R E P O R T R E S U M E S

ED 011 261 SP 001 066

A COMPARISON OF DIFFERENT MODELING PROCEDURES IN THE ACQUISITION OF A TEACHING SKILL.

BY- ALLEN, D.W. AND OTHERS

PUB DATE 18 FEB 67

EDRS PRICE MF-$0.09 HC-$0.96 24P.

DESCRIPTORS- BIBLIOGRAPHIES, *MICROTEACHING, *MODELS, OBSERVATION, *QUESTIONING TECHNIQUES, RATING SCALES, STATISTICAL ANALYSIS, TABLES (DATA), *TEACHER EDUCATION, *TEACHING TECHNIQUES, VIDEO TAPE RECORDINGS, NEW YORK CITY

TO DETERMINE THE RELATIVE EFFECTIVENESS OF THREE MODELING PROCEDURES FOR MODIFYING TEACHER BEHAVIOR (APPLIED, ILLUSTRATIVELY, TO HIGHER-ORDER QUESTIONING BEHAVIOR), VIDEOTAPES WERE MADE OF A SERIES OF FOUR MICROTEACHING SESSIONS REPRESENTING IN A 2 X 2 X 2 FACTORIAL DESIGN (N 103) THAT FURNISHED EIGHT EXPERIMENTAL GROUPS, (1) SYMBOLIC VERSUS PERCEPTUAL MODELING--SOME GROUPS READ WRITTEN SCRIPTS, WHEREAS OTHERS SAW THE ENACTMENT OF THE SCRIPTS, (2) PURE VERSUS MIXED LESSONS--POSITIVE INSTANCES ONLY VERSUS POSITIVE AND NEGATIVE INSTANCES OF THE BEHAVIOR TO BE LEARNED, AND (3) MATCHING IN THE SPECIFIC CASE VERSUS MATCHING IN PRINCIPLE--SOME GROUPS PERFORMED THE SAME LESSON AS THE MODE, WHEREAS OTHERS USED ANY LESSON THAT MATCHED THE MODEL IN PRINCIPLE. TRANSFER WAS TESTED BY REQUIRING TEACHER TRAINEES TO INCORPORATE QUESTIONING SKILL IN A DIFFERENT LESSON CONTEXT. AS MEASURED BY PERCENT OF HIGHER-ORDER QUESTIONS OUT OF TOTAL QUESTIONS ASKED IN A 5-MINUTE TEACHING SESSION, ALL GROUPS SHOWED SIGNIFICANT GAINS OVER SESSIONS. SPECIFIC FINDINGS WERE--(1) THE PERCEPTUAL AND SYMBOLIC MODES DID NOT DIFFER, (2) POSITIVE INSTANCES ONLY APPEARED TO LEAD TO GREATER TRANSFER, AND (3) EXACT MATCHING PRODUCED THE GREATER NUMBER OF HIGHER-ORDER QUESTIONS BUT DID NOT TRANSFER TO A NEW LESSON. THIS PAPER WAS PRESENTED AT THE MEETING OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION (NEW YORK, FEBRUARY 16-18, 1967). (HA)

A COMPARISON OF DIFFERENT MODELING PROCEDURES IN THE ACQUISITION OF A TEACHING SKILL

D.W.Allen, D.C. Berliner, F.J.McDonald, and F.T.Sobol

Stanford University

Paper presented at the meeting of the

American Educational Research Association

February 16 - 18, 1967

New York City, New York

by

David C. Berliner

Introduction

Questioning, as an instructional technique, has been recommended to teachers since Socrates first used it to draw out ideas from students. The steady stream of books and monographs on the "Art of Questioning" that have appeared over the years (e.g., Young, 1853; Landon, 1899; Monroe and Carter, 1923; Sanders, 1966) attest to the belief that appropriate questioning behavior is an important teacher characteristic. A common theme in this literature is that questioning is the means by which the teacher stimulates thinking, by which he elicits the higher order mental processes such as critical judgment. Dewey (1933) pointed out that thinking itself is questioning. The critical requirement for a "good" classroom question is that it prompts students to use ideas rather than just remember them. The generally accepted premise is that the form of the question serves as the stimulus for eliciting certain kinds of cognitive activities which may range from simple recall to highly complex inferences from data.

Although some teachers intuitively ask questions of high quality, far too many over-emphasize those that require only the simplest cognitive activity on the part of the students. The purpose of the study reported here was to test training methods designed to modify teachers' behavior so that they asked questions eliciting complex cognitive activity. We designed a procedure to sensitize the novice teacher to the effects of questioning on his students and provided practice in forming and using higher order questions. The dependent variable measured in this study is the acquisition of a particular teaching skill, higher-order questioning.

This study is the fifth of a series in a research project that has approached teaching as a set of skills (i.e., behavioral performances) which must be mastered. The major purpose of this research is to investigate the training conditions which foster quick and lasting acquisition of the skill. A key feature of this research has been the use of both videotape models and practice in micro-teaching situations (c.f. McDonald, Allen and Seidman, 1967; Cooper and Stroud, 1967) . In this particular study we focused our attention on one practice and two modeling procedures which might be beneficial in the acquisition and transfer of a teaching skill.

Design of the Study

Under study, then, were three important questions about the conditions which aid in the acquisition of a teaching skill.

Question 1: The first of these inquiries was related to differences in skill acquisition as a function of exposure to perceptual or symbolic models.* By "exposure to a perceptual model" we mean that the learner has observed the actual performance of another person who displayed the behaviors to be acquired. That such a procedure would result in behavior modification without discrimination of the relevant cues and in the absence of known reinforcers is well substantiated by Bandura's studies of vicarious processes (1965).

* The expression "symbolic modeling" has been used in the psychological literature to describe the modeling function of some written materials. The terms symbolic and written models are synonymous in this report.

But when we adapt these ideas to training we also study or utilize the effects of two other factors: (1) discrimination training (Orme, McDonald, Allen, 1966), which makes the behaviors to be learned more salient for the learner, and (2) reinforcement for the acquisition of these behaviors.

In this study a short teaching segment is performed by a teacher model who has been programmed to emit the behaviors of the dependent variable in as great a frequency and in as many diverse teaching situations as possible. This model performance is videotaped and shown to trainees. As trainees watch the videotape of the model, salient characteristics of the model's behavior are pointed out.

Perceptual models, either live or on TV tape, may be unnecessarily rich in behaviors (that is, there may be many different cues to which to attend) when the particular skill to be acquired is purely verbal. A written transcription of the model's behavior, however, presents only the verbal components of these stimuli. The presentation of a master teacher's performance through complete transcripts of his teaching act may be thought of as symbolic modeling. This form of model presentation can preserve the naturalness of a classroom performance insofar as it is an accurate record of student and teacher verbal behavior. Moreover, the symbolic or written model is not a real-time model, as is true of the videotape model, so that the information contained in the transcript is received by the trainee at his own pace. The written record also allows the trainee to review continually what has previously been learned. For this study then our definition of written modeling is to provide to a trainee the complete

transcript of the perceptual model's verbal performance together with whatever student and supervisory verbal behaviors are found in the perceptual modeling condition. Half our subjects were exposed to perceptual (video) models and half were exposed to symbolic (written) models during training.

Question 2: The second inquiry of importance to us in developing a training technology was whether presenting negative and positive instances or only positive instances of the behavior-to-be-acquired had a more significant training effect. A training model could be programmed to present only positive instances of the behavior (in this case, only higher order questions). This procedure insures that the behavior to be acquired is not masked by other behaviors.

A model lesson which presents only positive instances (a pure lesson) may be contrasted with a lesson presenting both positive and negative instances, i.e., a mixed lesson. A mixed lesson, in which higher and lower order questions occur, allows for discrimination training with contrasting stimuli. Such a procedure may be more beneficial to a trainee if the contrast provided by a negative case serves to clarify the characteristics of the behavior to be learned.

For this experiment videotapes were prepared in which the model was programmed to emit only higher order questions. An alternate tape was also prepared in which the same model cast performed almost the same lesson. The only modification was the addition of lower order questions, inserted into the lesson following almost every higher order question. Half the subjects (S's) who were exposed to a perceptual model viewed a "pure lesson", one

containing only higher order questions. The other half of the S's who observed a perceptual model viewed a mixed lesson, one containing both higher and lower order questions. Subjects who received training with the written models also read either pure or mixed transcripted lessons.

Question 3: Our third inquiry was concerned with the nature of the practice (or performance opportunities) provided to the S's. In many studies of modeling or imitation learning the learning effect has been inferred from a measure of matching behaviors when the learner is placed in a situation identical to the situation encountered by the model.

In adapting the modeling process to teacher training, we must decide how closely we want the trainee to match the model. If matching the model, say by performing the same lesson in the same way as the model, facilitates acquisition of the skill, would transfer of the skill to dissimilar situations (i.e., different lessons) also be facilitated? Or, if S's practice the behavior of forming higher-order questions within a lesson of their own design, will their performance during acquisition or in transfer tests be superior to that of S's who taught the same lesson as the model? These questions were investigated experimentally by having half of the S's practice the skill by performing lessons of their own design, while the other half were directed to practice the skill by performing the model's lesson. The two practice conditions were labelled "own" and "model" lessons.

Overall Design: The three two-way classifications resulted in a 2 x 2 x 2 factorial design requiring eight experimental groups. Differences in main

effects associated with these three questions were examined for perceptual vs. symbolic and pure vs. mixed model lessons. We also evaluated the main effect attributable to practicing the skill in model vs. subject's own lesson.

Procedure: Except for foreign language and physical education majors, the entire Stanford Secondary Education Intern class of 1966-67 served as a pool for S's. All members of this group (N = 120) have obtained at least a bachelor's degree, are predominantly recent college graduates, and are in residence at Stanford for one year to obtain a Master's of Arts in Education. The experiment was conducted during the Intern's first quarter of residence, Summer 1966, on five consecutive Saturdays. One additional experimental day was needed late in the quarter for those S's who missed their regular training session.

Subjects were stratified according to curriculum and a random procedure was then used to assign S's to one of eight experimental conditions. Random procedures were used to assign S's to experimental classrooms who contained video recording operators, machines, and playback equipment. Classrooms were located at the University, and were familiar to all the S's as a result of their frequent microteaching experiences in these rooms during the summer. Subjects were assigned to the regularly scheduled experimental days on the basis of administrative convenience. Table I describes the original sample pool and the final sample utilized in the study.

TABLE 1

ORIGINAL AND FINAL SAMPLE SIZE BY SUBJECT MATTER AREAS

	Original Pool by Experimental Group Assignment									Final Sample by Experimental Group								
	1	2	3	4	5	6	7	8	Total	1	2	3	4	5	6	7	8	Total
Social Studies	6	6	5	4	6	6	5	4	42	4	5	5	7	4	4	2	3	34
English	5	5	5	6	4	4	4	7	40	6	4	5	4	3	5	3	7	37
Science	2	2	2	2	1	1	4	2	16	1	2	2	2	2	1	3	2	15
Mathematics	1	1	2	2	3	1	1	1	12	1	1	2	2	3	1	1	1	12
Drama							1	1	2						1		1	2
Art	1	1	1	1	1	2	1		8	1		1			1			3
Total	15	15	15	15	15	15	15		120	13	12	15	15	12	13	9	14	103

Because of rescheduling at the end of the study (an effort to balance the number of S's per condition) occasionally more S's in a given curriculum area were assigned to an experimental condition other than that to which they were originally assigned. The drop in numbers if due to scheduling conflicts, equipment malfunction, distortion or erasure of videotapes, and, in the case of 10 S's, refusal to participate. The final sample used for statistical evaluation is given by practice/teaching session and experimental group in Table 2.

TABLE 2

NUMBER OF VIDEOTAPES WHICH WERE RATED BY PRACTICE SESSION AND EXPERIMENTAL GROUP

		Session 1	2	3	4
Experimental Group	1	11	13	13	13
	2	12	12	12	11
	3	15	15	13	13
	4	13	15	15	12
	5	12	12	10	11
	6	13	12	13	13
	7	7	9	9	9
	8	14	14	14	11
Total		97	102	99	93

Table 3 describes the different treatments for each of the eight experimental groups.

TABLE 3

METHODS OF TRAINING FOR EACH EXPERIMENTAL GROUP

Model	Written				Perceptual			
Lesson	pure		mixed		pure		mixed	
Practice	own	model	own	model	own	model	own	model
Group Number	1	2	3	4	5	6	7	8

Four high school students, generally of the same grade level and generally two male and two females, were assigned on the morning of an experimental day to a team. Teams were then assigned to classrooms by a procedure which insured that all teams appeared in all rooms for different training sessions throughout the day.

Schedule: A typical training schedule for S's on any experimental day contained 15 steps and required two hours and 20 minutes. Six S's at a time would begin a training sequence at 8:00, 10:25, 1:00 or 3:25 on an experimental day.

Step 1: Written pretest of verbal fluency and flexibility. Two tests from the French kit (1963) were used to assess possible correlates of the ability to use higher order questions. The data from the pretest is still being analyzed and results are not yet available.

Step 2: Teaching Session One. Subjects presented a five minute lesson of their own choice which they had prepared in accordance with instructions to include questions to their students. Data from teaching session one provides a record of S's performance under these minimal pre-experimental instructions.

Step 3: Brief Instructions and Description of the Training. The schedule of training, and the nature of the training task was presented to the S in written form by the T.V. operator within the classroom.

Step 4: Presentation of Model Lesson One. Depending on their assignment to an experimental condition, S's were exposed to the video or written versions of a model performing a pure or mixed lesson.

During the presentation of the video model, a separate Wollensak audio recorder (synchronized to the videotape) was used to provide a standard commentary for the discrimination of important behaviors in the model's lesson. These comments were also used on the written transcripts of the lesson and appeared in the same form and place in the lesson.

Step 5: Play-back of Teaching Session One. All S's viewed a video recording of their first teaching session. During this self-viewing time (a performance feedback condition), the audio tape was used to direct the S's to pay attention to the form of the questions used in their lesson. The form of this commentary was non-contingent but task orienting.

Step 6: Planning for Teaching Session Two. Subjects were informed at this time that they would either teach their own lessons again or that they would teach the same lesson as the model. In both cases subjects were instructed to practice the training task.

Step 7: Teaching Session Two. Subjects taught their own lesson or the model lesson to a different team of students.

Step 8: Presentation of Model Lesson Two: Subjects were exposed to a different lesson by the model, in the same form and via the same media used in step four.

Step 9: Play-back of Teaching Session Two. Same as step five.

Step 10: Planning for Teaching Session Three. All S's continued to plan either own or model lessons.

Step 11: Teaching Session Three. Same as step seven.

Step 12: Review of Model Lessons One and Two. All S's repeated their experience in steps four and eight.

Step 13: Playback of Teaching Session Three. Same as step five.

Step 14: Planning of New Lesson. All S's were instructed to plan a completely new lesson on any topic they chose. They were to demonstrate transfer of the skill in their next teaching session.

Step 15: Teaching Session Four. In this final teaching segment in the training sequence, all S's were evaluated for transfer effects.

Rating

After the study was completed stenographers transcribed the audio recording of the tapes, producing typed protocols of the subjects' questions in each of the four teaching sessions. Ratings of the questions were made from the typed protocols rather than from the video recordings of the teaching sessions. This procedure was followed because we anticipated difficulty in rating questions as high or low order and we believed that the static written protocol would be more amenable to reliable rating than the dynamic video recording.

Three male secondary school teachers, all in graduate study at Stanford, rated the typed protocols. The decision rules for rating a question as higher or lower order were based on two substitution rules.

(1) Can you substitute the verbs "remember" or "describe" into the question? If so, the question is lower order. If not, and such verbs as "infer", "interpret", and "synthesize" can be substituted into the questions, the question is higher order.

We chose to regard simple recall, where answers are available in memory

as lower order. To illustrate: the question, "How many people were in the room?" calls for concrete facts available through memory. The question, "How many people are in the room?" calls for an answer which is available to direct observation. In neither case is cognitive activity at its most complex. However, the question, "Can you infer something about the relationship of the people in the room?" requires considerably more cognitive work and is classified as higher order.

(2) Determine whether the question asks the student to apply a rule or produce examples of a principle (lower order questions) or whether it forces him to find a rule or discover a principle (higher order questions).

Raters also had available three other categories into which they could classify questions. The first and second of these were "repeat of higher order questions" and "repeat of lower order questions." In each case the category was used when the teacher had asked a question, or repeated the basic question in similar form or sought additional answers from the students. The third category, infrequently used, was for incompleted, unintelligible, unclassifiable, or obviously rhetorical questions. Questions placed in this category were ignored in subsequent data analysis.

Raters, as a group, were given approximately 10 hours training, and then instructed to work independently on the protocols. Each rater categorized the questions in 391 protocols, working blind with respect to the experimental group of the S's, but with knowledge of the teaching session. The design of the experiment was not fully explained until the rating had been completed.

The protocols used in rating were compressed versions of the lessons presented by a teacher, and included the opening teacher comments for context, and then only the teacher's questions. Additional context and student answers

were not included in these protocols. Ratings of the questions had to be made on the basis of the rules stated above and the categorization of higher order questions under these rating conditions was quite conservative.

RESULTS

The reliability of the ratings was analyzed for each session, using an analysis of variance model described by Winer (1962). This analysis provides two correlation coefficients: (1) the reliability of a single measurement which approximate the mean of the intercorrelations between any pair of judges and (2) the reliability of the mean rating by the judges. This latter coefficient may be interpreted as if it were the correlation bet ween two sets of mean ratings for the same people, where two random samples of judges were used. The assumption, requiring that the judges be a random sample from a population of judges was violated in this study, and the estimates of reliability are, to some extent, in error. However, this correlation is perhaps the most appropriate way to describe the data since the scores assigned to S's were the average of the ratings of the three judges. The mean ratings by judges were chosen to provide a more stable estimate of the true scores of S's.

Table 4 provides reliability information about the rating categories used in subsequent analyses. The adjusted correlations represent a correction for differences in frame of reference of the judges. Because the zero points on our scale of measurement are not as important as the dispersion or order of scores, any systematic variation between judges need not be considered as part of the error of measurement. Where large adjustments of the correlations appear in Table 4 it may be assumed that large differences in frame of reference between the judges are operating.

UNADJUSTED AND ADJUSTED RELIABILITY OF

THE AVERAGE SCORE ASSIGNED BY THREE JUDGES

TO FOUR CATEGORIES IN EACH TRAINING SESSION *

CATEGORY	TEACHING SESSION 1	2	3	4
Number of higher order questions	(.694) .745	(.564) .657	(.276) .660	(.325) .629
Number of higher and repeat of higher order questions	(.612) .683	(.539) .653	(.207) .678	(.360) .627
Number of lower order questions	(.859) .910	(.854) .889	(.312) .717	(.639) .874
Number of lower and repeat of lower order questions	(.880) .919	(.865) .896	(.333) .745	(.654) .865

*Unadjusted reliability coefficients are in parentheses

The reliability estimates of the average intercorrelation between any pair of judges which were also obtained, were low before adjustment and reflected only moderate inter-rater agreement after adjustment.

The dependent variable chosen for analysis was the total number of higher order (and repeat of higher order) questions, divided by the total number of questions asked in a teaching session. The transformation of the original values into percentages allows for ceiling effects which are imposed by the five minute time limit on a teaching session.

The data for Teaching session One were analyzed using a simple one-way analysis of variance between the eight experimental groups. Table 5 shows that in this test of initial differences between groups, the null hypothesis of no difference may be accepted.

TABLE 5

ONE WAY ANALYSIS OF VARIANCE ON EIGHT EXPERIMENTAL GROUPS IN SESSION ONE

Source	SS	DF	MS	F
Between Groups	.0561	7	.0080	.33
Within Groups	2.1415	89	.0241	
Total	2.1976	96		

Using Session I as a base, we may measure the effects of training over the four sessions. This information graphically displayed in Figure I.

FIGURE 1 (see next page, Page 16)

The training effects show, for example, that within experimental groups five and eight, the increase in per cent of higher order questions asked by a teacher has gone from 19 and 21 per cent in Session I to 51 and 55 per cent of all questions asked in Session III.

The Wilcoxon matched-pair signed-rank tests (Siegel, 1956) was used to determine the probability associated with changes between sessions. This information is included as Table 6.

TABLE 6

WILCOXON MATCHED-PAIRS SIGNED-RANK DATA ON CHANGES BETWEEN SESSIONS FOR EIGHT EXPERIMENTAL GROUPS

Difference	T Value	Probability	Direction of Difference
Session 1 and Session 2	0	.005*	S2 S1
Session 1 and Session 3	0	.005*	S3 S1
Session 1 and Session 4	0	.005*	S4 S1
Session 2 and Session 3	0	.005*	S3 S2
Session 2 and Session 4	18	NS	
Session 3 and Session 4	5	NS	

*One tail tests of significance

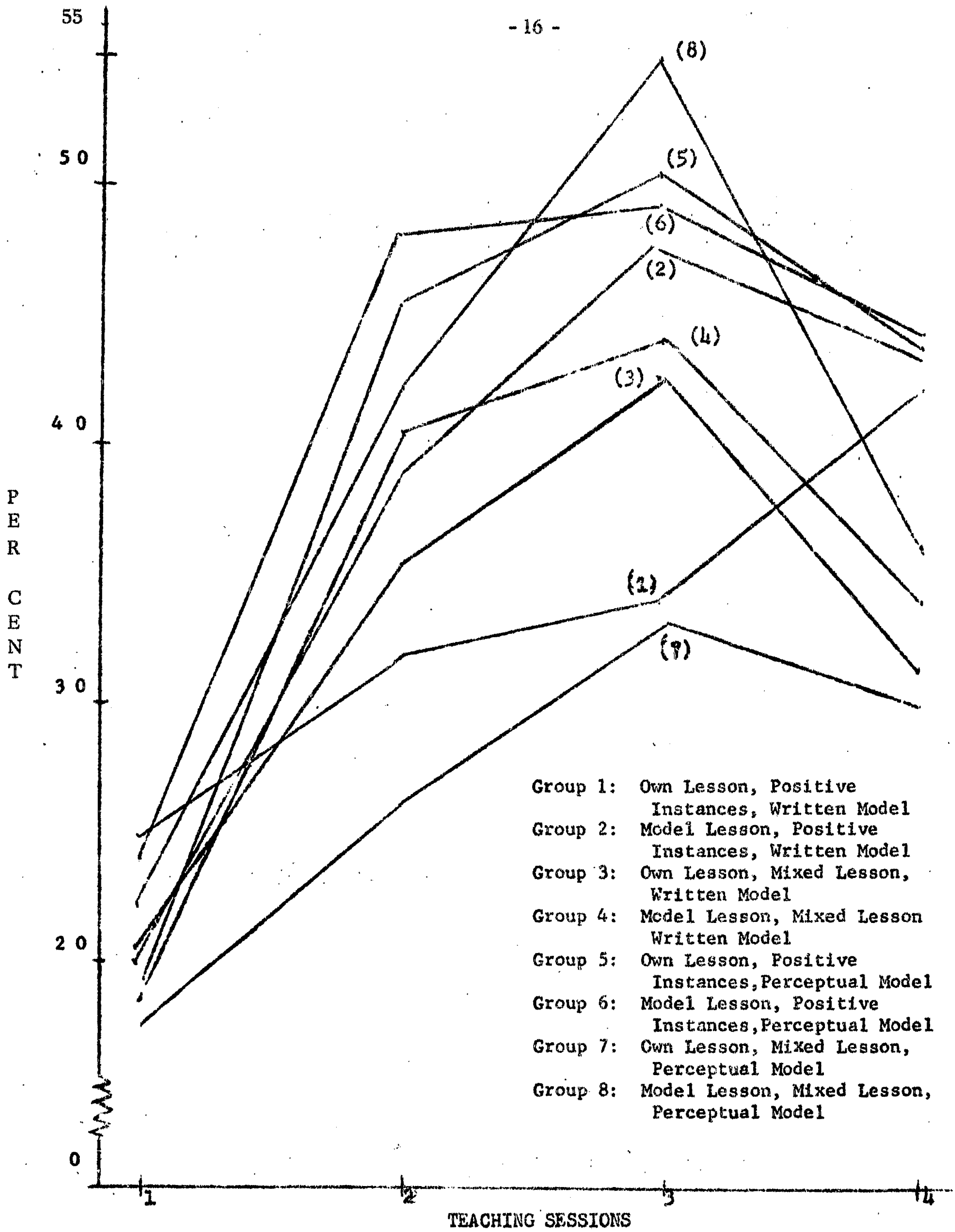

Fig.1. Mean percent of higher order questions for eight experimental groups on four teaching sessions

The pairs used in this analysis were the eight means in one session versus the eight means for another session. One-tail tests were appropriate for testing the significance of the changes between sessions since the direction of change was predicted. As predicted, and displayed in Figure one, acquisition of the skill continued over the three trials. A drop-off in performance in session four was also predicted since lessons were switched. However, with this test, the null hypothesis of no difference between session three and session four, and no difference between session two and session four, is accepted. While the first interpretation is to claim no significant drop off in session four, figure one shows this to be primarily a function of the continuous rise, over all training sessions, of Group one. Seven other experimental groups did fall off between sessions three and four.

Tables seven, eight and nine present the analysis of variance tables for Sessions two, three, and four.

TABLE 7

ANALYSIS OF VARIANCE FOR PER CENT OF HIGHER ORDER QUESTIONS IN TEACHING SESSION TWO

Source	DF	MS	F
Model	1	.03827	1.23
Lesson	1	.06247	2.00
Practice	1	.14987	4.81*
M X L	1	.13525	4.34*
M X P	1	.00819	.26
L X P	1	.03651	.69
M X L S P	1	.03651	1.17
Error	94	.00312	

* Significant beyond the .05 level.

TABLE 8

ANALYSIS OF VARIANCE FOR PER CENT OF HIGHER ORDER QUESTIONS IN TEACHING SESSION THREE

Source	DF	MS	F
Model	1	.05357	2.30
Lesson	1	.00711	.31
Practice	1	.18663	8.00**
M X L	1	.03799	1.63
M X P	1	.00712	.31
L X P	1	.01544	.66
M X L X P	1	.20357	8.73**
Error	91	.02332	

** Significant beyond the .01 level

TABLE 9

ANALYSIS OF VARIANCE FOR PER CENT OF HIGHER ORDER QUESTIONS IN TEACHING SESSION · FOUR

Source	DF	MS	F
Model	1	.00026	.01
Lesson	1	.27601	10.92**
Practice	1	.01549	.61
M X L	1	.00042	.02
M X P	1	.00097	.04
L X P	1	.00654	.26
M X L X P	1	.00144	.06
Error	85	.02528	

** Significant beyond the .01 level

The tests of main effects were our primary interest, and the significant F test for the practice variable in session two and again in session three indicates that for the acquisition phase, matching the model lesson very closely does produce a higher percentage of higher order questions. Interestingly though, on transfer to another teaching task, such consistent differences in practice do not hold up. The significant interaction of model and lesson-type in session two is difficult to interpret. The significant three-way interaction is session three is even more difficult to interpret.

In session four, on the transfer task, a main effect for lesson-type was found. Although this effect did not appear in the data for other sessions, the training done with the "pure" lesson, utilizing only positive instances of the skill, aided S's in utilizing the skill in new teaching situations.

No significant differences were found related to the media with which the model was presented. It appears that for a verbal skill, such as higher order questioning, the video technology used to present the model may be superfluous, and that the training may be readily accomplished through written models. (However, an overall training effect may be mediated by the constant treatment (in this study) of self-viewing by means of a videotape. We do not have information about such an effect.)

CONCLUSIONS

For this sample of Stanford secondary teaching interns, training in the use of higher order questions was successful. As measured by the variable per cent of higher order questions used in a five minute teaching session, all experimental groups showed significant training effects. It appears that for

this skill the perceptual model was no more efficient as a training agent than the written model. Whether this would be true for other skills, especially those involving motor behaviors, is not yet known.

The careful matching of the model lesson during acquisition was effective in producing a greater number of higher order questions. However, no transfer to a new lesson was noted for this kind of practice. Whether this lack of an effect would persist if more practice were given is not known. All curves were still rising on session three, indicating, perhaps, that another training session might have been useful before the attempt was made to transfer. The effectiveness of training with only positive instances of the skill appears to show up when a transfer test is made. This finding, if it remains after replication, may prove quite important in developing our training technology for pre-service teacher training.

Bibliography

Bandura, A. "Vicarious processes: A case of no-trial learning." in L. Berkowitz (ed), Advances in experimental social psychology. Vol II. New York: Academic Press, 1965. pp.1-55.

Cooper, J. and Stroud, T., The Stanford Summer Micro-Teaching Clinic, 1966. A paper presented at the American Educational Research Association annual meeting, February, 1967.

Dewey, J. How we think. Boston: D.C. Heath, 1933.

French, J. Manual for kit of reference tests for cognitive factors. Princeton, New Jersey: Educational Testing Service, 1963.

Landon, J. A manual of the art of questioning. Syracuse, N.Y.: C.W. Bordeen, 1899.

McDonald, F.J., Allen, D.W., and Orme, M.E., The effects of self-feedback and reinforcement on the acquisition of a teaching skill. A paper presented at the annual meeting of the American Educational Research Association, February, 1966.

McDonald, F.J., Allen, D.W., and Siedman, E., Televised models for Teacher Training. A paper presented at the American Educational Research Association annual meeting, February, 1967.

Monroe, W.S., and Carter, R. The use of different types of thought questions in secondary schools and their relative difficulty for students. Urbana, Ill.: Bureau of Educational Research, Univ. of Ill., Bull. No. 14, 1923.

Sanders, N.M. Classroom questions. New York: Harper and Row, 1966.

Siegel, S. Nonparametric statistics. New York: McGraw-Hill, 1956.

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

Young, W.T., The Art of putting questions. Syracuse, New York: C.W. Bordeen, 1853.