the guide and stimulator of pupil inquiring behavior. Such behavior proceeds as the result of internal self-directing processes initiated by a question. It presumably provides practice in higher-order thinking.

The assumption that proper questions will elicit higher-level thinking has not been empirically demonstrated, although Taba (1966, pp. 119-219) maintained that the use of questions which call for relationship thinking encourages pupil higher level responses. Similarly, Hudgins and Ahlbrand (1967, pp. 87-96) reported a significant positive relationship between the level of intellectual demand imposed by a teacher and the level of pupil thinking evidenced by student utterance.

Most of the studies of teacher-pupil classroom communications have been descriptive. The research of Bellack and his associates (1966); Hughes (1959); Gallagher (1965a); Aschner (1963); and Flanders and Amidon (1961); all sought to classify and code teacher verbal behavior as it occurred in the classroom situation. Other work has attempted to classify teacher-pupil interaction in terms of episodes or instructional units which are initiated by questions (Meux & Smith, 1964). Few descriptive studies have concentrated exclusively on teacher questioning behavior. Dodl (1965) classified teacher and pupil patterns of questioning behavior by type of responses given before and after a question is asked.

A few recent studies have attempted to change teacher questioning habits by some form of training. Taba (1966) compared teachers who had been trained to use a specific set of relationship questions with teachers who were

untrained and found that the trained group more frequently elicited abstract levels of thought from pupils than did the untrained teachers. Untrained teachers surpassed trained teachers only in seeking descriptive or factual information. Another study found that beginning secondary intern teachers could be trained to use a larger percentage of higher-order questions through the use of videotape or written models and microteaching sessions (Berliner, McDonald, Allen & Sobol, 1967). These studies suggest that higher-order questioning is a trainable skill.

A Model for Teacher Training

The instructional processes of this study are described by a paradigm combining adaptions of two models proposed by Glaser (1962) and McDonald (1965). Figure 1 shows the conceptual and operational framework of the teacher training program employed.

The Instructional Process Paradigm

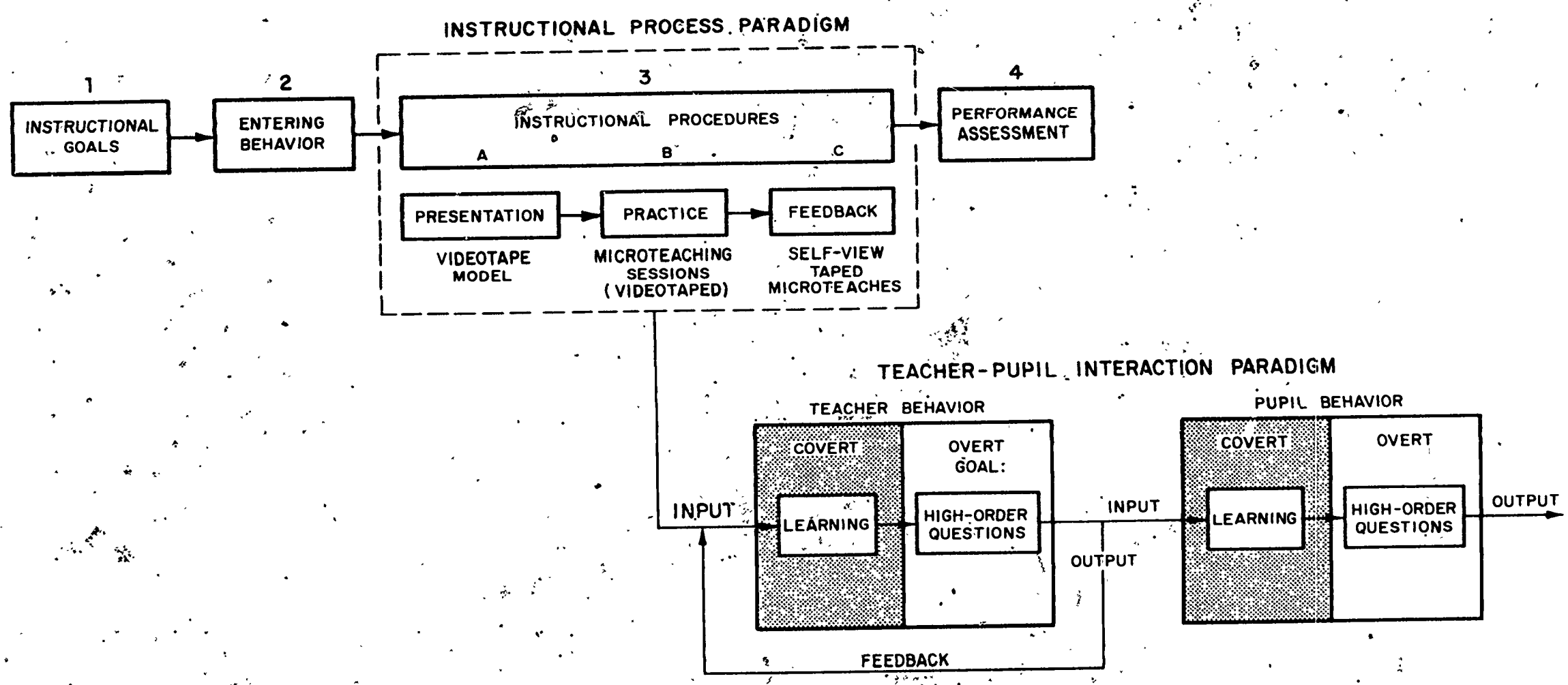
The instructional process paradigm describes the training procedures which became the input to the cybernetic system representing the teacher trainee in the training session.

Components of an instructional system. Glaser breaks the process of instruction into five basic components (Glaser, 1962, pp. 5-30). Initially, there is the setting of goals which will guide and establish the instructional procedures. Second, the student enters the system with an initial skill and knowledge repertoire, a particular set of aptitudes, and a prior educational background. Next, instructional procedures are employed to modify or guide the behavior of the student. Fourth, the terminal behavior repertoire of the student is evaluated in terms of the instructional goals. Over the basic instructional system, Glaser superimposed a research and development component which contributes to each of the other sequential segments. The top portion of Figure 1 represents a modification of Glaser's paradigm with the instructional procedures component expanded to focus on the instructional episode as described below.

The instructional episode. The learning experience occurring in the instructional procedures component of Glaser's model can be described as an instructional episode consisting of three components or stages: a response-guidance phase, a response-practice phase, and a feedback phase (McDonald, 1965, pp. 89-91). In Figure 1 these three phases are labeled (1) presentation, (2) practice, and (3) feedback. In the presentation phase, the learner receives some form of guidance as to the response to be learned. The practice phase entails active attempts on the part of the learner to perform the desired response. The feedback phase involves some form of information provided to the learner concerning the correctness of the response performance. In this study practice was held constant while presentation and feedback were modified experimentally.

FIGURE 1

A Model for Teacher Training Showing the Interrelationship of the Instructional Process Paradigm and the Teacher-Pupil Interaction Paradigm Within the Components of an Instructional System



There are many ways of implementing the elements of an instructional episode. This study chose to focus on modeling as the mode of presentation of the criterion behavior rather than telling the interns what the behavior would be, having them read written descriptions, or having them discuss and determine for themselves what the behavior was. The practice phase was implemented through microteaching sessions in which the interns practiced matching the model behaviors they had just seen. The feedback phase was implemented through self-view sessions wherein the trainees viewed their own performance to see how closely they actually matched the model behavior.

The Teacher-Pupil Interaction Paradigm

The teacher-pupil interaction paradigm describes the process by which a teacher, as a cybernetic system, could influence the behavior of a pupil, as another cybernetic system. This process paradigm could be indefinitely extended to include a network of relationships between pupils as well as between pupils and other agents in an instructional situation. It could also describe chains of "teacher-teaching--pupil-being-taught" episodes such as those suggested by Smith (1960). In this study it was used to describe a modeling-imitation process by which teacher trainees could learn to use higher-order questions by matching the behavior of a model teacher. It was also used to describe processes by which a pupil learns to ask higher-order questions by matching behaviors of a classroom teacher, or gains practice in higher-order thinking through external and internal feedback occurring as a result of answering teacher higher-order questions.

The dual input-output model shown in the lower portion of Figure 1 is based directly on a cybernetic decision-making model (McDonald, 1965, p. 60) and describes the modeling-imitation sequence of the teacher and pupil learning. Learning, which corresponds to the "plans" segment of the McDonald cybernetic model, can be seen as an intervening variable (Smith, 1960, p. 236). This interpretation allows for focus on a single teaching strategy or skill and a single pupil behavior. Conceptually, it can account for a stimulus-response (S-R) situation or for a cognitive restructuring situation (R-R).

Viewing learning as an intervening variable allows an investigator to focus on the overt teacher and pupil linguistic behaviors which are observable and codable and especially suited to recording media such as the videotape recorder and observational rating scales.

In this study, teaching was viewed as a repertoire of component behavioral skills which could be learned and practiced separately and then integrated into a complex teaching strategy by the professional teacher. Component task analysis requires the teacher to focus on a particular skill during the preactive phase of teaching where goals and strategies are determined (Jackson, 1966). It logically should then be easier for the teacher to utilize the skill in a deliberate, more effective way during the interactive phase of teaching in the classroom.

Independent variables in the input of the teacher training segment of the paradigm could be of two kinds: internal and external, with the internal variables consisting of overt and proprioceptive feedback from the pupils, and the external emanating from outside agents such as texts, supervisors, or performance rating scales. This study focused on the external variables or the cues provided by the supervisors in the model and feedback sessions of the training sequence. A "cue" is a stimulus (auditory, visual, or of some other dimension) which directs attention to the salient features of the model immediately prior to the model's response. In this study the salient features of the model presentation were determined by training objectives.

The Relevance of Observational Learning Theory

The paradigms of this study approach the training of a complex teaching skill from the framework of observational learning theory. In terms of training a complex skill, modeling appears to be particularly successful in two types of cases: where novel responses are involved, and where the desired responses are too weak or too infrequent to be directly reinforced. Many complex human skills are learned more readily by observation and imitation than by other methods such as successive approximation (Hilgard & Bower, 1966, pp. 534-538). This is particularly true of novel or complex social behavior for which there is no reliable eliciting stimulus except cues provided by others as they exhibit the behavior (Bandura & Walters, 1963). In a series of experiments with young children from 1961 to 1966, Bandura and his associates studied the processes by which models transmit behavioral repertoires, change existing response patterns, or provide cue responses for later elicitation of specified action strategies. These studies have shown that the observation of models has had important effects on the behavior of observers.

In a study testing the effects of modeling and social reinforcement, Bandura and McDonald (1963) found that models were more effective in changing moral judgment responses than were operant procedures. The study suggested that where the desired responses were initially weak or occurred infrequently, modeling was more efficient in changing behavior than the application of reinforcement. The authors pointed out that even though the behaviors to be learned in a training session might already be in the subject's repertoire, reinforcement procedures alone would not be powerful enough to elicit the desired responses without some procedures for highlighting the salient cues prior to the elicitation of the behavior.

Several explanations have been advanced to explain how modeled behaviors are learned. Miller and Dollard (1941) provided probably the first integrated behavior theory for modeling-imitation concepts in their discussion of three types of identificatory behavior: "Matched-dependent behavior," described as imitation at the subconscious level; "copying behavior," in which the imitator consciously tries to replicate the behavior of a model; and "same behavior," which is conscious behavior attributed to mass or crowd actions.

In their book, Social Learning and Imitation, Miller and Dollard (1941) saw imitative learning as contingent upon reinforcing stimuli administered either to the model or to the observer as he performs a close approximation to the matching response. According to Bandura and Walters (1963), stress upon successive approximation places a severe limitation on the Miller and Dollard theory. Imitative behavior is often acquired in situations where a model's responses are not performed by observers nor is reinforcement applied to models or observers during acquisition. Bandura and Walters suggest that an observer may learn new responses from a model even though he performs no overt responses nor receives direct reinforcement. This applies particularly to the learning of novel responses or complex skills.

Considerable interest in the effects of modeling and the role of reinforcement in imitation learning has been generated since 1960. Mowrer's (1960) work on feedback led to his distinguishing two types of imitative learning. In the first, reinforcement is applied directly to the observer. In the second, the observer receives reinforcement vicariously--i.e., the model receives reinforcement for appropriate responses. Hill (1960) concurred in the view that an observer empathetically experiences reinforcement when viewing a model who is reinforced.

Sheffield (1961) developed a contiguity theory to account for this vicarious acquisition of imitative responses. The contiguous association of sensory events emanating from the model's sequence of behaviors enables the observer to acquire "cue responses" which can elicit behavior corresponding to the model's behavior some time after demonstration. Bandura (1963, 1965a, 1965b) suggested that contiguity accounts for the acquisition of model behaviors by the subject, but that reinforcement of the model rather than of the observer influences the performance of the imitatively learned responses. McBrearty, Marston and Kanfer (1962) showed that the behavior of the observer may match the behavior of the model in some cases, even though no direct reinforcement is administered.

Most of the previously mentioned studies have used children or teenagers as subjects. A series of studies conducted at Stanford University (McDonald & Allen, 1967) were concerned with the training of pre-service intern teachers in complex teaching behaviors. Using the portable videotape recorder as a modeling and feedback device, these studies demonstrated that models of the desired skills were successful in changing teaching behavior in the training sessions. Orme (1966) reported that perceptual modeling (i.e., with videotapes) was more effective in eliciting a desired behavior than any other form of behavior portrayal. Studies by Bandura, Ross and Ross (1963a, 1963b, and 1963c), demonstrated the effectiveness of film-mediated models in transmitting certain behavior patterns, and results from the Krumboltz, Varenhorst and Thoresen study (1967) suggested that presentation of video models in counseling sessions effectively increases the frequency and variety of information-seeking behavior.

The Use of Cueing in Training Studies

The McDonald and Allen (1967) studies found that those experimental groups which experienced supervisor cues during model and feedback sessions showed more improvement than those which experienced the self-view conditions. This was expected on the basis of those studies which stressed the importance of discrimination training, prompting, and cues in learning situations (Cook & Kendler, 1950; Angell & Lumsdaine, 1961; Sheffield & Maccoby, 1961; Wulff & Kraeling, 1961). It therefore seemed appropriate to predict in this study that the supervisor-viewed conditions would produce greatest changes in trainee behaviors. Tolman's (1959) cognitive map and Sheffield's (1961) perceptual blueprinting conceptions suggest that greater emphasis should be placed on viewing the model over viewing the feedback. Supervisor cues and prompting during the model presentation should facilitate the development of cognitive structures or maps which would unify and guide future use of the learned skills.

This prediction is in accord with Bandura's (1968) theory that modeling is an acquisition variable. Imitative learning of a complex verbal behavior involves the processes of sensory registration and symbolic encoding. The sensory registration of visual images could be accomplished by use of a film-mediated model while symbolic encoding (of a subject's attending responses) presumably could be enhanced by overt verbal responses provided by an external agent prior to or during the performance of a criterion behavior.

The Use of Self-View as Feedback

Recent studies on observation of models and self-reward patterns of children have reaffirmed the finding that social rewards dispensed to a model have greater effect on the performance of the observer than if there is no consequence for model behavior (Bandura, Grusec & Menlove, 1967). The conception of vicarious reinforcement presents the question that perhaps use of videotape recordings in the training of intern teachers could be as effective using reinforced models as with live supervisor cues. In addition, the study by Gewirtz and Baer (1958), found evidence that the effectiveness of an adult in reinforcing a child's behavior decreased with the familiarity of the adult. Due to the initial difficulties of training with videotape recordings, a supportive supervisor is considered desirable. The microteaching supervisor in the supervisor-cue situation could have less effect on the behavior of the teacher trainee than the more distant model in the self-view condition (Bandura, Grusec & Menlove, 1967).

The Study

The Dependent Variable: Questioning Behavior

The dependent variable in this study was questioning behavior, divided into higher order and lower order, with the primary focus on the higher-order question.

Although there are many views on the structure of a hierarchy of questions, the most recent proposals are based on the structure of knowledge presented in the Taxonomy of Educational Objectives, 1 (Bloom, 1956). If the Taxonomy is in fact hierarchical, the higher the cognitive category, the more complex the intellectual operations involved in skill used. Since the Taxonomy was originally devised as an ordering system for questions asked on the Graduate Record Examinations, it seemed an appropriate guide for classifying the questions asked by teachers in their classroom discourse.

Questioning behavior was divided into six higher-order and two lower-order categories on the basis of the logical distinction occurring between the broad categories of knowledge and comprehension in the Taxonomy, as shown in Table 1.[2] The gross dichotomous distinction is similar to McDonald's (1965) discrimination between factual and relationship questions and to the Siegels' (Siegel & Siegel, 1967) distinction between factual and conceptual levels of achievement. In the present study, lower-order questions corresponded to factual questions associated with the first and seventh category labels in the Taxonomy as revised, employing the operations of recall or recognition. Higher-order questions corresponded to relationship or conceptual questions, which elicited pupil responses that exceeded the rote-memory level of thinking. Higher-order questions were labeled according to the higher level cognition skill categories listed in the Taxonomy. The terms "higher" and "lower" were assigned on the basis of the Taxonomy and imply no value orientation.[3]

The Independent Variables

Independent variables were the presence or absence of cueing by supervisors during presentations of the videotape model or viewing of self-performance in videotape playback sessions (feedback). Supervisors cued trainees on the criterion behavior, higher-order questioning. Four treatment groups involving various combinations of two training mode variables were used. The two training modes or techniques were (a) "supervisor-view," which was the condition in which the intern was cued by a supervisor while watching either the videotape model or the playback of his own performance, and (b) "alone," which was the condition in which the intern received no external cues in model or feedback sessions.

---

[2] Sanders (1966) changed the Bloom category of Knowledge to Memory and extracted the subcategories of Translation and Interpretation (omitting Extrapolation) from the Comprehension category of the Taxonomy.

[3] Intercorrelations of the eight categories of questions at each teaching session and across all sessions indicated that the categories, as used in this study, were largely independent. The few significant correlations could have been due to chance or to the relationship between frequencies of questions asked rather than to any systematic relations between the categories. Additional information on this point may be found in Claus (1968a, pp. 47-54).

TABLE 1

Summary of the Question Classification System and the Characteristics of Questions

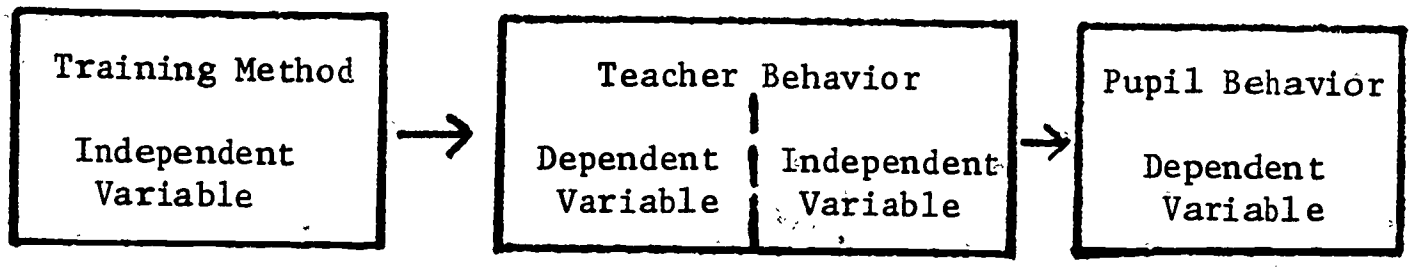
Category	Operation Performed	Examples of Behavior
*1. Memory:	Recognition or recall of factual information.	"Where does Congress meet?"
2. Translation:	Changing ideas from one form of communication to another.	"Can you draw a picture of democracy at work?"
3. Interpretation	Relationships determined between facts, generalizations, definitions, values, and skills.	"How is the weather different in Nome than in Fairbanks?"
4. Application:	Use a whole range of skills with a minimum of instruction or direction; problem solving.	"How would you go about solving the traffic problem in Chicago?"
5. Analysis:	Problem-solving method employed to arrive at conclusion; ideas are separated into component parts and reasons are found to explain relationships between components.	"Why is there more rain on this side of the mountain?"
6. Synthesis:	Imaginative, original thinking, divergent approach.	"How many hypotheses can you suggest to explain why Khrushchev is no longer Premier?"
*7. Lower-order evaluation:	Evaluation of quality evidenced by simple "yes" or "no" response; an opinion.	"Did you like the story?"
8. Evaluation:	Two major steps are involved: (1) set up standards or value structure and (2) determine how closely idea or object meets standards or values.	"Are wars necessary?"

*These are lower-order question categories.

Figure 2 illustrates two interlocking independent-dependent variable sets, schematically represented in terms of a modeling-imitation paradigm.

FIGURE 2

Interlocking Independent-Dependent Variable Sets in This Study



The independent variables in the first part of the modeling paradigm were combinations of training techniques using a videotape recorded model or performance playback with or without supervisor cues. The dependent variable used to test the success of the training method was teacher higher-order questioning behaviors as exhibited in microteaching sessions. These teacher questions then hypothetically became the independent variable for pupil learning of questioning skills. Pupil questions were tabulated in the same manner as teacher questions.

The Hypotheses

The major hypotheses tested in this study were that the frequency of teacher higher-order questioning behavior would be increased more (a) by using cued modeling procedures than by using non-cued modeling procedures, and (b) by using cued feedback procedures than by using non-cued feedback procedures. In addition, it was suspected that trainees receiving the treatment combination in which a supervisor provides discrimination training in both model and playback conditions would show a greater gain in incidence of higher-order questioning than would trainees who received cueing in only one viewing condition or who experienced no supervisor cues at all.

Current modeling theory (Bandura, 1968) would support the hypothesis that the cued model treatments might be more effective than cued feedback in training the use of a skill which was new to the subjects.

The Design

A 2 x 2 fixed effects design was employed to test effects of selected combinations of cueing and non-cueing procedures within modeling and self-view treatments. Subjects were randomly assigned to the four treatment groups summarized in Table 2.

Subjects: The Ss were 40 pre-service elementary intern teachers attending the 1967 summer training session at a California state college in the San Francisco Bay Area.

TABLE 2

Summary of the Modeling and Feedback Treatment Groups

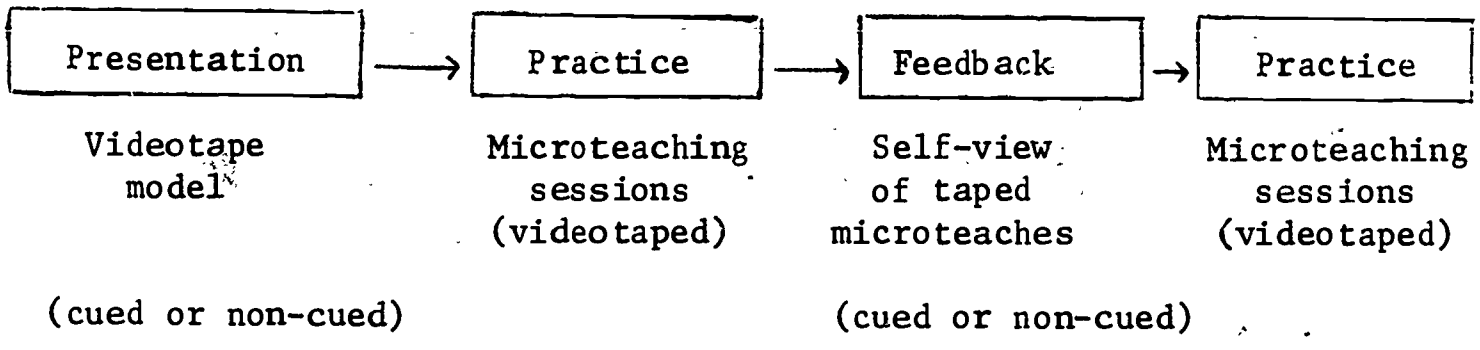
Group Code	Treatments	
	Presentation of videotaped model	Presentation of videotaped microteaching sessions (self-performance feedback)
SS	Trainee viewed videotaped model with experimenter cueing questioning behavior (cued)[a]	Trainee viewed videotape of self-performance, experimenter cueing questioning behavior (cued)
SA	Viewed with experimenter (cued)	Viewed alone (non-cued)
AS	Viewed alone (non-cued)	Viewed with experimenter (cued)
AA	Viewed alone (non-cued)	Viewed alone (non-cued)

[a]Cued indicates that the experimenter orally classified questions immediately following the emission of the behaviors on the videotape, e.g. "That was an application question."

Procedure. Training was conducted in the microteaching format using a modified model-practice-evaluation instructional procedures paradigm (Figure 3).

FIGURE 3

Instructional Procedures Paradigm for Microteaching Sessions



The modeling session consisted of a videotape presentation of a master teacher displaying the criterion behavior, higher-order questioning.[4] Ss then attempted to match the model behaviors in a practice microteaching session. The feedback session consisted of the playback of S's own performance on the criterion behavior. All Ss saw the same model twice and playbacks of their own performance on their second and third trials (teaching sessions). Each S taught a total of four ten-minute lessons to four different groups of seven sixth-grade pupils. The total training sequence consisted of 12 steps summarized in Table 3. Photographs of some of the training steps are shown in Figure 4.

Rating. Typewritten transcripts of the teaching sessions were randomly and independently rated by trained raters. Using the eight-category classification outlined in Table 1, all questions asked by teachers and pupils in a ten-minute lesson period were coded into one of the two lower-order or six higher-order categories. A consensus category score was given to each question based on over 50 percent agreement between the ratings on a question.

Analysis of variance techniques for determining interrater reliability (e.g., Winer, 1962, pp. 124-132) assume the data are in a ratio scale (e.g., frequency scores), that scores are independently drawn from normal treatment populations each having the same variance, and that the errors associated

---

[4]The model displayed as many desired stimuli as possible during the 10-minute session. The rationale for this approach was based upon W. K. Estes' (1959) approach to learning theory which emphasizes the frequency of stimuli as the major independent variable in producing learning rather than a reward-punishment paradigm. Bandura (1968) has also stressed that a high occurrence of observing responses and presence of adequate discriminative cues will enable a learner to acquire relatively complicated linguistic response patterns through observation. A detailed description of the model tape and the taping procedures may be found in Claus (1968a, pp. 34-37).

TABLE 3

Summary of Steps in Treatment by Experimental Group[a]

Treatment Step	Experimental Group 1	2	3	4	Minutes in Treatment
1. Set induction[b]	X	X	X	X	60
2. Teach 1 (pretest)	X	X	X	X	10
3. Model 1	A	A	S	S	10
4. Replan lesson	X	X	X	X	15
5. Teach 2	X	X	X	X	10
6. Self-view of Teach 2	A	S	A	S	10
7. Model 2	A	A	S	S	10
8. Replan lesson	X	X	X	X	15
9. Teach 3	X	X	X	X	10
10. Self-view of Teach 3	A	S	A	S	10
11. Replan lesson	X	X	X	X	15
12. Teach 4 (posttest)	X	X	X	X	10

[a]An "X" indicates that the group underwent this treatment without experimental modification. An "S" designates the experimental treatment wherein a supervisor was present to cue on the desired behavior. An "A" indicates that the subject was viewing the model or playback tapes alone.

[b]Set induction refers to a lecture and discussion period prior to the main body of the treatment sequence wherein all interns received a "set" on the task and information concerning general lesson planning. For some interns there were several days between this set induction period and their skill training.

with any pair of operations are independent. In determining the reliability of the ratings of several raters on a series of questions, the available data consisted of either (1) the codes of the eight different question categories for each question by each rater, or (2) the total frequency of questions within each category for each rater. In the first instance, the category codes were nominal, thus violating the assumption of a ratio scale. In the second instance, the sum of the frequencies of all categories for a given transcript was fixed for all raters as a consequence of the decision to pre-number all questions. Thus the assumption of independent sampling was questionable. Another problem was the fact that distributions of scores in categories of behaviors which occurred very infrequently would probably violate the normality assumption.

It was decided to use percent agreement as a measure of reliability. This agreement figure is represented by the ratio:

$$\frac{\text{Total number of agreements}}{\text{Total number of opportunities to agree}}$$

The method of computation is illustrated in Table 4.

Because transcripts had differing numbers of raters, separate percent agreement figures were calculated for each number of original raters (see Table 5). These figures ignore the additional ratings which were obtained when necessary to derive consensus scores. Table 5 reports the percent agreement found for original ratings using the complete eight-category system.

TABLE 4

Illustration of Computation of Percent Agreement on Question Categorizations for Four Questions Rated by Three Raters

Transcript Question Number	Raters A	B	C	Agreement: Number of agreements/ opportunities to agree
TQ1	6[a]	6	6	3/3
TQ2	3	2	2	2/3
TQ3	5	5	4	2/3
TQ4	1	1	1	3/3
		Totals		10/12
		Percent Agreement		.83

[a]Numbers inside the table are question category code numbers

TABLE 5

Percent Agreement for Ratings of Eight Question Categories by Original Sets of Raters

	Number of Original Raters				
	2	3	4	8	Total
No. of agreements / No. Opp. to agree	6276 / 7816	5500 / 6673	3979 / 5007	106 / 144	15681 / 19640
Percent agreement	80.3	82.4	79.5	73.6	80.8

Results

The Hypotheses

The major hypothesis of this study was tha. the frequency of higher-order questioning behavior could be increased more by using cued modeling techniques than by using non-cued modeling techniques. A second hypothesis was that the frequency of higher-order questioning behavior could be increased more by cued feedback procedures than by non-cued feedback. It was found that cued modeling was significantly ($p < .05$) more effective than non-cued modeling in producing desired behavior change. The feedback treatments produced no significant effect.

Comparison of the number of subjects who stayed the same or went down in the incidence of the desired behavior by treatment group suggests a relationship (Table 6). Group SS, the group with the strongest cueing, ranks highest among the groups who increased their use of the skill. Group AA, the weakest group in terms of cues received, ranked last in increase of the skill. The differences in rankings, however, were not statistically significant.

TABLE 6

Increase or Decrease in Teacher Higher-Order Questioning Behavior from Teach 1 to Teach 4 by Treatment Group

Group	AA (N=10)	AS (N=10)	SA (N=10)	SS (N=10)
No. increasing	5	6	6	8
No. decreasing or remaining same	5	4	4	2

Treatment Effects on Teachers

Initial skill level of the subjects. A two-way analysis of variance on Teach 1 showed no significant initial differences between treatment groups (Table 7). Therefore, it was assumed that treatment groups started with approximately the same level of skill incidence.

Analysis of variance on the difference scores (T4 - T1) of combined categories of questions. The major interest of this study was to produce a change in questioning behavior from the first teaching session to the final teaching session. Therefore, analysis of the difference scores between Teach 4 and Teach 1 on incidence of higher-order questions was the first analysis done.

Two questions were considered before using change scores to test treatment effects: (1) Are change scores unreliable? and (2) Why use change scores in preference to analysis of covariance with Teach 1 scores as the covariate?

TABLE 7

Summary of Two-Way Analyses of Variance for Modeling and Feedback Effects on Pretest and on (Posttest-Pretest) Difference Scores of Teacher Higher-Order Questioning

Source of Variation	SS	df	MS	F
Pretest (T1)				
Modeling Effects (A)	25.600	1	25.600	0.34
Feedback Effects (B)	1.600	1	1.600	0.02
Interaction (A x B)	1.600	1	1.600	0.02
Residual (Error)	2563.200	36	71.200	
Posttest-Pretest (T4-T1)				
Modeling Effects (A)	240.250	1	240.250	5.222*
Feedback Effects (B)	23.361	1	23.261	0.503
Interaction (A x B)	38.028	1	38.028	0.819
Residual (Error)	1,486.000	32	46.438	

*$p < .05$

The unreliability of change scores has been discussed by Lord (1963), who cites two major sources of confusion in studies of change: errors of measurement and regression effects (see also Bereiter, 1963, and Webster & Bereiter, 1963). If two measurements are highly positively correlated, their difference has a smaller variance than if they were independently related. This variance, however, is almost all error and it is in this sense that a difference score is unreliable. Also, any one change score is unreliable due to regression toward the mean of the group. It is therefore difficult to estimate true change for each subject from individual observed change scores. However, where effects of separate treatments are to be compared, estimates of group mean changes can be used. Lord (1963, p. 37) cautions, however, that "analysis of observed gains results in a built-in bias in favor of whatever treatments happen to be assigned to initially low-scoring groups." Random assignment of individuals to treatments is a prerequisite to control for this bias.

The second question concerned the preferability of change score analysis over analysis of covariance. Analysis of covariance assumes that the initial score will influence the amount of improvement. Covariance also assesses treatments in terms of the "average" abilities of the subjects--a significance test between the means of the distributions of the adjusted y-scores. If there is no difference in change due to a subject's initial score, then change scores can be used just as well as covariance.[5] Difference scores assume that the amount of improvement is a constant.

Analysis of variance performed on difference scores has the advantage of avoiding certain assumptions made by analysis of covariance. These assumptions are: regressions are homogeneous; treatment effects and regression effects are additive; and residuals are normally and independently distributed with zero means and equal variances (Winer, 1962, p. 586). In addition, difference scores do not present unusual reliability problems when used to test the significance of differences between groups rather than to describe individual performances.

a. Higher-order questions. Figure 5 shows that all groups increased or maintained their mean frequency of higher-order questions (HOQ). Two-way analysis of variance on HOQ (T4 - T1) difference scores showed that cued model groups increased their use of higher-order questions significantly more ($p < .05$) than did non-cued model groups[6] (Table 7). Cued feedback treatments did not yield significant differences.

---

[5]Analysis of covariance on Teach 4 using Teach 1 as a covariate yielded the same results as analysis of variance on the (T4 - T1) difference scores. Modeling effects were significant at the .05 level. Feedback effects were non-significant.

[6]Due to the variability of some individuals' question-frequency patterns across time, natural log transformations of the raw scores were obtained. It was hoped that analysis of the differences between the Teach 4 and Teach 1 transformations would eliminate possible non-homogeneity and yield a stronger result. Analysis of variance on the differences between logs yielded the same results as analysis of variance on the raw difference scores: Modeling effects were significant ($.05 < p < .10$); feedback and interaction were non-significant.

b. Lower-order questions. As Figure 6 indicates, the non-cued model groups showed a decrease in the mean frequency of lower-order questions (LOQ). Cued groups showed an increase. Analysis of variance on (LOQ) (T4 - T1) difference scores showed that cued model groups differed significantly from non-cued model groups ($p < .01$). as seen in Table 8.

TABLE 8

Two-Way Analysis of Variance on
Teacher Lower-Order Questions

Source of Variation	SS	df	MS	F
Modeling Effects (A)	1092.03	1	1092.03	9.05*
Feedback Effects (B)	235.23	1	235.23	1.95
Interaction (A x B)	50.62	1	50.62	0.42
Residual (error)	4339.90	36	120.60	

*$p < .01$

FIGURE 5

Mean Frequency of Teacher Higher-Order Questions on Teach 1 and Teach 4 for Each Treatment Group. (S indicates cued condition; A indicates non-cued condition.)

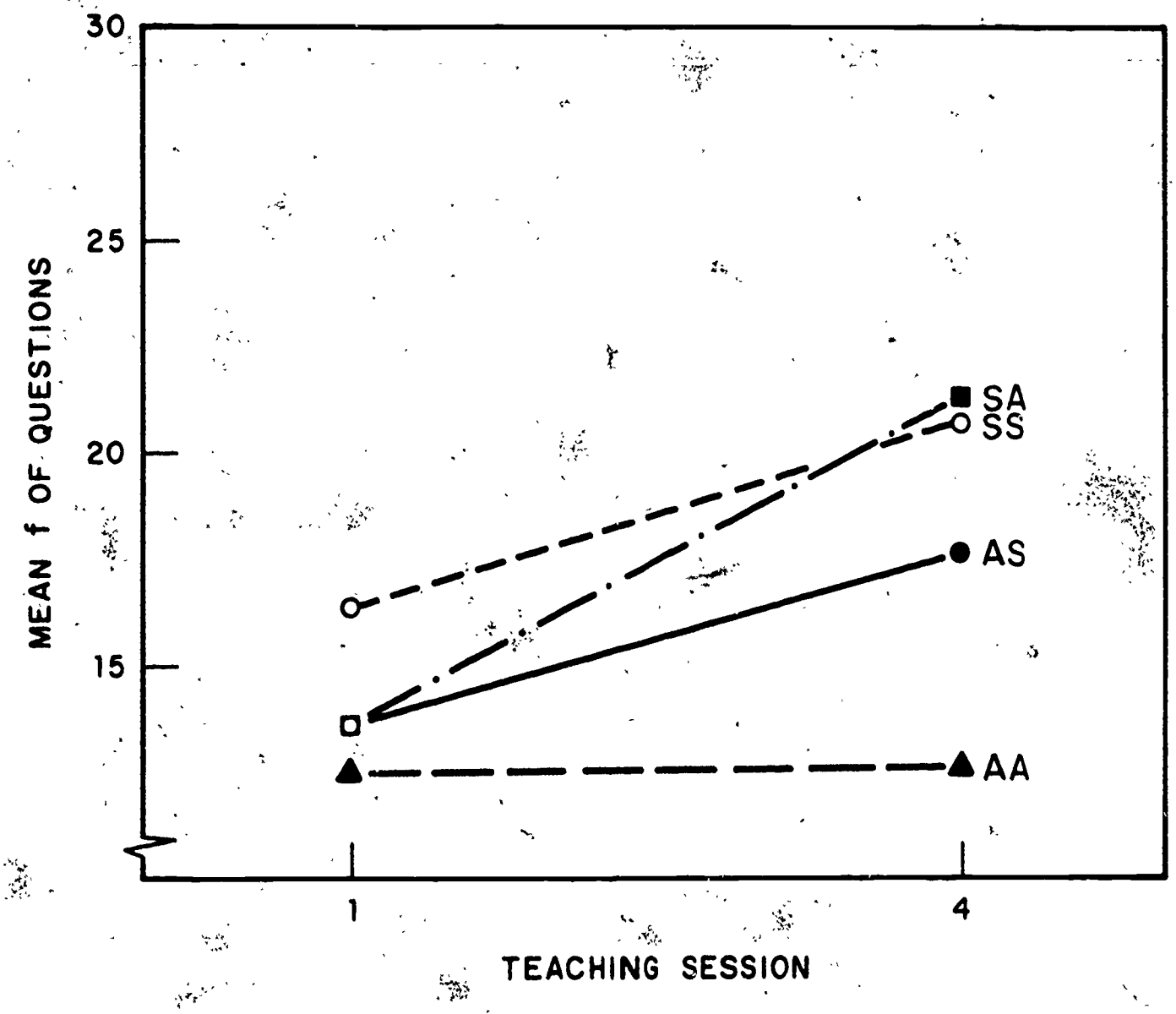
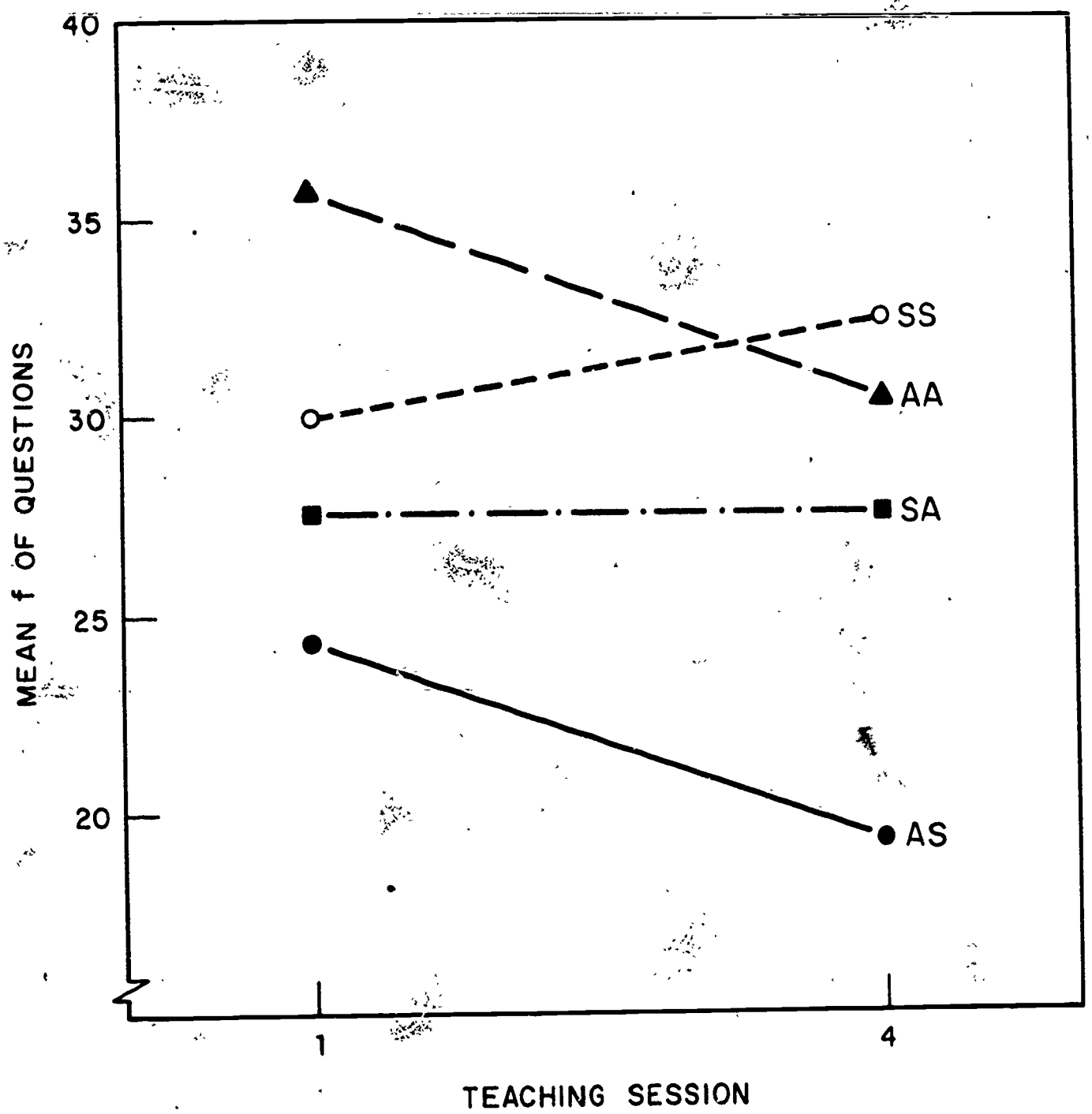


FIGURE 6

Mean Frequency of Teacher Lower-Order Questions on Teach 1 and Teach 4 for Each Treatment Group. (S indicates cued condition; A indicates non-cued condition.)



c. Total questions. As expected, frequency of total questions for cued model groups increased while frequency of total questions decreased for non-cued model groups (Figure 7). Analysis of variance on the (T4 - T1) difference scores showed that the inverse relation between the positive difference scores of the cued model groups and the negative difference scores of the non-cued model groups was significant ($p < .005$) (Table 9).

TABLE 9
Two-Way Analysis of Variance on
Total Teacher Questions

Source of Variation	SS	df	MS	F
Modeling Effects (A)	1918.22	1	1918.22	9.56*
Feedback Effects (B)	308.02	1	308.02	1.54
Interaction (A x B)	225.62	1	225.62	1.12
Residual (error)	7221.09	36	200.58	

*$p < .005$

Figure 8 illustrates the relationship between the (T4 - T1) difference scores for treatment groups on higher-order questions, lower-order questions, and total questions. (Means and standard deviations of [T4 - T1] difference scores are found in Appendix A.) Cued model groups showed more increase in HOQ from Teach 1 to Teach 4 than did non-cued model groups. On LOQ, non-cued groups showed negative change from their initial Teach 1 scores, cued model groups showed a positive change from their initial scores. Difference scores indicate that use of lower-order questions appears to be strongly affected by cueing or not cueing a model presentation. Difference scores on total questions used between Teach 1 and Teach 4 indicate that the combined frequency of question usage also appears to be affected by presence or absence of cueing on the model.

FIGURE 7

Mean Frequency of Teacher Total Questions on Teach 1 and Teach 4 for Each Treatment Group. (S indicates cued condition; A indicates non-cued condition.)

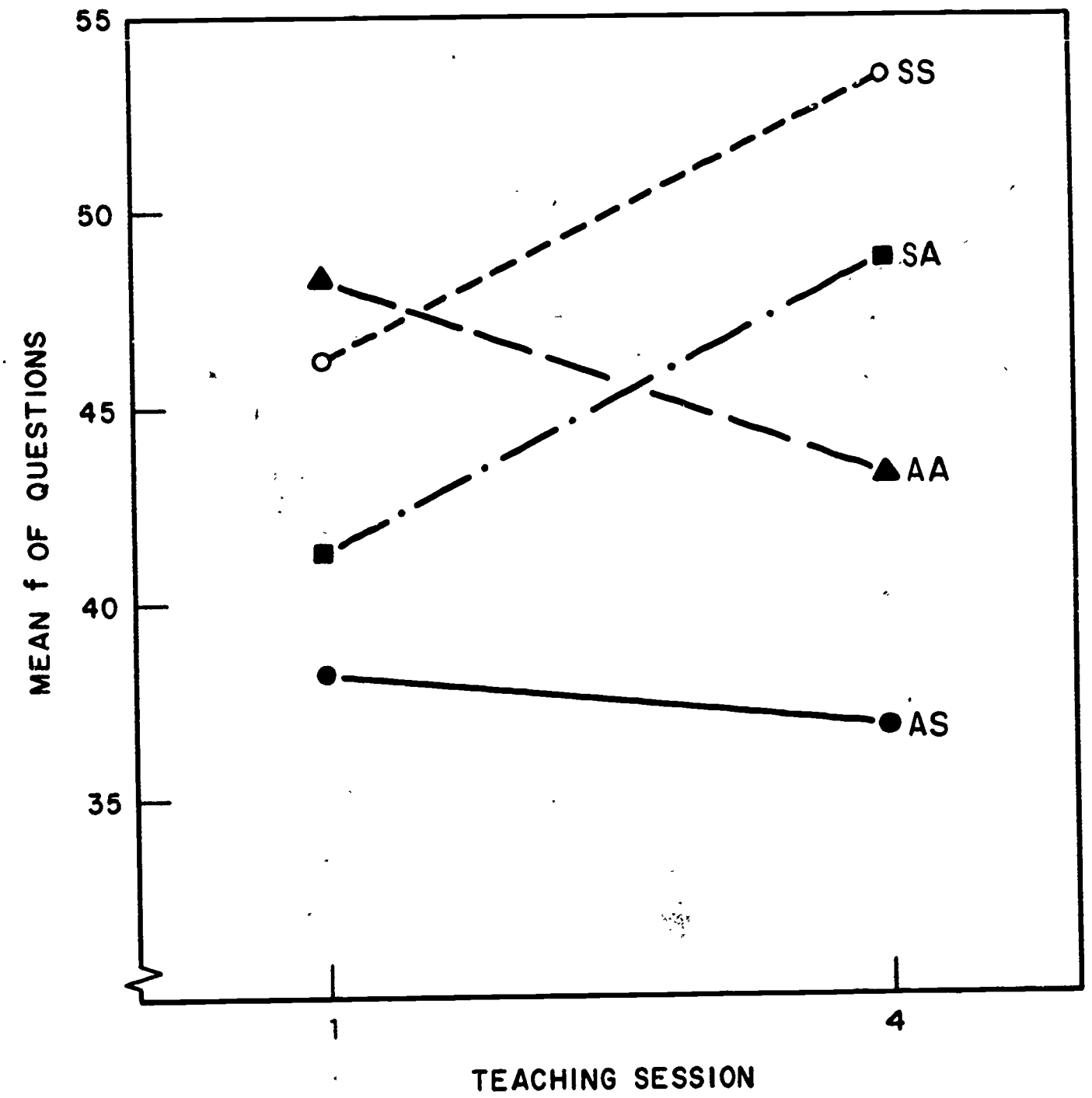


FIGURE 8

Histograms Showing Mean (T4 - T1) Difference Scores for Higher-Order, Lower-Order and Total Questions by Treatment Groups

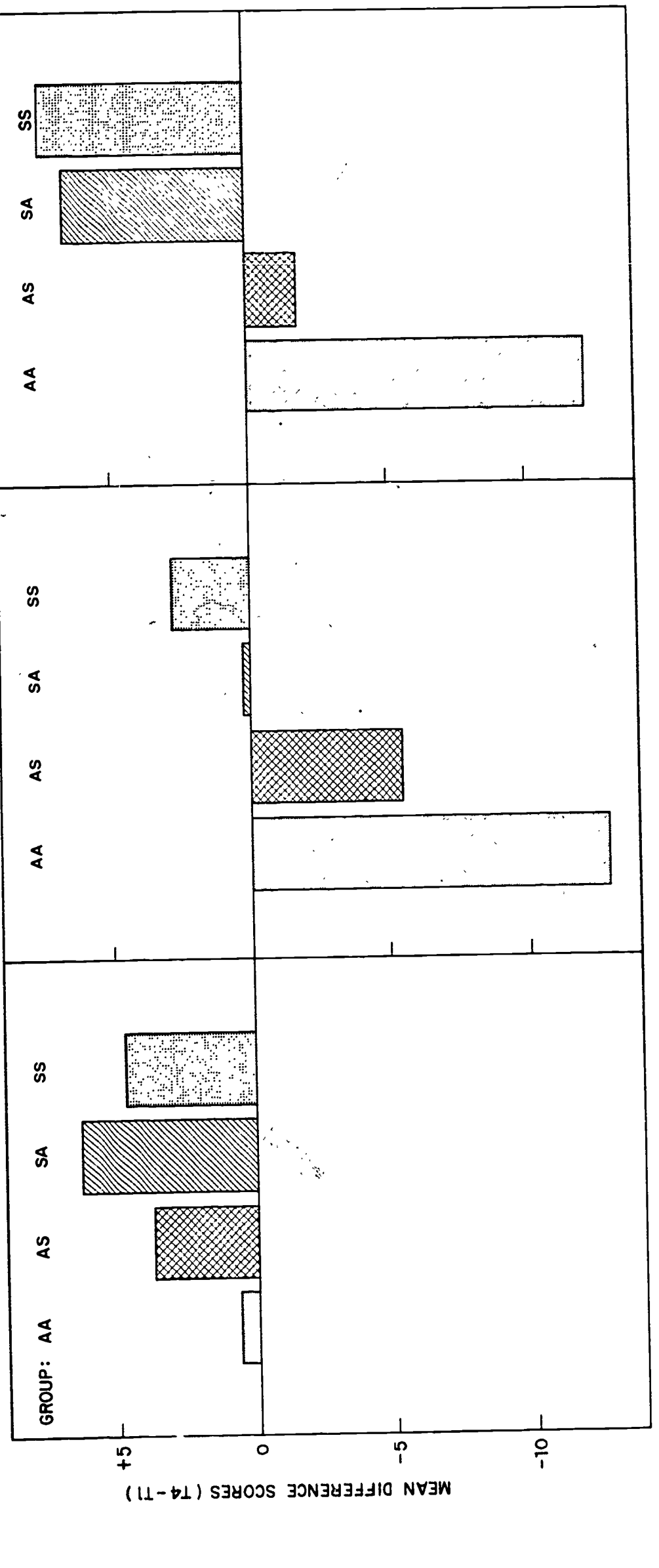


Figure 9 shows the treatment effects at each teaching session. The plot-points for each of the groups were derived from the mean differences in frequency of higher-order questions between the second, third, and fourth teaching sessions and Teach 1, the pretest. Pretest mean scores for each group were plotted at zero (T1 - T1) to illustrate cumulative treatment effects within the two-hour continuous training period (note the learning curve of Group SS). Intervening treatments are indicated along the horizontal axis between teaching sessions. Mean frequencies of higher-order questions, lower-order questions, and total questions by treatment group on all four teaching sessions are shown in Appendix B.

Analysis of variance on the proportions of higher-order questions to lower-order questions. A Friedman two-way analysis of variance on the rank-ordered proportions (Siegel, 1956, pp. 166-173) of higher-order questions to lower-order questions over the four teaching sessions yielded significant treatment differences ($p < .05$) with cued model groups ranking higher than non-cued model groups (Appendix C).

Analysis of variance on the separate question categories. Analysis of variance was performed on the differences between Teach 1 and Teach 4 frequencies for all eight categories of questions. Analysis questions showed a significant increase for cued model groups over non-cued model groups ($p < .05$) (Table 10). Application questions also increased in frequency from Teach 1 to Teach 4 but differences between cued model groups and non-cued model groups were only marginally significant ($p < .10$)(Table 11).

Effects on Pupils

A secondary interest of this study was to determine whether an increase in teacher higher-order questioning behavior would influence pupil questioning behavior to change in the direction of greater use of higher-order questions. The frequency of pupil questions was quite low (see Table 12), precluding an adequate statistical test of differences. Figure 10 illustrates the graphic relation of teacher mean HOQ to pupil group mean HOQ frequency. Pupil means and standard deviations by treatment group and teaching session are found in Appendix D.

Correlations of teacher higher-order questions with pupil higher-order questions. Correlations of teacher higher-order questions with pupil higher-order questions on Teach 1 yielded two negative and two positive coefficients (Table 13). None was significant. On Teach 4, however, all teacher-pupil correlations were negative. One of these, that for Group AA ($r = -.73$), was significantly so ($p > .05$). In addition, the negative correlation of teacher total questions with pupil total questions on Teach 4 ($r = -.29$) approached significance ($p > .10$).

FIGURE 9

Cumulative Treatment Effects at Each Training Session

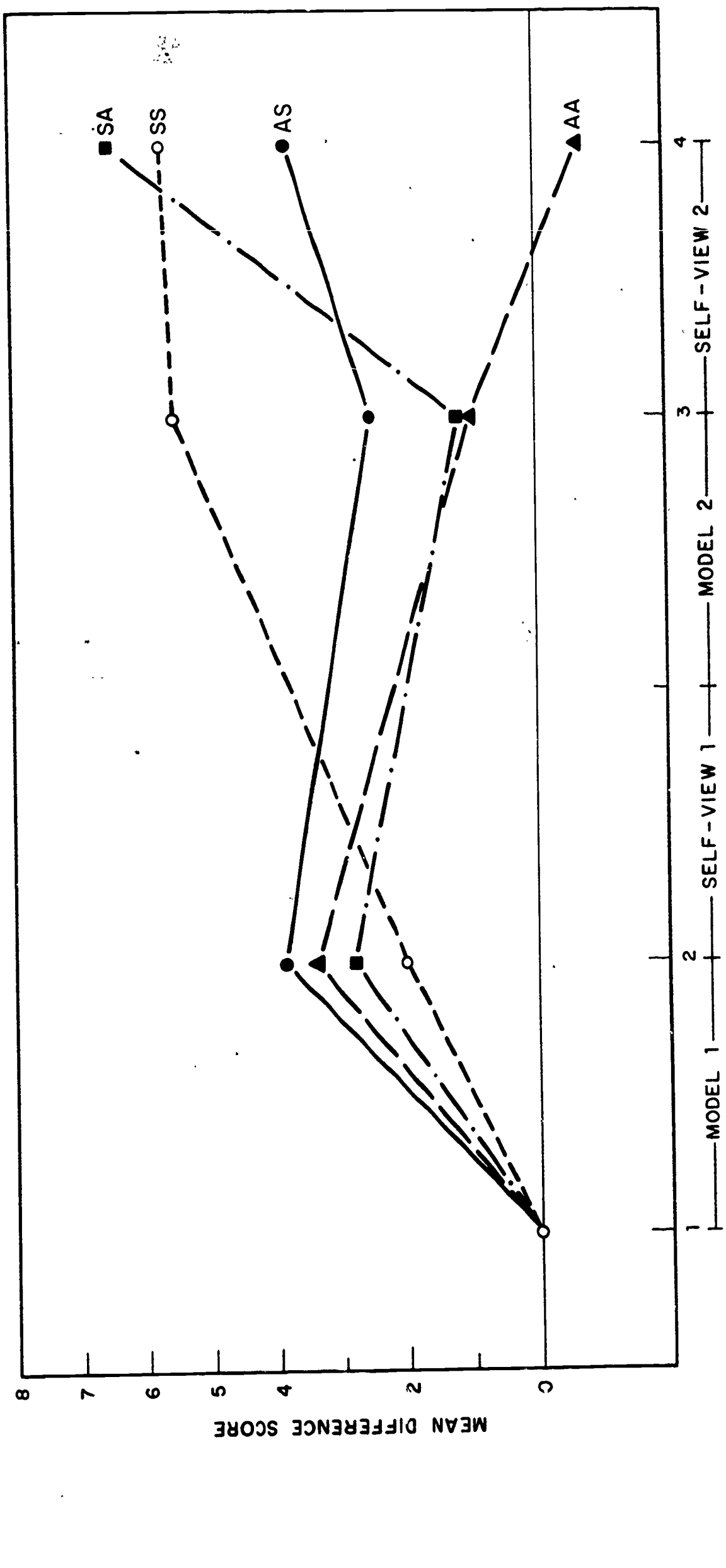


TABLE 10

Analysis of Variance on Difference Scores for Analysis Questions

(Teach 4 - Teach 1)

Source of Variation	SS	df	MS	F
Modeling Effects (A)	32.40	1	32.40	5.36*
Feedback Effects (B)	.40	1	.40	.07
Interaction (A x B)	19.60	1	19.60	3.25
Residual (error)	217.60	36	6.04	

*$p < .05$

TABLE 11

Analysis of Variance on Difference Scores for Application Questions

(Teach 4 - Teach 1)

Source of Variation	SS	df	MS	F
Modeling Effects (A)	140.6	1	140.6	2.90*
Feedback Effects (B)	7.2	1	7.2	
Interaction (A x B)	24.0	1	24.0	
Residual (error)	1300.5	36	36.1	

*$p < .10$

TABLE 12

Number of Pupil Questions in Each Category Over All Interns and All Teaches

Questions Category	Memory 1	Translation 2	Interpretation 3	Application 4	Analysis 5	Synthesis 6	Lower-Order Evaluation 7	Evaluation 8	Total
Number of Questions	302	1	9	87	58	1	2	1	461
% of Total Questions	66	.2	2	19	13	.2	.4	.2	100

FIGURE 10

Mean Number of Teacher and Pupil Higher-Order Questions at Each Teaching Session by the Treatment Group

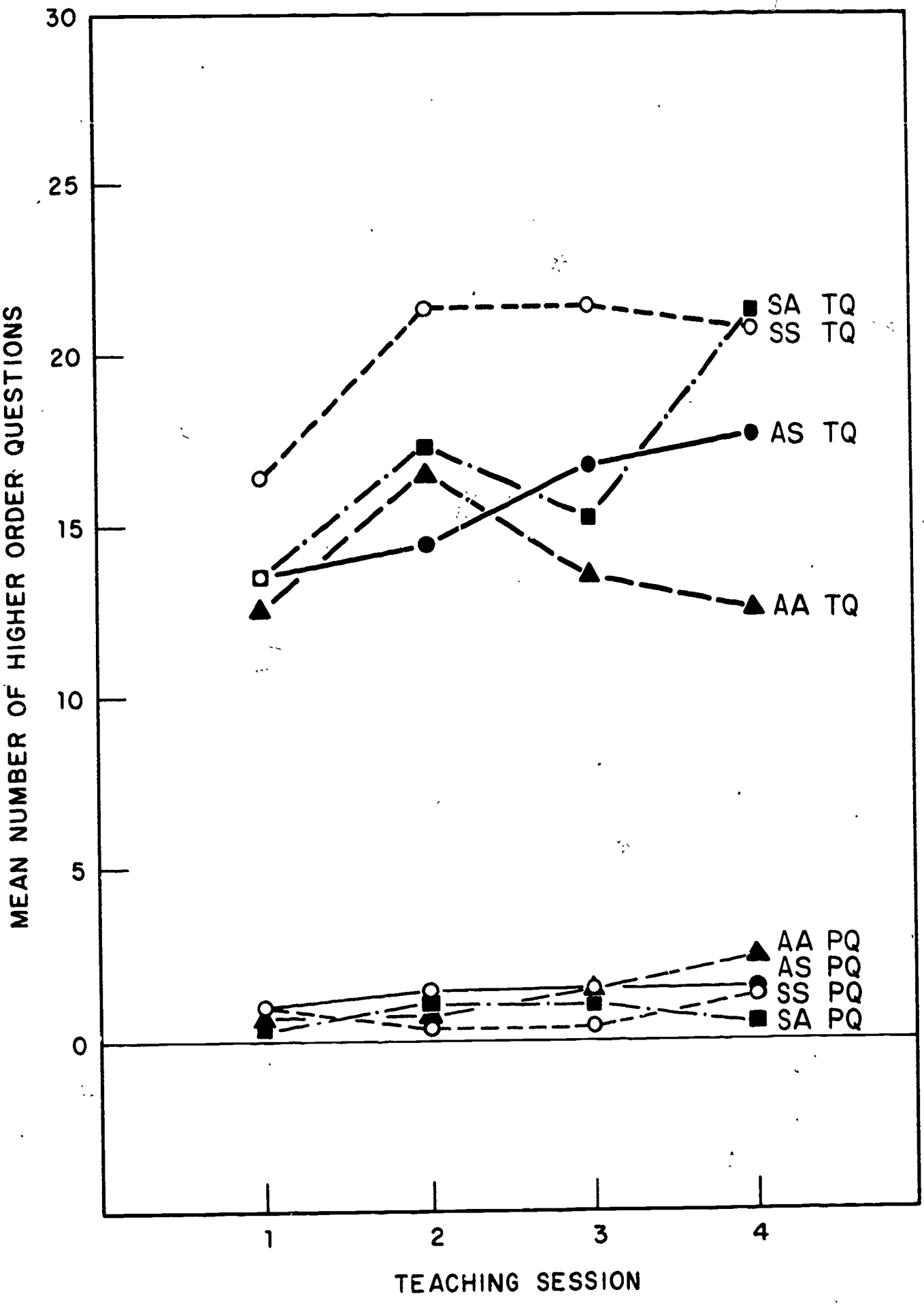


TABLE 13

Correlations of Teacher Higher-Order Questions with Pupil Higher-Order Questions on Teach 1 and Teach 4 by Treatment Groups

Group	AA (N=10)	AS (N=10)	SA (N=10)	SS (N=10)	Total (N=40)
Teach 1	.48	-.29	-.38	.19	-.06
Teach 4	-.73**	-.12	-.01	-.43	-.29*

*p < .10
**p < .05

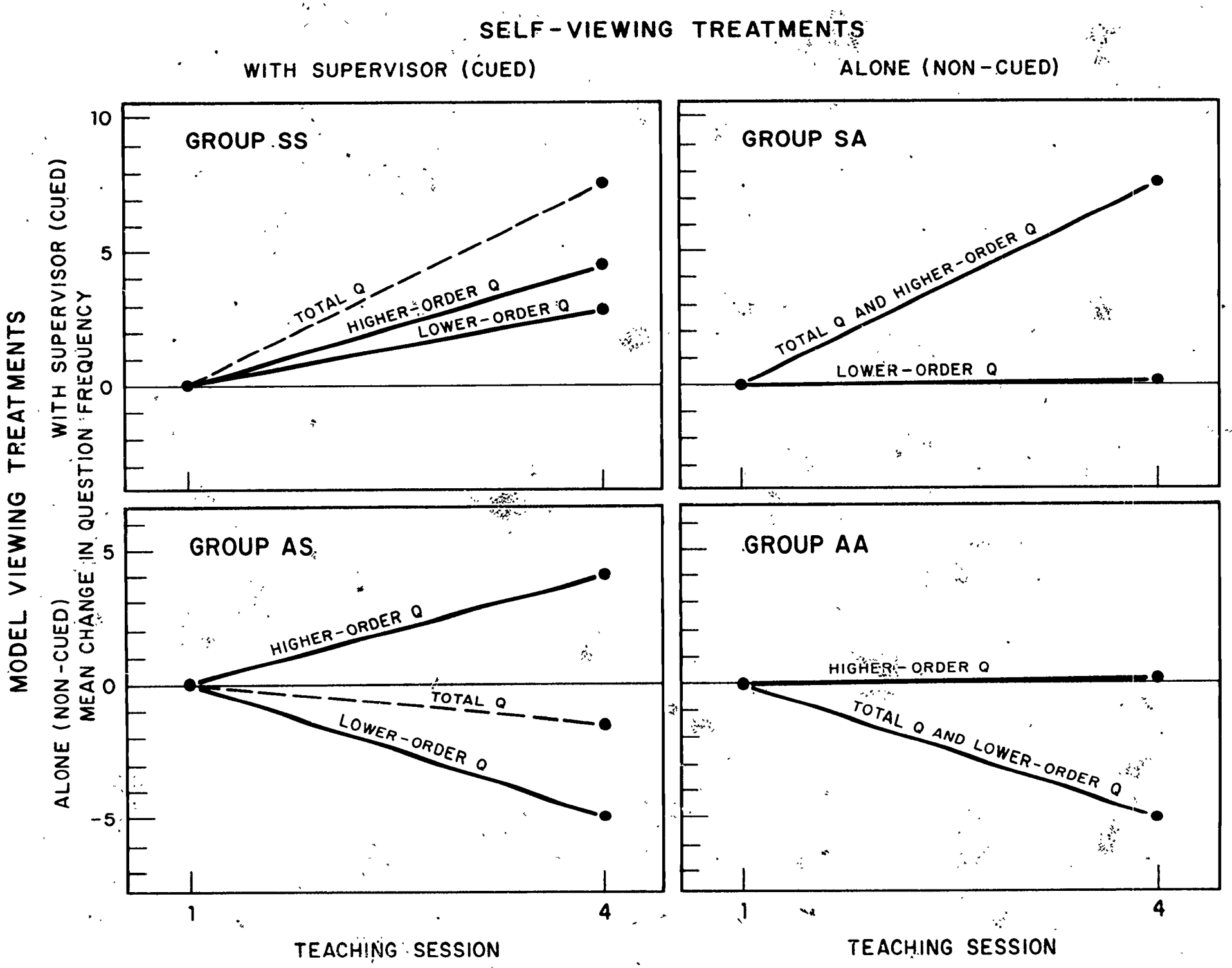
In the AA and SS groups, the correlation of teacher with pupil questions changed from positive in Teach 1 to negative in Teach 4. For group AA, that change from positive to negative correlation coefficients was significant ($p < .01$) using Fisher's z (Guilford, 1965, pp. 190-191). Tests on the differences between the Teach 1 to Teach 4 correlation coefficients of the other three treatment groups and on the total Teach 1 to Teach 4 correlation were not significant. The lack of significant findings here could be a function both of the small cell size ($n = 10$) and of the short time (ten minutes) allowed for each training session.

Summary of Results

This study investigated the use of cueing procedures in modeling and feedback treatments on the acquisition of a teaching behavior. The major hypothesis stated that cues from an experimenter on the desired behavior during modeling and feedback treatments would increase the frequency of the teaching behavior being learned. Findings indicated that treatments involving modeling with cues were clearly more effective in training teachers to use higher-order questions than non-cued modeling treatments. No significant effect was found for the feedback treatments.

FIGURE 11

Summary of Mean Rates of Change in Frequency of Higher-Order, Lower-Order and Total Questions Asked from Teach 1 to Teach 4 for Each Treatment Group



Frequency data from the study indicated that both teachers and pupils used mainly factual questions (Appendix E). Around 60 percent of the questions asked by either teachers or pupils were memory questions. The next largest proportion of teacher or pupil questions was accounted for by questions calling for the application of knowledge. Analysis questions were third in frequency. Very little of the teacher-pupil interaction involved translating, comparing, synthesizing, or evaluating.

Discussion

Developing the cognitive potential of children is a central aim of education. Since the elementary school years are considered crucial to the development of skills of reasoning and inquiry, it seems imperative that elementary school teachers should teach in ways that encourage development of and provide practice for their pupils' intellectual skills.

Questioning has been considered a central teaching tool for eliciting relationship, or higher-order, thinking in children. It is the foundation of any type of directed discussion activity (Fenton, 1967) and the first step in problem solution (Dewey, 1933; Suchman, 1962; Taba, 1966). Several studies have indicated that the use of questions to elicit certain types of thinking is important (Davis & Tinsley, 1967; Davis & Hunkins, 1966; Pfeiffer & Davis, 1965). It is generally postulated that teachers' actions are one of the most important influences on the thought processes of children--encouraging freedom to explore alternatives or constricting thought to factual specifics. It is important, therefore, that an efficient and effective method be developed to teach teachers how to use questioning techniques or at least to sensitize them to the varied range of question types available for strategizing learning activities which seek to develop the higher cognitive skills.

Results of the present study generally confirmed the hypothesis that cueing model presentations would help beginning teachers focus on specific behaviors to be learned. Teachers were taught to increase their use of higher-order questions by a procedure which involved showing a videotape model of a teacher using questioning skills and by providing verbal cues on the various types of higher-order questions occurring during the demonstration. Cueing on the teacher's own use of higher-order questions, as displayed in the videotapes of their own practice sessions, did not prove to be significantly effective.

Each study has limitations in reaching its goals. This study utilized a relatively small sample of pre-service elementary intern teachers. Generalizability of the results must therefore be weighed in terms of the limitations of low-frequency data and in terms of the typicalness of the sample of interns in relation to the general population of pre-service intern teachers or teachers in general. Certain assumptions as to the chain of events from teacher learning to pupil learning were also made. These assumptions were stated in terms of the input-output process paradigm guiding the study.

A basic assumption was that higher-order questioning is a skill not usually possessed by intern teachers. The basic paradigm involved a training situation with modeling and feedback conditions interspersed between microteaching sessions. In the learning of complex cognitive skills, modeling is an acquisition or learning variable, while feedback is a performance variable (Bandura, 1968). This distinction is supported in this study by the finding that cued vs. non-cued modeling produced differential effects, while cued vs. non-cued feedback produced no differential effects. Since modeling is an acquisition variable and if teachers could not use the questioning skill effectively at the start of the training, then it would be reasonable to expect cued modeling to be a more effective learning treatment than non-cued modeling. If feedback primarily governs performance rather than learning, it becomes less surprising that during acquisition, cued feedback is no more effective than non-cued feedback.

It is interesting to note that plots of the difference scores at each teaching session illustrate learning curves which are in accord with predictions of social learning theory as described by Bandura (1965c, 1967, 1968). (See Figure 9, p. 28.)

Group SS showed a rapid acquisition of the criterion behavior and then asymptotic performance (which could be a function of a ceiling effect imposed by the ten-minute performance session). According to Bandura's (1968) suggestion that modeling is an acquisition variable, it is predictable that subjects in a cued model condition would gain in incidence of a new skill faster than subjects who received no discriminative cues during model presentations. If feedback is a performance variable, it is not surprising that cueing the feedback session helped to maintain the level of performance for the SS group.

Group SA, although gaining sharply on the final teaching session, showed a slow and less predictable acquisition curve, even though this group also received cues from a supervisor during modeling sessions. Perhaps the absence of cueing on the first feedback session allowed more forgetting and the effects of cueing the model were slower in becoming visible. The process could be described by an all-or-none learning theory (Hilgard & Atkinson, 1967) in which the cumulative effects build up until a great spurt of learning occurs.

Group AS appears to have maintained a level of performance (due to cued feedback) after the initial acquisition with only slight increases on each teaching session (see Appendix F, which shows the mean frequency of higher-order, lower-order and total questions at each teaching session).

Group AA, as a type of control group, received no cues in either model or feedback conditions. Figure 9 shows a steady decrease in the difference scores for this group. Perhaps this decreasing curve indicates extinction of behavior which was learned during the set induction period which occurred prior to Teach 1. Symbolic encoding of the visual images may have been too difficult without external assistance.

It was not possible to run control groups who received neither modeling nor feedback. No conclusions can be drawn from this study, therefore, regarding the absolute effects of modeling or feedback alone. If self-view feedback is a performance variable for complex cognitive tasks, and if a subject has mastered the task skills, then he may get sufficient cues to regulate his performance from non-cued feedback.

There are several alternative explanations for the finding that cueing was effective with modeling but not with feedback procedures. The supervisors may have been able to cue more effectively on the model behaviors than on self-view feedback. They were thoroughly familiar with the model tape before the training sessions began: they knew the location and category of each question on the model tape. Cues on the interns' own performance were necessarily more spontaneous, since the supervisor had no opportunity to see the interns' tapes in advance of the feedback sessions.

Another possibility is that cues from the supervisor may have interfered with the interns' own perceptions of the type of questions which were asked. It seems likely there would be a greater emotional response to interference occurring on one's own product than on a model's.

Yet another possibility is the "cosmetic effect" so often observed when people see videotape or film playbacks of their own behaviors. In this study, the interns may have been preoccupied with their own general appearance in the self-view sessions and failed to attend to cues given by supervisors.

Analysis of the teacher-pupil question relationships revealed negative correlations between teacher higher-order questions and pupil higher-order questions in the final teaching session, some of which were significant and all of which showed the same trend. When teachers use higher-order questions before pupils are ready for them, it may be that the frequency of pupil lower-order questions rises, since it is then incumbent upon the pupil to fill in the necessary factual information in order to answer the relationship questions. Conversely, perhaps pupils use more higher-order questions when teachers use more lower-order questions. This finding was not what was expected on the basis of general observational learning theory and the paradigm of this study. Perhaps modeling procedures are limited in their effectiveness when small blocks of time are involved in the training sequence.

Modeling by teachers may not be as effective <u>on pupils</u> as modeling by peers in a ten-minute segment of classroom interaction. Also, pupils asking higher-order questions may be strongly influenced by variables other than modeling, e.g., teacher reinforcement of higher-order questions when they occur or the unfamiliarity of the teacher. Even though care was taken to select talkative students for the practice sessions, there still may have been a hesitancy on the part of the pupils to ask questions of a strange teacher.

The ten-minute time limitation presents other interesting problems. One is the possible depressant effect of a ten-minute session when a complex behavior is involved. There may be some plateau level of performance which is governed by the time period in which the task is to be performed, as suggested by Keller (1958) in his review of research on learning of Morse Code. As teacher higher-order questions increased in a ten-minute session, it would be natural that pupils would spend less time asking questions and more time responding to complex questions. Ten minutes may also hold the number of certain types of behavior to a minimum merely because of anticipation of the end of the session.

A ten-minute session may not allow for a change in question type by practice, resulting in a new, coherent teaching style. Such a short time period could not offset other variables which would influence a teacher's strategy. As pointed out by Taba at the conclusion of her study (1966), teachers' personal capacities for logical thinking; their skill in formulating relationship questions; their facility in managing the thought produced by those questions; and their capacity for rapid and accurate diagnosis of pupil responses in terms of quality and content are all factors which directly affect the training results. Cued modeling groups significantly changed in the total number of questions asked in a session. It may be possible to model change in frequency of behavior in a short period of time but a longer period of practice may be needed to change the type of behavior through modeling and feedback procedures.

Consistent with other studies of questioning behavior cited earlier, this study found that teachers utilize primarily factual questions in planned classroom discourse. This is a firm habit, appearing even in a training situation where teachers are told to concentrate on using higher-order questions. Also, there is great reliance on the "how to do it" type of question which calls for application of knowledge. The model tape exhibited a larger percentage of these two types of behaviors. It could, therefore, be that modeling effects on these behaviors were very powerful in this study. An alternative explanation is that these two types of questions are the easiest to ask, hence their higher frequency of occurence.

Implications for Teacher Training

Training and retraining of teachers is a continuing problem. Any finding is welcome which might suggest another way to accomplish this task more effectively and efficiently. Cued modeling procedures are not only useful in the one-to-small-group microteaching context, but also could find application in a training situation where several teachers meet to learn about a complex teaching skill. Large lecture and small discussion groups alike could find a training situation useful in which a supervisor or consultant cues on a desired behavior. Feedback could be provided in a variety of forms such as videotaping in the individual classrooms, some form of written evaluation, or group critique of a performance record. This feedback could also be a form of cued videotape record for use by the classroom teacher at her own leisure.

In short, cued modeling procedures could be adapted to a variety of forms allowing considerable flexibility in the planning of teacher-training sequences. Use of modeling, to be most effective in the training of a complex cognitive task, probably should be accompanied by cues which point out salient characteristics and help avoid initial skill confusion.

Much research is needed to determine what effects certain teacher behaviors have on pupils. If there is an inverse relationship between teacher use of certain types of questions and pupil use of the same types of questions, then teachers should be aware of this relationship.

Future Research Possibilities

This study suggests questions for further investigation. One of these concerns the use of a supervisor to do the cueing. Modeling of a complex skill was found effective when cued by an experimenter who identified the salient characteristics of the desired behavior when it occurred in a complex interactional setting. Without this help, only the most common of behaviors may have been easily observed. It is important in a cued modeling or cued feedback session to have supervisors who can discriminate between the various behaviors. To test the effects of various types of cueing compared with the effects of a model presentation without cueing, it would be necessary to introduce various combinations of supervisor cues with various types of model presentation. An example would be to show a model illustrating all of the behaviors to be learned and to vary the cueing from total cues on all behaviors, to cueing on only selected behaviors, to absence of any cueing as a control.

A related question is whether cueing should include affective elements. Still another question is whether cueing could be done by a means other than a live supervisor. For instance, comments on a video- or audiotape might be used to cue relevant behavior. One could test effects of reinforcement of the model (vicarious reinforcement for the subject) against direct reinforcement of the subject's attending responses as he views the model.

Feedback may be more effective if placed at a point in the training sequence different from that used in this study. Associated with the question of sequence of training is the problem of length of training time. Would feedback become more effective in longer training sequences?

Other questions relate to model tape characteristics. Is a model which displays a high incidence of behaviors more effective than a model which illustrates each category slowly and methodically? Would models teaching regular-sized classes be more effective than models teaching micro-classes? Is model effectiveness related to the sex of the model and the sex of the subjects?

In this study, it was not feasible to gather field data on the subjects' performance on the job long after training. This is information which is needed to test transfer and to determine the effectiveness of the skill training. A longitudinal study in the field could also determine if there are cumulative effects on pupils from the continued use by teachers of particular types of questions. What effect does a learner's asking of higher-order questions have on his own learning abilities? Pupils may learn just as well without overtly asking higher-order questions. McDonald's (1968) model of mathemagenic behaviors suggests that perhaps covert asking of questions by pupils can be one of a set of intervening variables operating between teacher verbal discourse and pupil overt behavior such as that displayed on an achievement test. Heuristic teaching, which may involve modeling of higher-order questioning skills by a teacher, should induce this type of covert behavior on the part of pupils which initiates patterns of conceptual thinking.

Conclusion

The major question to be answered in this study concerned the effectiveness of cueing during modeling and feedback presentations. Results of the study indicated that cued modeling treatments were clearly more effective in training teachers to use higher-order questions than were non-cued modeling treatments. Cueing the feedback presentations, however, did not appreciably change the use of the criterion behavior. These findings lend support to the distinction that modeling is an acquisition variable and feedback is a performance variable.

References

Angell, D. & Lumsdaine, A. A. Prompted plus unprompted trials versus prompted trials alone in paired-associate learning. In A. A. Lumsdaine (Ed.), Student response in programmed instruction: A symposium. Washington, D. C.: National Academy of Sciences - National Research Council, Publication No. 943, 1961. Pp. 389-399.

Aschner, M. J. The analysis of verbal interaction in the classroom. In A. A. Bellack (Ed.), Theory and research in teaching. New York: Bureau of Publications, Teachers College, Columbia University, 1963. Pp. 53-78.

Aschner, M. J., Gallagher, J. J., Perry, J., Afsar, S., Jenner, W. C., & Farr, H. A system for classifying thought processes in the context of classroom verbal interaction. Urbana: Institute for Research on Exceptional Children, University of Illinois, 1965.

Bandura, A. J. The role of imitation in personality development. Journal of Nursery Education, 1963, 18, 207-215.

Bandura, A. J. Behavioral modifications through modeling procedures. In L. Krasner & L. P. Ullmann (Eds.), Research in behavior modification. San Francisco: Holt, Rinehart, and Winston, 1965. (a)

Bandura, A. J. Influence of model's reinforcement contingencies on the acquisition of imitative responses. Journal of Personality and Social Psychology, 1965, 1, 589-595. (b)

Bandura, A. J. Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), Advances in experimental social psychology. Vol. 2. New York: Academic Press, 1965. (c)

Bandura, A. J. Principles of behavioral modification. New York: Holt, Rinehart, and Winston, 1967.

Bandura, A. J. Social learning theory of identificatory processes. In D. A. Goslin & D. C. Glass (Eds.), Handbook of socialization theory and research. Chicago: Rand McNally, 1968.

Bandura, A. J., Grusec, J. E., & Menlove, F. L. Some social determinants of self-monitoring reinforcement systems. Journal of Personality and Social Psychology, 1967, 5, 449-455.

Bandura, A. J., & McDonald, F. J. The influence of social reinforcement and the behavior of models in shaping children's moral judgments. Journal of Abnormal and Social Psychology, 1963, 67, 274-281.

Bandura, A. J., Ross, D., & Ross, S. A. Imitation of film-mediated aggressive models. Journal of Abnormal and Social Psychology, 1963, 66, 3-11. (a)

Bandura, A. J., Ross, D., & Ross, S. A. A comparative test of the status envy, social power, and secondary reinforcement theories of identificatory learning. Journal of Abnormal and Social Psychology, 1963, 67, 527-534. (b)

Bandura, A. J. & Walters, R. Social learning and personality development. New York: Holt, Rinehart, & Winston, 1963.

Bandura, A. J., Ross, D. & Ross, S. A. Vicarious reinforcement and imitative learning. Journal of Abnormal and Social Psychology, 1963, 67, 601-607. (c)

Bellack, A., Kliebard, H., Hyman, R., & Smith, F. The language of the classroom. New York: Teachers College Press, 1966.

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed), Problems in measuring change. Madison: University of Wisconsin Press, 1963. Pp. 3-20.

Berliner, D. C., McDonald, F. J., Allen, D. W., & Sobol, F. T. A comparison of different modeling procedures in the acquisition of a teaching skill. Paper presented at the meeting of the American Educational Research Association, New York, February, 1967.

Bloom, B. S. (Ed.) Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain. New York: David McKay Co., 1956.

Burton, W. H. The guidance of learning activities. (3rd ed.) New York: Appleton-Century-Crofts, 1962.

Butler, F. A. The improvement of teaching in secondary schools. Chicago: University of Chicago Press, 1939.

Claus, K. E. The effects of modeling and feedback variables on higher-order questioning skills. (Doctoral dissertation, Stanford University) Ann Arbor, Michigan: University Microfilms, 1968, No. 69-207. (a)

Claus, K. E. Effects of cueing during modeling and feedback sessions on learning a teaching skill. Paper presented at the 76th meeting of the American Psychological Association, San Francisco, August 1968. (b)

Cook, J. O. & Kendler, T. S. A theoretical model to explain some paired-associate learning data. In A. A. Lumsdaine & R. Glaser (Eds.), Teaching machines and programmed learning. Washington, D. C.: National Education Association, 1950. Pp. 602-604.

Davis, O. L. & Hunkins, F. P. Textbook questions: What thinking processes do they foster? Peabody Journal of Education, 1966, 43, 285-293.

Davis, O. L. & Tinsley, D. C. Cognitive objectives revealed by classroom questions asked by social studies student teachers. Peabody Journal of Education, 1967, 45, 21-26.

Dewey, J. How we think. (2nd ed.) Boston: D. C. Heath, 1933.

Dodl, N. R. Pupil questioning behavior in the context of classroom interaction.(Doctoral dissertation, Stanford University) Ann Arbor, Michigan: University Microfilms, 1965, No. 66-2512.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), Psychology: A study of a science. Vol. 2. General systematic formulations, learning, and special processes. New York: McGraw-Hill, 1959. Pp. 380-491.

Fenton, E. The new social studies. New York: Holt, Rinehart, & Winston, 1967.

Flanders, N. & Amidon, E. Direct and indirect teacher influence on dependent prone students learning geometry. Journal of Educational Psychology, 1961, 52, 286-291.

Gallagher, J. J. Expressive thought by gifted children in the classroom. Elementary English, 1965, 42, 559-568. (a)

Gallagher, J. J. Productive thinking of gifted children. Urbana: University of Illinois, U. S. Office of Education Cooperative Research Project No. 965, 1965. (b)

Gallagher, J. J. & Aschner, M. J. A preliminary report on analyses of classroom interaction. Merrill Palmer Quarterly, 1963, 9, 183-194.

Gewirtz, J. L. & Baer, D. M. Deprivation and satiation of social reinforcers as drive states. Journal of Abnormal and Social Psychology, 1958, 57, 165-172.

Glaser, R. Psychology and instructional technology. In R. Glaser (Ed.), Training research and education. New York: John Wiley & Sons, 1962. Pp. 1-30.

Guilford, J. P. The structure of intellect. Psychological Bulletin, 1956, 53, 267-293.

Guilford, J. P. Fundamental statistics in psychology and education. (4th ed.) New York: McGraw-Hill, 1965.

Guszak, F. J. Reading comprehension development as viewed from the standpoint of teacher questioning strategies. Paper presented at the meeting of the American Educational Research Association, Chicago, February, 1966.

Hilgard, E. R. & Atkinson, R. C. Introduction to Psychology (4th ed.) New York: Harcourt, Brace & World, Inc., 1967.

Hilgard, E. R. & Bower, G. H. Theories of learning. (3rd ed.) New York: Appleton-Century-Crofts, 1966.

Hill, W. F. Learning theory and the acquisition of values. Psychological Review, 1960, 67, 317-331.

Hudgins, B. B. & Ahlbrand, W. P., Jr. A study of classroom interaction and thinking. Interim report, 1967. Saint Louis: Central Midwestern Regional Educational Laboratory, Inc. U. S. Office of Education Contract No. OEC 3-7-062875-3056.

Jackson, P. W. The way teaching is. In The way teaching is. Washington, D. C.: Association for Supervision and Curriculum Development and the Center for the Study of Instruction of the National Education Association, 1966. Pp. 7-28.

Keller, F. S. The phantom plateau. Journal of Experimental Analysis of Behavior, 1958, 1, 1-13.

Krumboltz, J. D., Varenhorst, B., & Thoresen, C. E. Non-verbal factors in the effectiveness of models in counseling. Journal of Counseling Psychology, 1967, 14, 412-418.

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963. Pp. 21-38.

McBrearty, J. F., Marston, A. F., & Kanfer, F. H. Conditioning a verbal operant in a group setting: Direct vs. vicarious reinforcement. American Psychologist, 1962, 16, 425-434.

McDonald, F. J. Educational psychology. (2nd. ed.) Belmont: Wadsworth, 1965.

McDonald, F. J. & Allen, D. W. Training effects of feedback and modeling procedures on teaching performance. Stanford: Stanford University, U. S. Office of Education Research Grant No. OE-6-10-078, 1967.

McDonald, F. J. A model of mathemagenic behaviors as intervening variables in classroom communication. Paper presented at the meeting of the American Educational Research Association, Chicago, February, 1968.

Meux, M. & Smith, B. O. Dimensions of teaching behavior. In B. Biddle & W. Ellena (Eds.), Contemporary research on teacher effectiveness. New York: Holt, Rinehart, & Winston, 1964. Pp. 127-164.

Miller, N. E. & Dollard, J. Social learning and imitation. New Haven: Yale University Press, 1941.

Mowrer, O. H. Learning theory and the symbolic processes. New York: John Wiley & Sons, 1960.

Orme, M. E. The effects of modeling and feedback variables on the acquisition of a complex teaching strategy. (Doctoral dissertation, Stanford University) Ann Arbor, Michigan: University Microfilms, 1966, No. 67-4417.

Pfeiffer, I. & Davis, O. L. Teacher-made examinations: What kind of thinking do they demand? National Association of Secondary School Principals Bulletin, 1965, 45, 1-10.

Rivlin, H. N. Teaching adolescents in secondary schools. (2nd ed.) New York: Appleton-Century-Crofts, 1961.

Sanders, N. M. Classroom questions: What kinds? New York: Harper & Row, 1966.

Sheffield, F. D. Theoretical considerations in the learning of complex sequential tasks from demonstration and practice. In A. A. Lumsdaine (Ed.), Student response in programmed instruction: A symposium. Washington, D. C.: National Academy of Sciences--National Research Council, Publication No. 943, 1961. Pp. 13-32.

Sheffield, F. D. & Maccoby, N. Summary and interpretation of research on organizational principles in constructing filmed demonstrations. In A. A. Lumsdaine (Ed.), Student response in programmed instruction: A symposium. Washington, D. C.: National Academy of Sciences--National Research Council, Publication No. 943, 1961. Pp. 117-131.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Siegel, L. & Siegel, L. C. The instructional gestalt. In L. Siegel (Ed.), Instruction: Some contemporary viewpoints. San Francisco: Chandler Publishing Co., 1967. Pp. 261-290.

Smith, B. O. A concept of teaching. Teachers College Record, 1960, 61, 229 -241.

Stevens, R. The question as a measure of efficiency in instruction: A critical study of classroom practice. New York: Teachers College, Columbia University, Contributions to Education Series, No. 48, 1912.

Suchman, J. R. The elementary school training program in scientific inquiry. Urbana: University of Illinois, U. S. Office of Education Cooperative Research Project No. 2660, 1962.

Taba, H. Teaching strategies and cognitive functioning in elementary school children. San Francisco: San Francisco State College, U. S. Office of Education Cooperative Research Project No. 2404, 1966.

Tolman, E. C. Principles of purposive behavior. In S. Koch (Ed.), Psychology: A study of a science. Vol. 2. General systematic formulations, learning, and special processes. New York: McGraw-Hill, 1959. Pp. 92-157.

Webster, H. & Bereiter, C. The reliability of changes measured by mental test scores. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963. Pp. 39-59.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wulff, J. J. & Kraeling, D. Familiarization procedures used as adjuncts to assembly-task training with a demonstration film. In A. A. Lumsdaine (Ed.), Student response in programmed instruction. Washington, D. C.: National Academy of Sciences--National Research Council, Publication No. 943, 1961. Pp. 141-154.

APPENDICES

APPENDIX A

Means and Standard Deviations of Difference Scores on Teacher Questions (T4 - T1) By Treatment Group

Treatment Group		AA (N=10)	AS (N=10)	SA (N=10)	SS (N=10)
Higher-order questions	M	0.6	3.7	6.3	4.6
	SD	2.2	11.8	3.1	4.5
Lower-order questions	M	-12.7	-5.5	.2	2.7
	SD	10.6	10.3	8.7	13.7
Total Questions	M	-12.1	-1.8	6.5	7.3
	SD	13.7	15.2	14.2	13.5

APPENDIX B

Mean Number of Teacher Higher-Order, Lower-Order and Total Questions Asked in Each Teaching Session by Treatment Group

	Higher-Order	Lower-Order	Total
Teach 1			
AA	12.5	35.7	48.2
AS	13.7	24.4	38.1
SA	13.7	27.6	41.3
SS	16.2	30.0	46.2
Teach 2			
AA	16.5	36.3	52.8
AS	14.5	23.2	37.7
SA	17.1	34.5	51.6
SS	21.2	25.4	46.6
Teach 3			
AA	13.6	28.6	42.2
AS	16.9	26.5	43.4
SA	15.1	23.5	38.6
SS	21.1	27.6	48.7
Teach 4			
AA	12.6	30.5	43.1
AS	17.7	19.2	36.9
SA	21.1	27.8	48.9
SS	20.8	32.7	53.5

APPENDIX C

Ratios of Frequency of Higher-Order Questions to Frequency of Lower-Order Questions in Each Teaching Session by Treatment Group

Treatment Group	Session				
	Teach 1	Teach 2	Teach 3	Teach 4	Total
SS	.620	.740	.768	.636	.086
SA	.428	.478	.613	.705	.552
AS	.547	.542	.612	.690	.601
AA	.350	.455	.476	.465	.432
Total	.479	.541	.616	.626	
$\frac{\Sigma \text{ Ratios}}{4}$	.486	.554	.617	.624	

APPENDIX D

Means and Standard Deviations of Pupil Questions Occurring in Each Ten-Minute Teach Session by Treatment Group

Treatment Group		AA	AS	SA	SS
		N = 10 microclasses for each treatment group[a]			
Teach 1					
Higher-order	M	.60	1.00	.40	1.00
	SD	.84	1.49	1.26	.81
Lower-order	M	1.80	2.30	2.20	2.20
	SD	2.39	2.58	3.46	1.87
Total Questions	M	2.40	3.30	2.60	3.20
	SD	7.82	14.68	12.49	5.96
Teach 2					
Higher-order	M	.70	1.22	1.00	.40
	SD	1.06	1.71	1.15	.52
Lower-order	M	1.60	3.11	1.40	1.50
	SD	1.65	2.67	.84	1.84
Total Questions	M	2.30	4.33	2.40	1.90
	SD	6.01	10.75	2.27	3.43
Teach 3					
Higher-order	M	1.40	1.40	.70	.70
	SD	1.28	1.50	1.06	.68
Lower-order	M	3.00	.90	1.60	1.30
	SD	3.05	.57	1.42	.45
Total Questions	M	4.40	2.30	2.80	2.00
	SD	16.01	1.79	4.23	2.89
Teach 4					
Higher-order	M	2.30	1.50	.40	1.30
	SD	2.06	1.65	.52	.95
Lower-order	M	1.70	1.80	2.40	1.90
	SD	2.26	2.48	1.78	1.85
Total	M	4.00	3.30	2.80	3.20
	SD	13.33	8.68	3.73	5.73

[a]Pupil means are averaged over ten microclasses for each treatment group. Each microclass contained on the average of seven pupils.

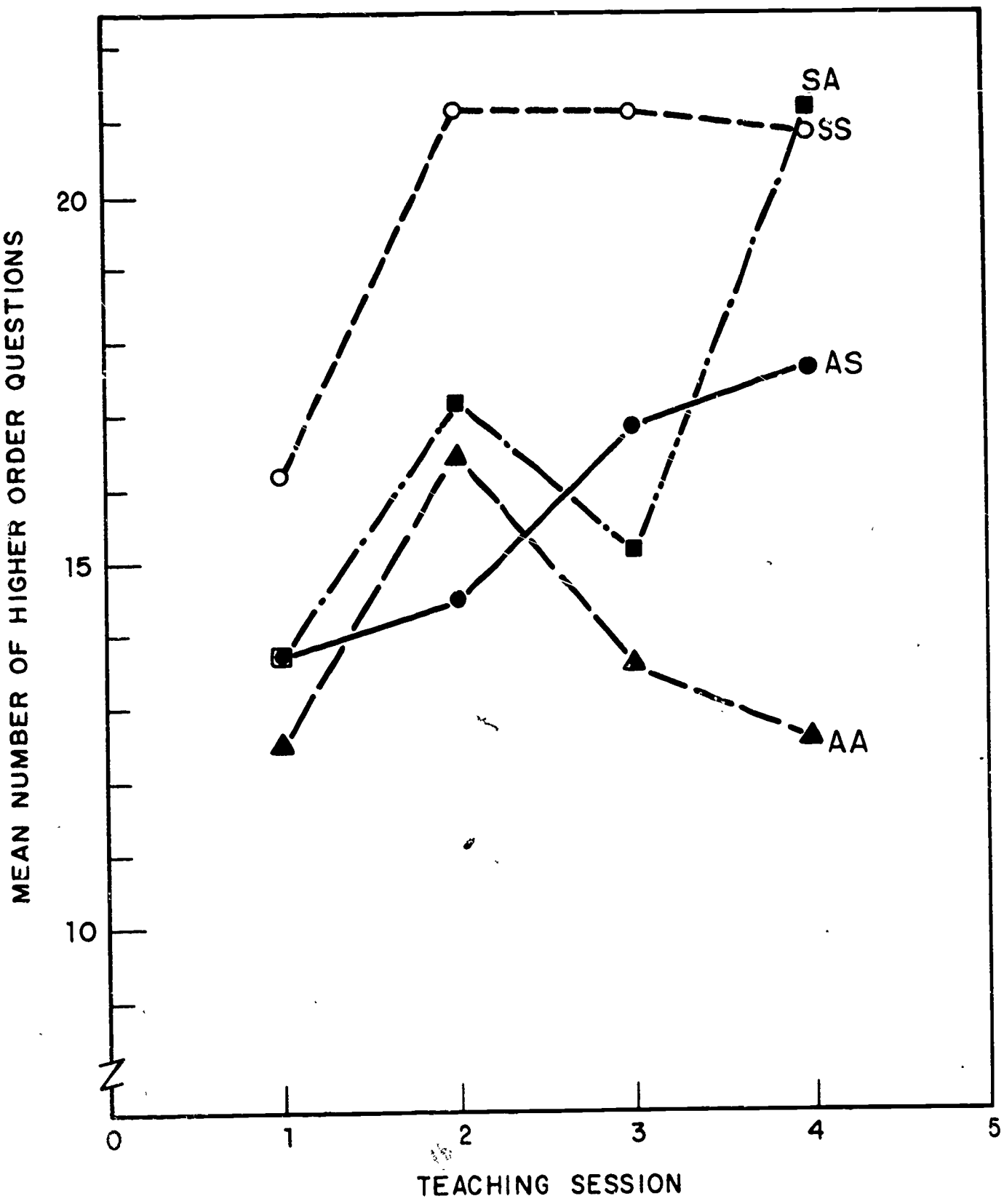
APPENDIX E

Number of Teacher and Pupil Questions in Each Category Over All Interns and All Teaches

		1 Memory	2 Translation	3 Interpretation	4 Application	5 Analysis	6 Synthesis	7 Lower-Order Evaluation	8 Evaluation	Total
Question Category										
Number of Questions	Teachers	4332	118	382	1345	529	183	188	61	7138
	Pupils	302	1	9	87	58	1	2	1	461
% of Total Questions	Teachers	61	2	5	19	7	3	3	1	100
	Pupils	66	.2	2	19	13	.2	.4	.2	100

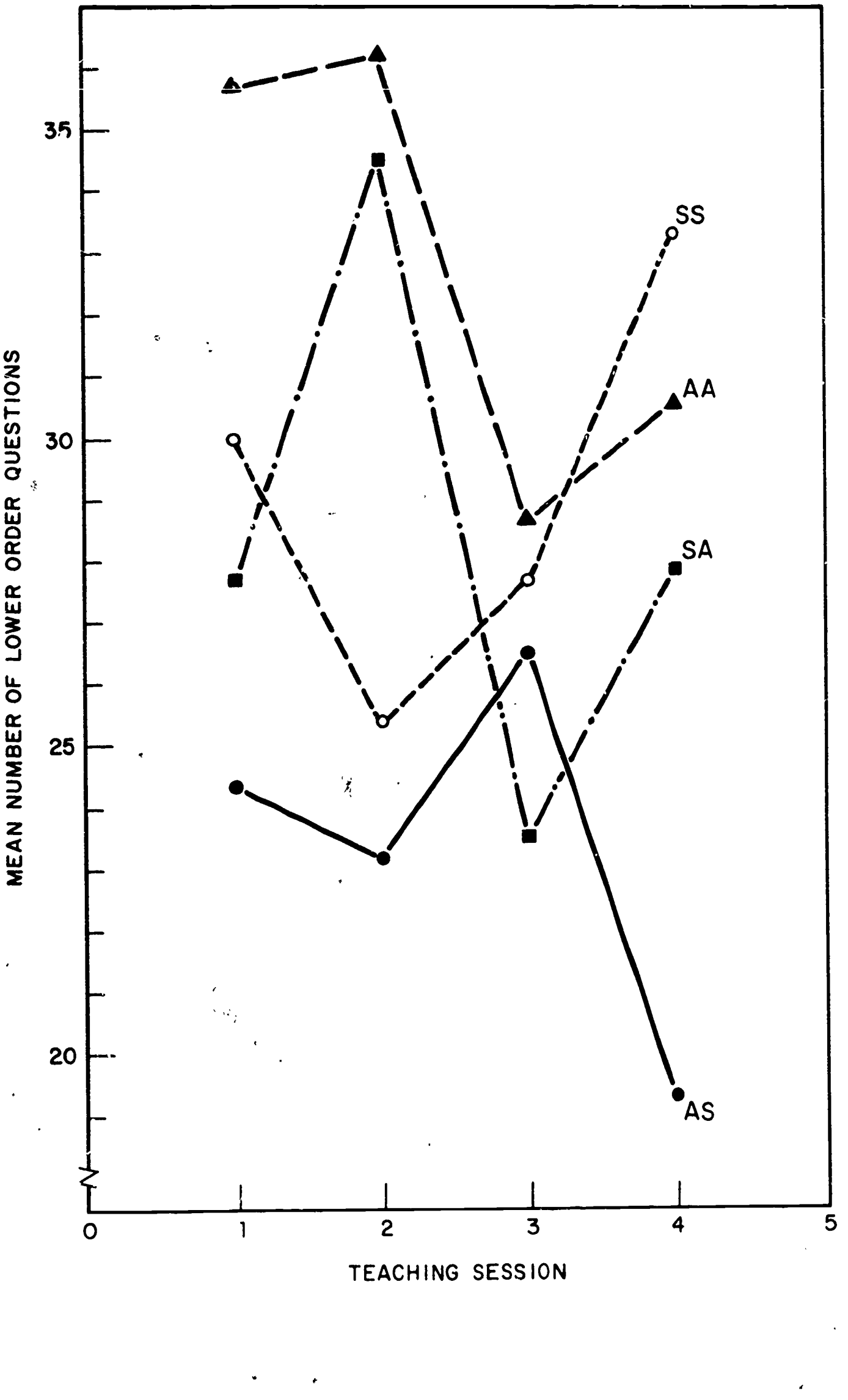
APPENDIX F 1

Mean Number of Teacher Higher-Order Questions
at Each Teaching Session by Treatment Group



APPENDIX F 2

Mean Number of Teacher Lower-Order Questions
at Each Teaching Session by Treatment Group



APPENDIX F 3

Mean Number of Teacher Questions Asked at
Each Teaching Session by Treatment Group