DOCUMENT RESUME

ED 117 089 95 SP 009 789

AUTHOR Clark, Christopher M.; And Others
TITLE Three Experiments on Learning to Teach. Research and Development Memorandum No. 140.
INSTITUTION Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Dec 75
CONTRACT NE-C-00-3-0061
NOTE 32p.

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS Academic Achievement; Affective Behavior; Effective Teaching; *Learning; *Students; *Teachers; *Teaching Experience; Teaching Procedures; *Teaching Skills; Teaching Styles

ABSTRACT

This paper reports on three studies of the changes in teacher effects on student learning and attitudes as a function of teacher practice. Two studies were performed in a tutorial context, the first with 4 tutors and 16 students, the second with 17 tutors and 68 students. The third study used small group instruction with 12 teachers and 288 students. In each study, the same teachers taught a short curriculum unit three or four times, teaching different students in each trial. The results of the studies showed marked individual differences among teachers in the extent to which their students learned more or less. Few teachers showed marked increases in student learning with practice. It was also shown that, on the average, student achievement decreased with teacher practice. Student attitudes toward the learning situation and toward themselves as learners changed in a curvilinear fashion. (Several graphs are included to illustrate the data.) (Author/RC)

STANFORD CENTER
FOR RESEARCH AND DEVELOPMENT
IN TEACHING

Research and Development Memorandum No. 140

THREE EXPERIMENTS ON LEARNING TO TEACH

Christopher M. Clark, Richard E. Snow,
and Richard J. Shavelson

School of Education
Stanford University
Stanford, California

December 1975

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the National Institute of Education, U. S. Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. NE-C-00-3-0061.)

Introductory Statement

The Center's mission is to improve teaching in American schools. Its work is carried out through three research and development programs--Teaching Effectiveness, The Environment for Teaching, and Teaching and Linguistic Pluralism--and a technical assistance program, the Stanford Urban/Rural Leadership Training Institute. A program of Exploratory and Related Studies includes smaller studies not included in the major programs. The ERIC Clearinghouse on Information Resources is also a part of the Center.

This paper is part of the work of the Program on Teaching Effectiveness.

Contents

THREE EXPERIMENTS ON LEARNING TO TEACH[1]

Christopher M. Clark
Stanford University

Richard E. Snow
Stanford University

Richard J. Shavelson
UCLA

To what extent is training really necessary for teachers? Beginning teachers frequently criticize their professional preparation on the grounds that it is largely irrelevant to their needs in the classroom. Veteran teachers often feel that experience was their best teacher--that the best preparation for teaching is teaching itself. Teachers at the junior high school, secondary, and college levels who teach several sections of the same course often report that their second or third time through a lesson seems more effective than the first. In this connection, Medley argues that

> the effective teacher will differ from the ineffective teacher primarily in his control over the repertory of competencies he commands; in his ability to adapt his behavior to the pupils, the purpose, and the situation in which he operates; and most important of all in his ability to learn from his own experience. (Medley, 1970)

The process of learning from experience has also been investigated in the psychological laboratory. In 1949, Harlow published a paper entitled "The Formation of Learning Sets," in which he showed that monkeys and children improve with practice across a series of similar learning problems. That is, they learned how to learn a particular class of tasks. Learning-to-learn became an important concept in learning research (see,

---

[1]Based on "Experiments on Learning to Teach," presented at the Annual Meeting of the American Psychological Association, Chicago, September 1975.

for example, Harlow, 1959; Reese, 1964). The present research sought to determine whether teachers learned to teach, using a paradigm similar to that of the Harlow studies. The basic question was: Do teacher effects on student learning improve as teachers gain experience in a particular teaching situation? That is, do teachers teach more effectively (as measured by amount and kinds of student learning) the second or third time they teach a given curriculum unit than they do the first time? Three studies of this question were undertaken. The first examined this question alone in a tutorial context. The second added a training variable, again in a turorial context. The third was conducted in a small group instruction situation. The studies were performed as part of the Program on Teaching Effectiveness, Stanford Center for Research and Development in Teaching.

The general hypothesis for the three experiments is represented in Figure 1. In each study, the same teachers taught a short curriculum unit three or four times (trials I-IV). Different students were taught in each trial. The predicted effect of teacher practice was that students taught on the later trials would learn more and express more favorable attitudes about the situation and themselves than students taught in the earlier trials.

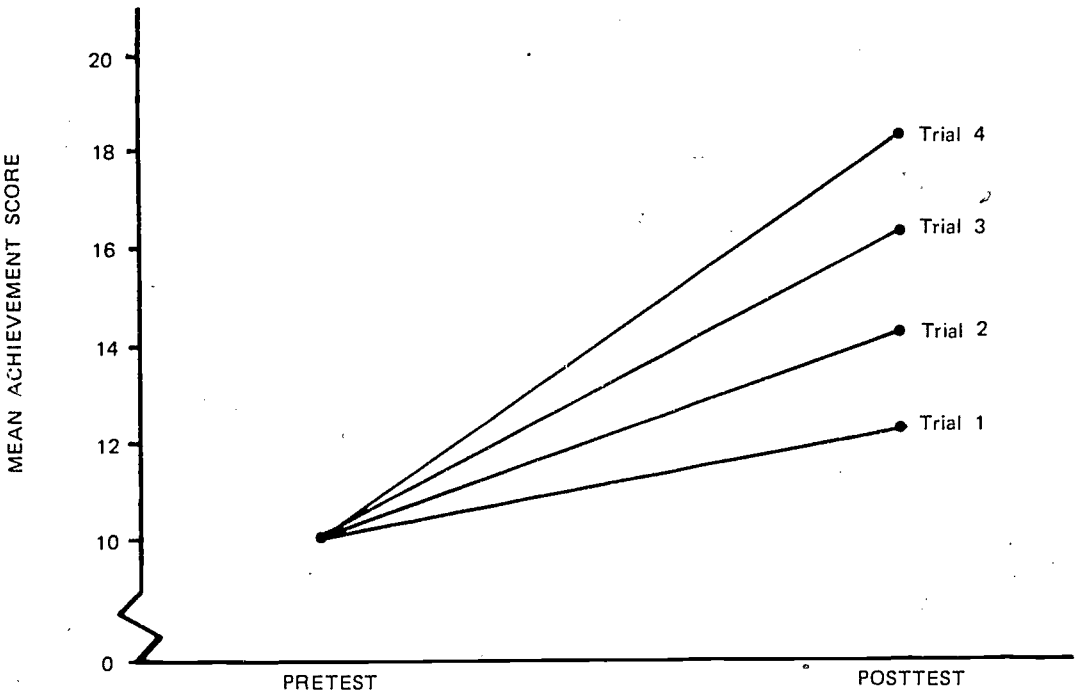

Fig. 1. Predicted pattern of change in student learning as a function of teacher practice over four trials.

The First Study: Tutoring

In the first study, four advanced graduate students in education tutored junior high school students, one at a time, in physics. Each tutor taught one student for five one-hour sessions, and then was assigned randomly to a new student the following week to repeat the material. Each tutor taught four students successively. Thus, 16 students participated. Figure 2 shows the design of the first study.

Each student took an achievement test before and after instruction. Table 1 summarizes the results; Figure 3 shows the changes in mean student achievement by week. These results indicate that, on the average, the

Tutor	Week I	II	III	IV
1	S-1	S-5	S-9	S-13
2	S-2	S-6	S-10	S-14
3	S-3	S-7	S-11	S-15
4	S-4	S-8	S-12	S-16

Fig. 2. Design of the first study. Four tutors taught one student per week over four weeks.

TABLE 1

Students' Pretest and Posttest Achievement Scores (possible range, 0-36)

Week	Tutor	Student	Achievement Pretest Score	Achievement Posttest Score	Difference
I	A	1	12	10	-2
	B	2	7	13	6
	C	3	16	15	-1
	D	4	7	11	4
II	A	5	14	22	8
	B	6	9	14	5
	C	7	9	8	-1
	D	8	12	22	10
III	A	9	6	15	9
	B	10	14	18	4
	C	11	15	22	7
	D	12	14	19	5
IV	A	13	12	18	6
	B	14	13	13	0
	C	15	25	27	2
	D	16	17	19	2

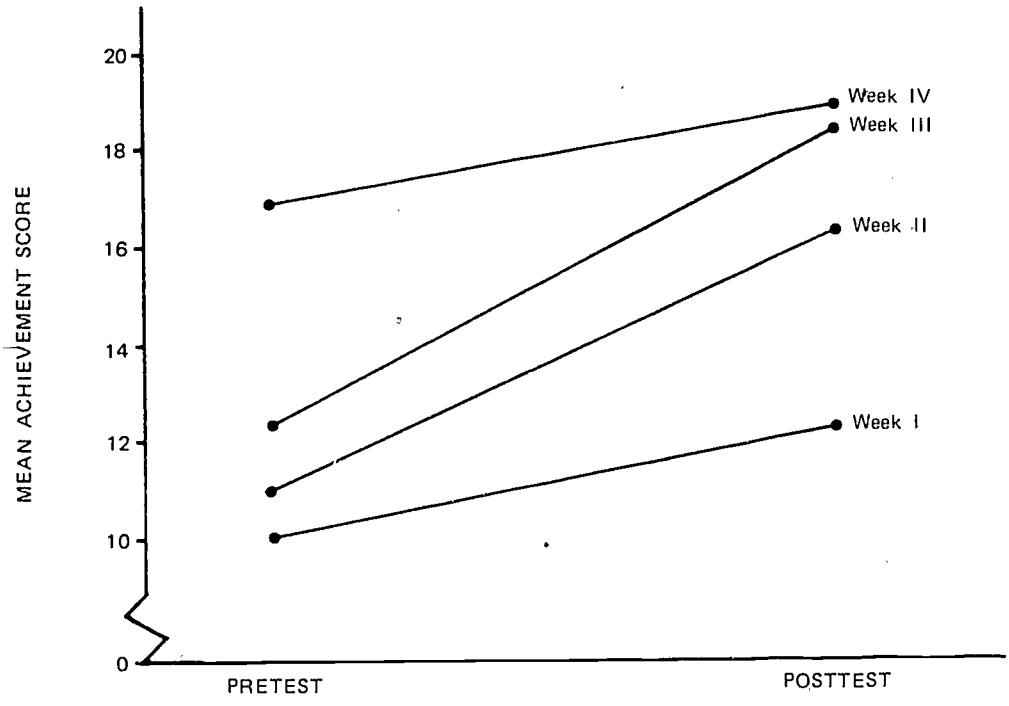

Fig. 3. Changes in mean student achievement from pretest to posttest by week. (N=16 students, 4 per week)

teachers obtained higher achievement from their students in each successive week. But the effect was weak and was marred by one student's unusually high pretest score in Week IV. The results were considered encouraging enough to attempt a larger experiment.

The Second Study: Tutoring

The second study was similar in design to the first. Here, however, 17 participating tutors were volunteer undergraduates, nine of whom received 12 hours of training in tutoring skills before teaching. The tutor training consisted of exercises on questioning, listening, and explaining in a tutorial context (see Clark, 1972).

Tutors	Tutor Training Workshop	Week I	II	III	IV
Experimental Group (N = 9)	X	S1-S9	S18-S26	S35-S43	S52-S60
Control Group (N = 8)	0	S10-S17	S27-S34	S44-S51	S61-S68

Fig. 4. Design of the second study. Seventeen tutors (nine trained and eight untrained) taught one student per week over four weeks. Total N = 68 students (S).

Again, junior high school students (N = 68) were randomly assigned one at a time to a tutor for one week's physics instruction; this continued for four weeks (see Figure 4). Because laboratory space was not available, tutors had to meet their students outside of school. An attempt was made to record all tutoring sessions on audiotape for analysis of teaching process variables. Achievement and attitude measures were administered pre and post.

Table 2 shows the mean pre and post achievement scores for students of tutors with and without training. Figure 5 shows the same results graphically.

Table 3 shows the students' mean pre and post scores on their ratings of themselves as learners (General Scale), themselves as students of physics (Physics Scale), and themselves as students in a tutoring situation (Tutoring Scale). Figure 6 shows these changes in student self-rating scores graphically.

TABLE 2

Means and Standard Deviations of Physics Achievement Scores

Week	Group		Pretest	Posttest
I	E (N=9)	$\bar{X}$	7.89	12.56
		SD	2.98	3.71
	C (N=8)	$\bar{X}$	10.38	13.13
		SD	2.20	1.55
	Both (N=17)	$\bar{X}$	9.06	12.82
		SD	2.86	2.83
II	E	$\bar{X}$	7.11	11.67
		SD	1.27	2.00
	C	$\bar{X}$	8.38	12.63
		SD	2.13	4.31
	Both	$\bar{X}$	7.71	12.12
		SD	1.79	3.22
III	E	$\bar{X}$	9.67	13.44
		SD	2.87	2.40
	C	$\bar{X}$	9.13	14.25
		SD	2.23	2.31
	Both	$\bar{X}$	9.41	13.82
		SD	2.53	2.32
IV	E	$\bar{X}$	7.89	11.00
		SD	2.32	2.40
	C	$\bar{X}$	10.25	14.38
		SD	2.43	2.92
	Both	$\bar{X}$	9.00	12.59
		SD	2.60	3.10

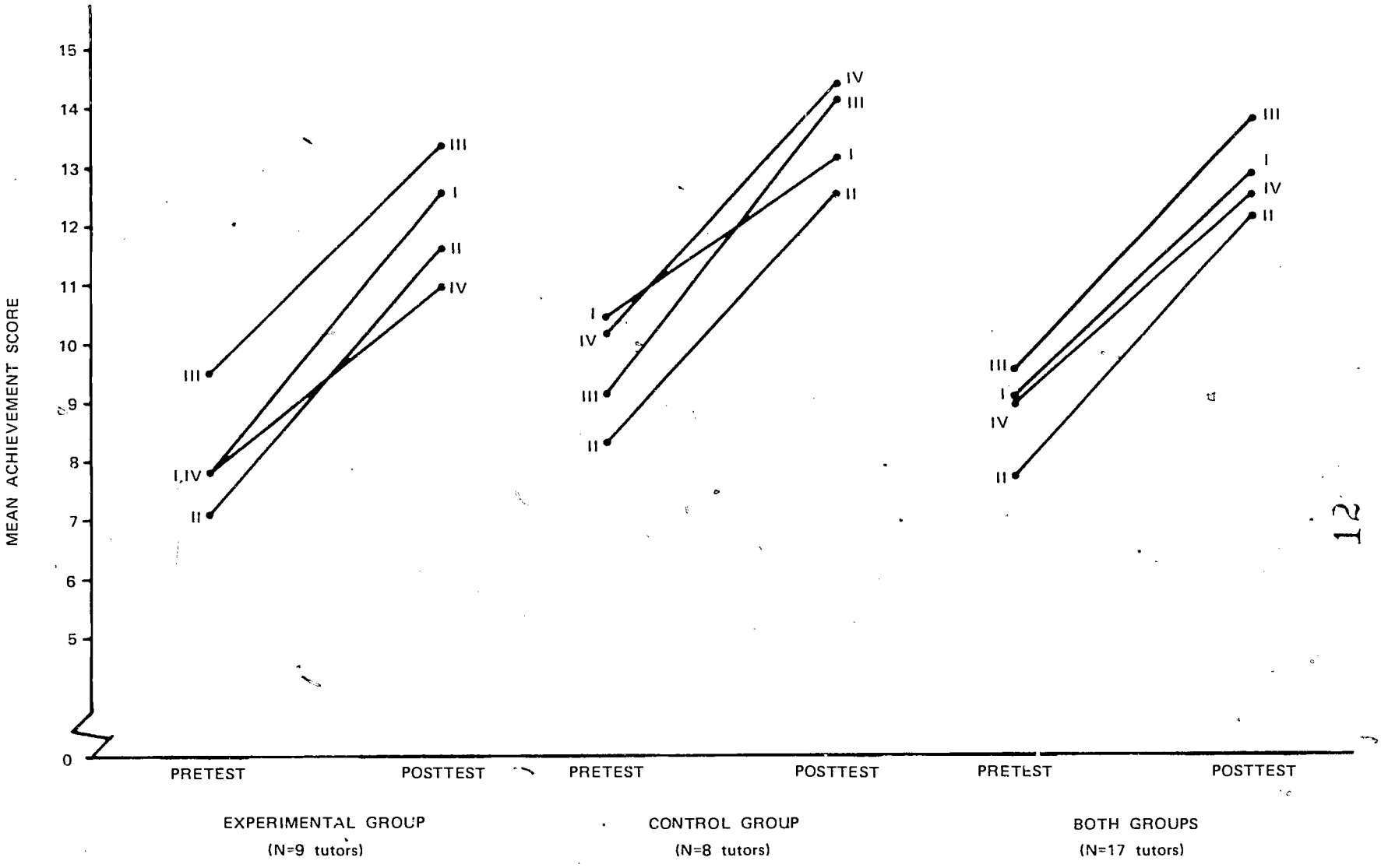

Fig. 5. Changes in mean student achievement from pretest to posttest by week. (Roman numerals indicate weeks.)

TABLE 3

Means and Standard Deviations of Students' Scores on Self-Rating Inventory

Week	Self-Rating Scale		Pretest	Posttest
I	General	$\bar{X}$	42.88	43.24
		SD	11.44	14.20
	Physics	$\bar{X}$	51.53	57.24
		SD	18.20	21.40
	Tutoring	$\bar{X}$	33.59	36.76
		SD	9.43	11.57
II	General	$\bar{X}$	44.88	46.65
		SD	8.31	8.91
	Physics	$\bar{X}$	54.12	58.24
		SD	13.69	10.89
	Tutoring	$\bar{X}$	35.29	37.24
		SD	8.14	6.61
III	General	$\bar{X}$	42.94	42.53
		SD	8.94	10.21
	Physics	$\bar{X}$	48.59	53.76
		SD	11.71	13.00
	Tutoring	$\bar{X}$	33.06	36.59
		SD	7.42	8.65
IV	General	$\bar{X}$=	44.35	48.47
		SD	8.50	10.82
	Physics	$\bar{X}$	49.76	58.41
		SD	13.22	14.44
	Tutoring	$\bar{X}$	35.47	40.82
		SD	6.82	7.68

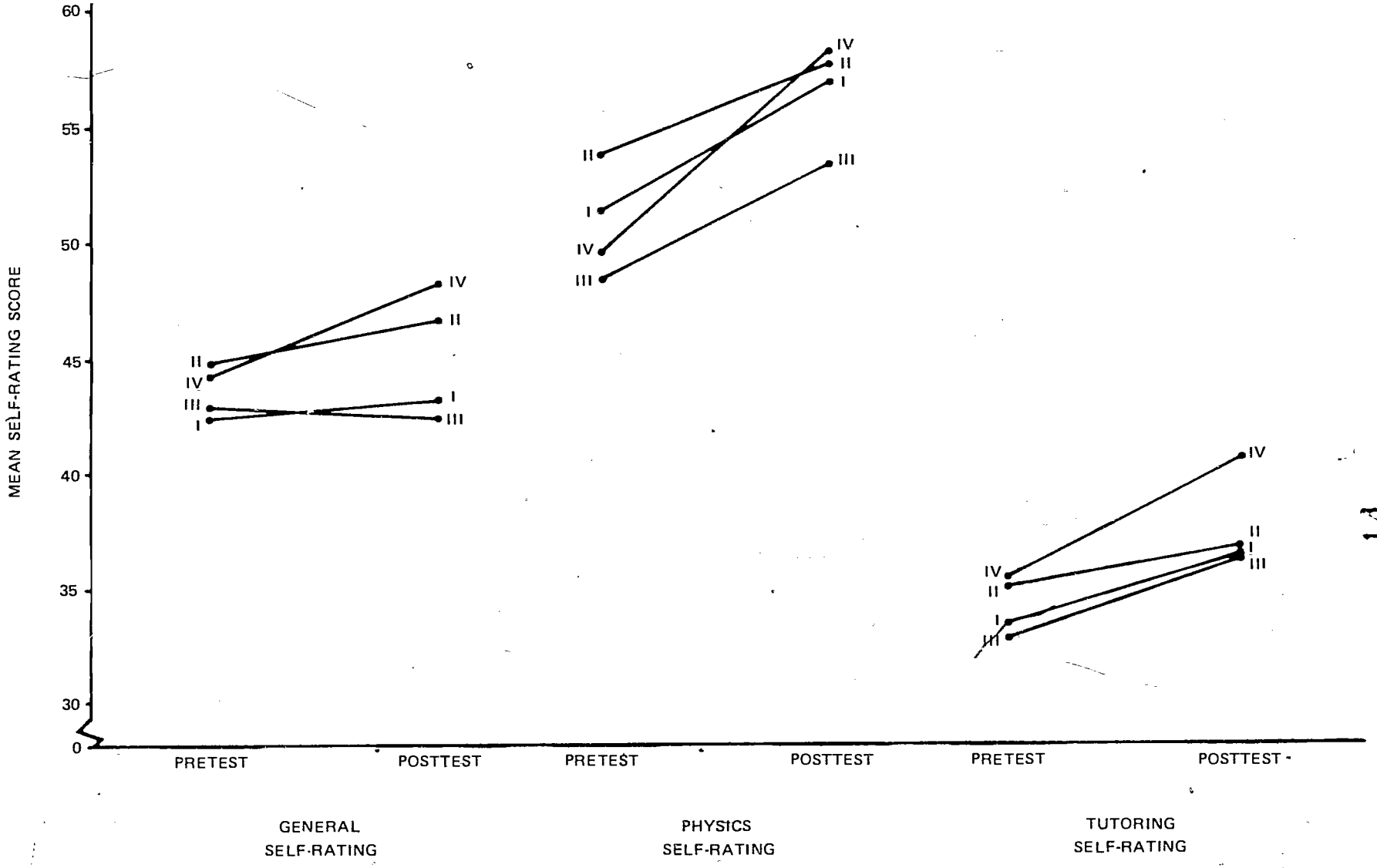

Fig. 6. Changes in mean student scores on the Self-Rating Inventory from pretest to posttest by week. (N=68 students, 17 per week)

The results of the second study were discouraging. Although some tutors improved with practice on some student outcome measures, performance curves were erratic and there was no apparent average improvement. In fact, analyses of variance and covariance (not shown) revealed that on the average, control tutors (those who were not specially trained) obtained higher achievement from their students than did trained tutors. There was some indication that this occurred because the trained tutors spent more time on their own skill practice and less on instruction than the control tutors. But this possibility could not be verified because the audiotapes were of poor quality. It was also noted that the tutoring situation allowed variation in interaction patterns so wide as to smother other effects of interest. In addition, since students had access to the physics text material, the effects of tutoring and the effects of studying were confounded.

The Third Study: Small Group Instruction

The third study was conducted in laboratory facilities which allowed small group instruction and extensive videotaping. Experienced teachers (six men, six women) taught the same social studies unit to three successive groups (hereafter called classes) of eight students each (N = 288 students). For each class, the social studies unit was taught in three 50-minute periods in one day. Students were posttested at the end of each day using a multiple-choice recall test, an essay test, and an attitude inventory. Achievement pretests were not used, but aptitude data were collected on all teachers and students before instruction. (For a more detailed report of this study, see Clark, forthcoming.)

The students (116 boys and 196 girls) were paid volunteers who had just completed seventh, eighth, or ninth grade. They were assigned to classes on a stratified random basis, stratifying on verbal ability and sex, but not grade level. The design of the third study is shown in Figure 7.

Three days before teaching for the first time, each teacher was given two hours to read the printed curriculum material and examine a set of color transparencies which were to be used in the teaching sessions. In addition, each teacher was given a list of objectives to be achieved by the students. At the beginning of each teaching day, each teacher was given 90 minutes to plan the teaching session. The students were given one hour to read the printed curriculum material and view the color transparencies before entering class.

The student posttest data for each day of teaching are summarized in Table 4. To test the effects of teacher practice on student posttest variables, a two-factor analysis of variance (Teacher X Day) was performed on each variable. Scores on the Recall Test were adjusted for differences in vocabulary and reasoning ability by analysis of covariance.

	Day I			Day II			Day III		
Teacher	Students Read Text Material	Teachers Teach Three 50-minute Periods	Student Posttests	Students Read Text Material	Teachers Teach Three 50-minute Periods	Student Posttests	Students Read Text Material	Teachers Teach Three 50-minute Periods	Student Posttests
1	X	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X	X
4	X	X	X	X	X	X	X	X	X
5	X	X	X	X	X	X	X	X	X
6	X	X	X	X	X	X	X	X	X
7	X	X	X	X	X	X	X	X	X
8	X	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	X	X
10	X	X	X	X	X	X	X	X	X
11	X	X	X	X	X	X	X	X	X
12	X	X	X	X	X	X	X	X	X

Fig. 7. Experimental design for the third study: Two factors (Practice X Teacher) with repeated measures on the teacher factor. A new group of eight students was assigned to each teacher each day.

TABLE 4

Means and Standard Deviations of Student Posttest Variables

Posttest Variables		Day I	Day II	Day III
Recall Test[a]	$\bar{X}$	39.51	40.44	37.56
	SD	7.35	7.08	9.20
Essay Test, Concrete	$\bar{X}$	12.22	11.67	11.15
	SD	5.69	4.74	5.35
Essay Test, Abstract	$\bar{X}$	2.04	2.09	2.20
	SD	1.74	1.66	1.88
Attitude Toward the Situation	$\bar{X}$	57.79	59.15	55.32
	SD	11.06	8.84	10.44
Attitude Toward Self	$\bar{X}$	15.37	15.89	15.21
	SD	2.88	2.68	2.72

Note: N = 288 students, 96 each day.

[a]Recall test scores are adjusted for vocabulary and reasoning ability.

For the Recall Test, on the average, adjusted class means were relatively stable from Day I to Day II and then decreased significantly from Day II to Day III. That is, the classes taught on the first two days tended to score highest, and the classes taught on the third day tended to score lowest. The absence of a significant teacher effect and Teacher X Day interaction indicates that, within days, teacher effectiveness as measured by the Recall Test was similar across teachers.

The essay test yielded two scores: number of Concrete themes (i.e., number of facts mentioned) and number of Abstract themes (i.e., number of principles, extrapolations from facts, etc., mentioned). A two-way analysis of variance (Teacher X Day) indicated that these scores

were stable across teacher and day. That is, neither amount of teacher practice nor the effects of particular teachers were reflected by these variables.

The attitude measure also had two scales. Scores on Attitude Toward the Situation (i.e., toward the teacher, subject matter, and learning environment) showed a mean increase from Day I to Day II and a decrease from Day II to Day III. That is, the classes taught on the teachers' second day of teaching expressed more positive attitudes toward the teaching situation than did the classes taught on the first and third days of teaching. Within a given day, class mean scores on this variable differed significantly from one another, indicating a significant teacher effect. This is especially noticeable on Day III, where mean scores for different teachers range from 37.75 to 63.62. For this variable, the Teacher X Day interaction was also significant. That is, the unique combination of a particular teacher and a particular day was associated with a particular Attitude Toward the Situation.

Mean scores on Attitude Toward Self as learner (e.g., satisfaction with the student's own participation) did not change significantly across days. A significant teacher effect for this variable; however, indicated that over all days class means were significantly different from one another.

In all of these trends, there were some individual differences among teachers. The scores of some increased with practice on some variables, suggesting positive "learning to teach." Others showed negative "learning to teach." Figures 8 through 12 display class means for each teacher for each day on each student posttest variable.

For adjusted Recall Test mean scores (Figure 8), two patterns predominated. Four teachers (Teachers 1, 2, 4, and 8) were relatively ineffective on Day 1, but then were among the most effective on Day II. On Day III, these four teachers declined in effectiveness, but not to the low level of Day I (with the exception of Teacher 4). The remaining eight teachers tended to decline progressively from Day I to Day III. Only one of the 12 teachers (Teacher 3) improved in effectiveness from Day II to Day III, and this was only a very slight improvement.

For Essay Test Concrete, the picture is more complex (Figure 9). Four teachers' (Teachers 2, 3, 4, and 8) class scores decreased on Day II and then increased on Day III. Teachers 1, 7, and 11 exhibited the opposite patterns, increasing on Day II and dropping on Day III. The mean class scores increased systematically with practice for two teachers (6 and 10), and decreased systematically with practice for three teachers (5, 9, and 12).

The patterns for Essay Test Abstract (Figure 10) are similar to those for Essay Test Concrete. Four teachers (Teachers 2, 3, 9, and 12) increased in class scores on Day II and decreased on Day III. None of these teachers had exhibited this pattern for Essay Test Concrete. Three teachers' effectiveness decreased on Day II and increased on Day III (Teachers 4, 5, and 11). Only Teacher 4 had previously exhibited this pattern on the Essay Test Concrete variable. Teachers 1, 7, and 8 systematically increased in effectiveness with practice, and Teachers 6 and 10 changed very little with practice.

The dominant pattern for students' scores on Attitude Toward the Situation (Figure 11) shows an increase on Day II and a decrease on Day

III (Teachers 1, 2, 7, and 8). The opposite pattern, decreasing on Day II and increasing on Day III was manifested by Teachers 3, 5, and 12. Generally decreasing trends were shown by Teachers 6, 9, and 11, with Teachers 9 and 11 decreasing precipitously on Day III and Teacher 6 decreasing more sharply on Day II than on Day III.

Attitude Toward Self mean scores (Figure 12) produced patterns similar to Attitude Toward the Situation. Again, the dominant pattern (Teachers 2, 7, 10, and 11) involved an increase on Day II and a decrease on Day III. Only Teacher 6 showed the opposite pattern, decreasing on Day II and increasing again on Day III. Four teachers (Teachers 1, 4, 9, and 12) showed a generally decreasing pattern with practice, and three (Teachers 3, 5, and 8) showed increased scores with practice.

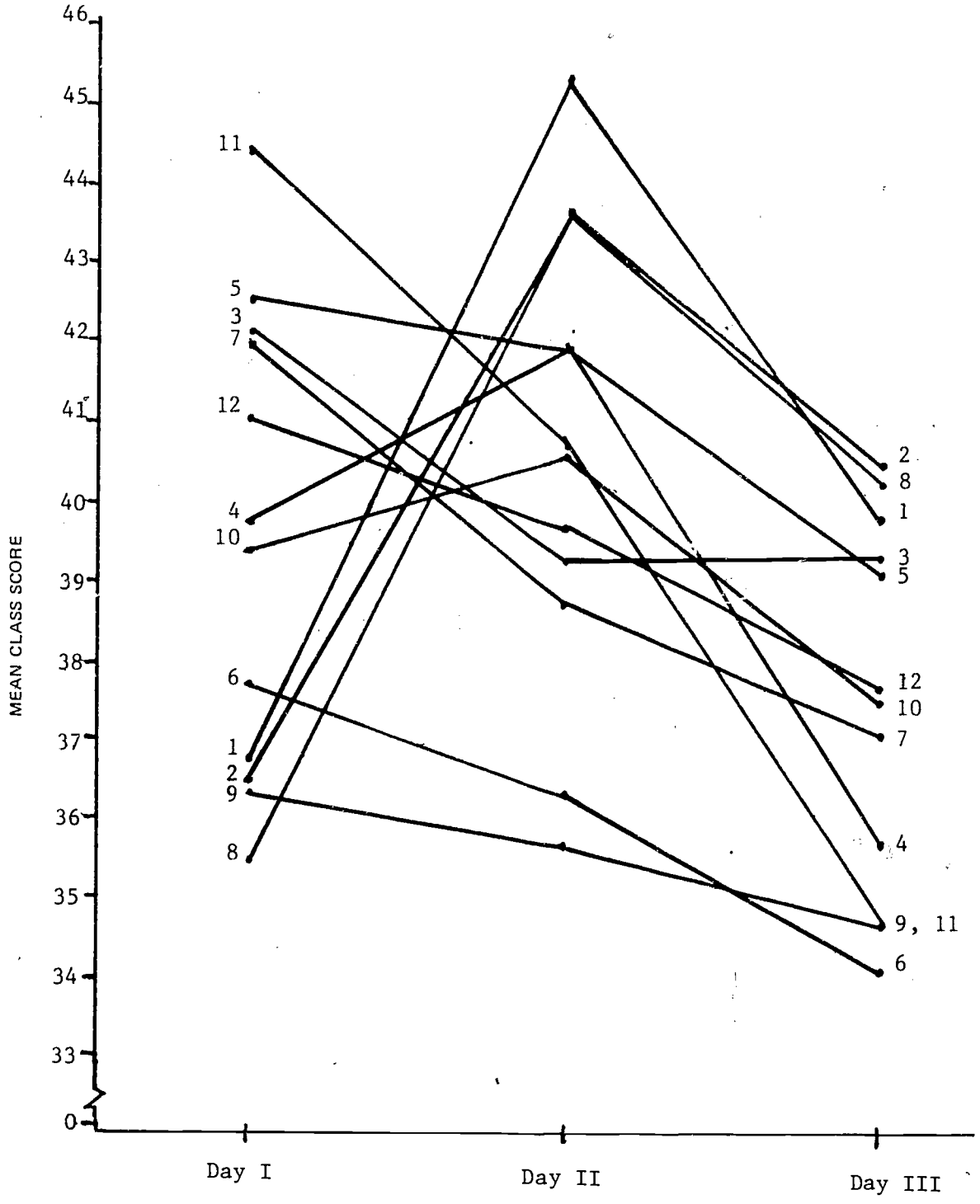

Fig. 8. Mean class scores on Recall Test for Days I to III, adjusted for vocabulary and reasoning ability. (Numbers identify individual teachers.)

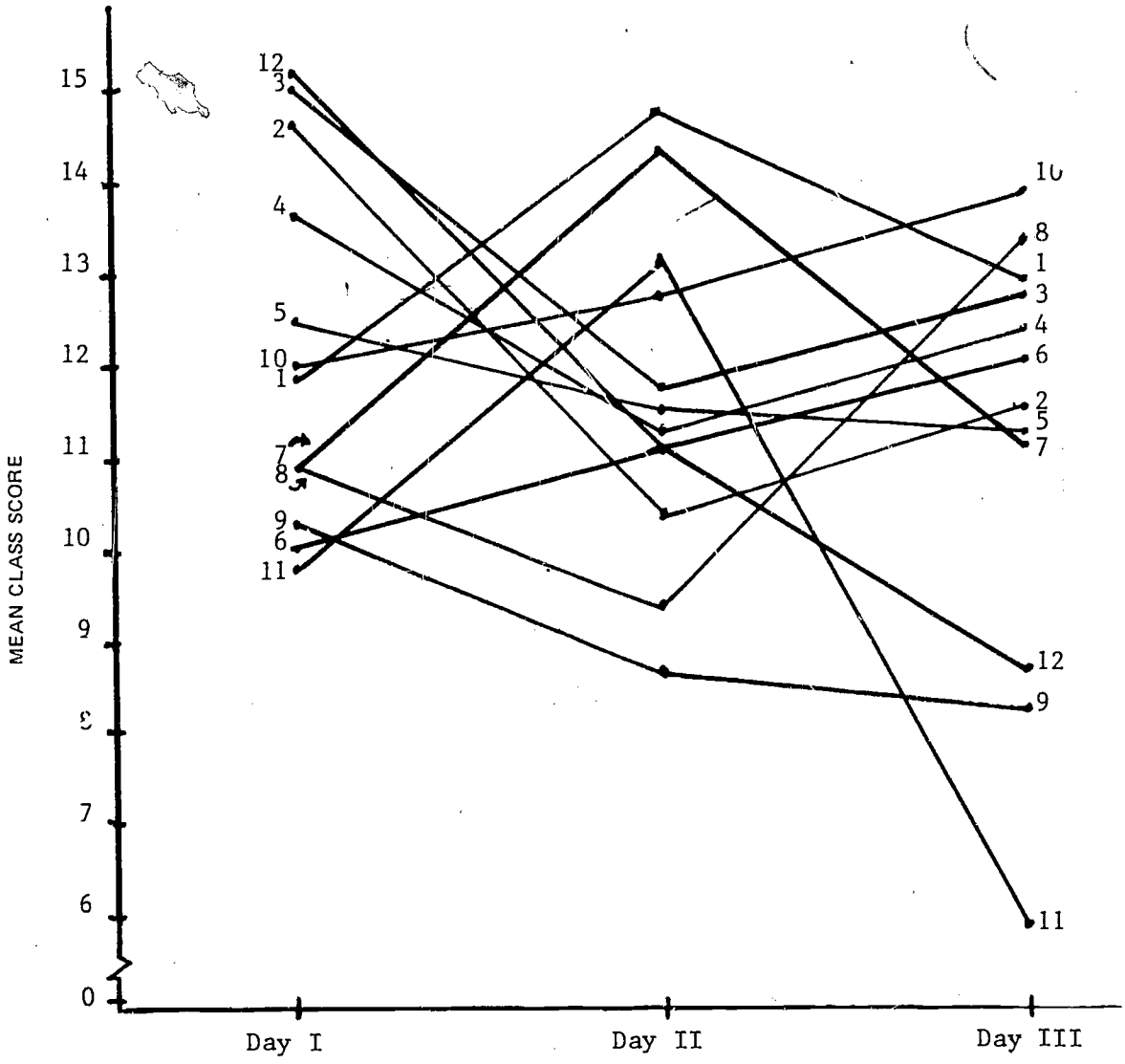

Fig. 9. Mean class scores on Essay Test Concrete for Days I to III. (Numbers identify individual teachers.)

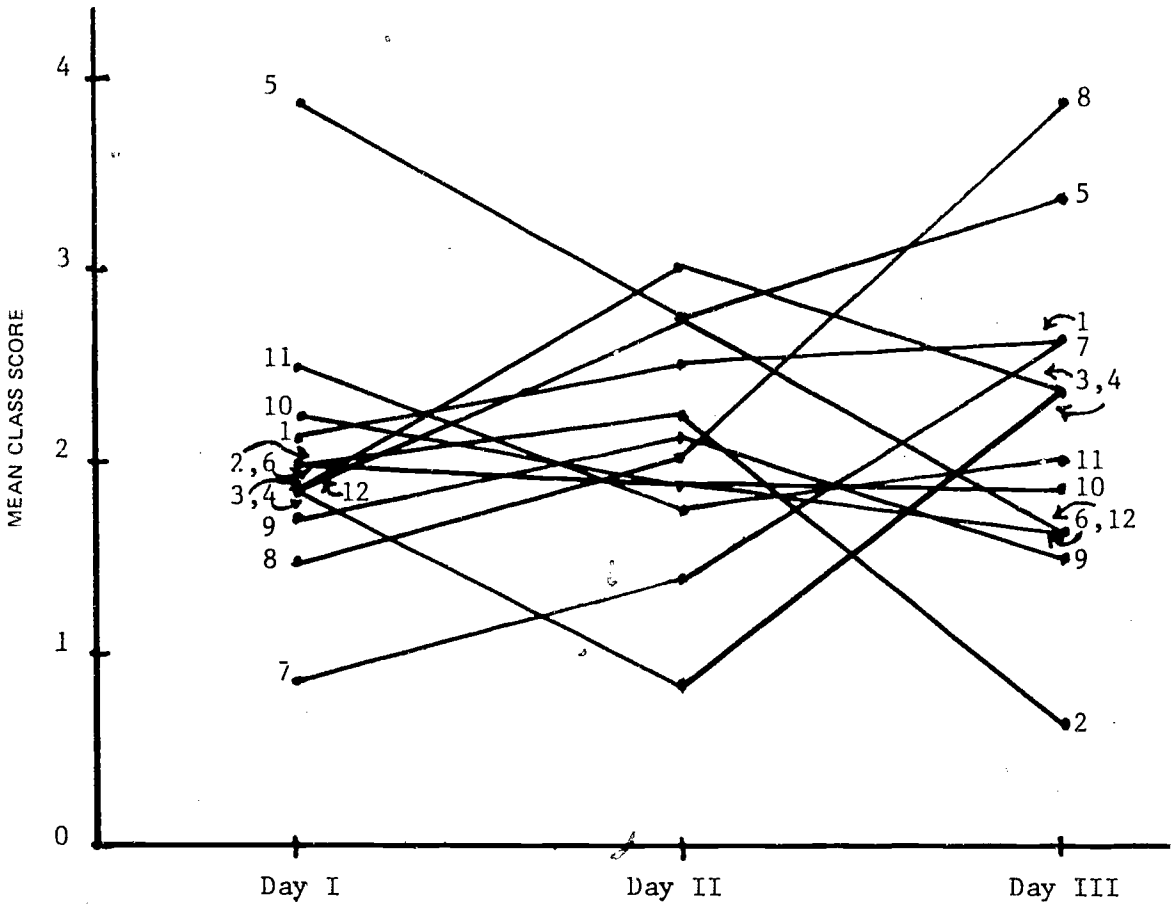

Fig. 10. Mean class scores on Essay Test Abstract for Days I to III. (Numbers identify individual teachers.)

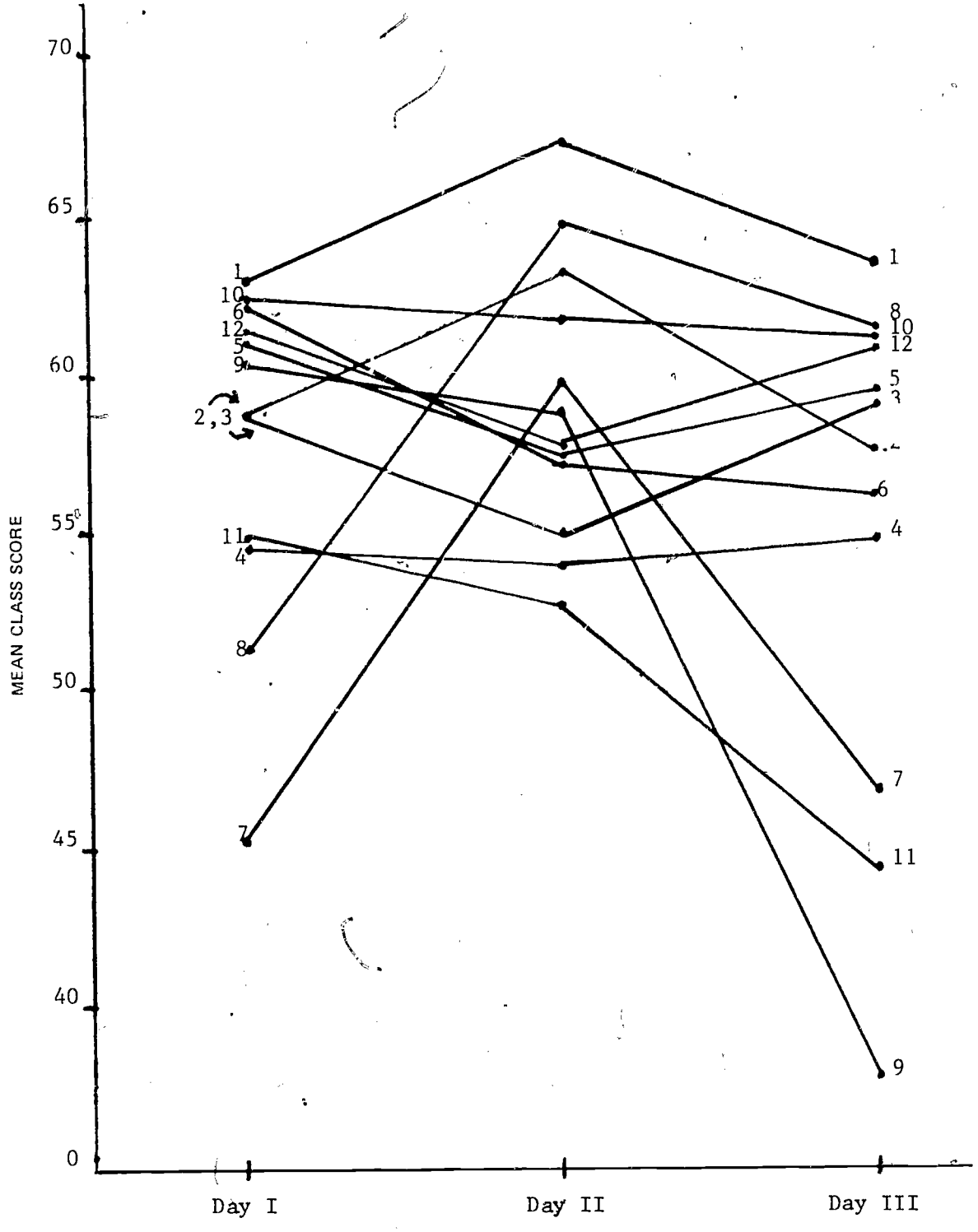

Fig. 11. Mean class scores on Attitude Toward the Situation tests for Days I to III. (Numbers identify individual teachers.)

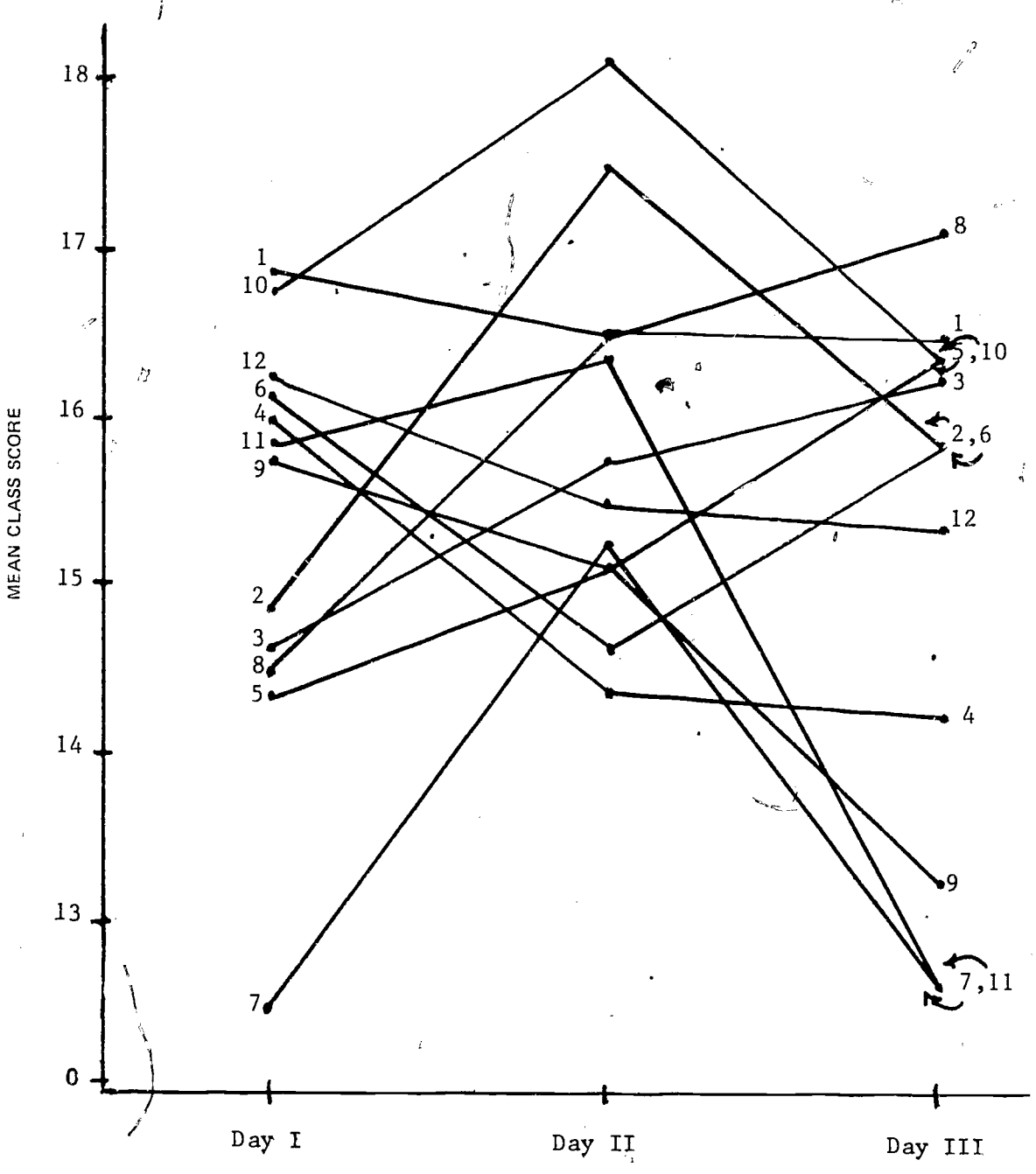

Fig. 12. Mean class scores on Attitude Toward Self tests for Days I to III. (Numbers identify individual teachers.)

Discussion and Conclusions

Why did relatively few teachers show marked increases in student learning with practice? We expected that as teachers' familiarity with the content to be taught, the teaching situation, and the reactions of students increased, their effectiveness would increase. Certainly the teachers did become more familiar with the content, teaching situation, and student reactions. But that was apparently not sufficient for improving teaching effectiveness.

One of the most striking effects in these results is the decrease (in the third study) from Day II to Day III of student adjusted Recall Test scores (11 of 12 teachers) and student Attitude Toward the Situation scores (8 of 12 teachers). A comparison of the teaching process variables for Days II and III indicated that, for the categories of teacher and student verbal behavior measured by an interaction analysis system (not described here), teaching on Day II was very similar to teaching on Day III. The drop in Day III student posttest scores was probably due to qualitative differences in teacher-student interaction not measured by the interaction analysis system. Possibly the teachers were bored with the task and the curriculum and communicated their lack of enthusiasm to the students. Teacher preparation and planning may also have been more perfunctory on Day III, leading to poorly organized teaching.

Another explanation for the results of these studies is related to the extent to which the Harlow learning-to-learn paradigm of the psychological laboratory can be effectively translated into designs for research on teaching. Learning-to-teach, as investigated in the research reported

here, may be very different from the learning-to-learn phenomenon investigated by Harlow. Harlow used discrimination tasks of a relatively low order in his investigations of learning in rhesus monkeys and children. In the present studies, the teachers were faced with a much more complex task, namely, combining teaching skills within a particular strategy in such a way as to teach a particular learner or learners most effectively. Teaching does involve a number of discrimination tasks: for example, discrimination of important learner characteristics, of important facts, principles, and concepts imbedded in the subject matter, and of teaching skills and strategies likely to be useful in teaching the content. But teaching is more than the sum of a series of discrimination tasks. Teaching is more akin to a higher level problem-solving task, in which the teacher must make a series of decisions about how to behave, given a great deal of complex information about subject matter, students, and the learning situation.

A second difference between Harlow's work and ours has to do with feedback. In Harlow's studies, the subjects were given immediate feedback after each very brief trial. In the learning-to-teach experiments, the teachers were not given any information about student achievement. We thought that the teachers would be receiving sufficient feedback from the interaction with their students to make judgments about the effectiveness of their teaching behavior. Further, we assumed that the teachers would be able to use the feedback from their interaction with the students to change their teaching behavior in ways that would improve their subsequent effectiveness. It is not clear from the data where this logic broke down in practice. It may be that explicit and accurate

feedback about student achievement and attitudes would have helped the teachers to improve their effectiveness. On the other hand, teachers might not know how to use even explicit and accurate feedback in ways that would improve their performance. Thus, we do not know whether the absence of learning to teach in our results was due to inadequate feedback to the teachers or to the teachers' inability to capitalize on the information available to them.

A third area of contrast between the present studies and Harlow's learning-to-learn experiments is in the extent to which successive trials for any given subject are comparable. In Harlow's work, each subject was interacting with a mechanism that operated on a fixed and predetermined set of rules. In the learning-to-teach experiments, the curriculum, learning objectives, and size of group were held constant in each trial. But in our experiments the teacher was interacting with different students in each trial--students who were not behaving in controlled or predetermined ways. It may be that experience with a particular student or group of students is not very useful in improving the effectiveness of interaction with a subsequent student or group of students. In other words, familiarity with the content to be taught, the teaching situation, etc., may contribute only a small part to the effectiveness with which a teacher interacts with a new group of students in a similar situation. This question could be investigated experimentally by having teachers teach the same students over a number of comparable curriculum units.

Our studies and Harlow's both used highly controlled laboratory settings. Important variables such as time, materials, and environmental factors were controlled in both sets of studies. Such highly structured

situations as we created in our experiments severely restricted the freedom of action of our teachers. Limited time and materials and prescribed objectives probably limited their freedom to experiment with different teaching approaches, thereby limiting their opportunity to learn to teach. Thus, the demand characteristics of the learning-to-teach experiments may, ironically, have prevented the phenomenon from occurring.

Finally, the learning-to-teach studies involved only a few trials (three or four) for each teacher, whereas Harlow's studies employed hundreds of trials for each subject. It may be that the hypothesized effects of teacher practice as measured by student achievement do not appear until after much more practice than our experimental designs allowed for. Or, it may be that student achievement as we measured it is not the "place" to look for the effects of teacher practice.

In the face of evidence to the contrary, it is still possible to retain some optimism that situation-specific practice can indeed help a teacher become more effective. The studies reported here represent early attempts at exploring and demonstrating the learning-to-teach phenomenon. We have learned at least three lessons from this series of studies: (1) our conceptualization of what is learned by teachers as they gain experience must be broadened and clarified; (2) our methodology for measuring the effects of such teacher learning on students must be improved; and (3) the usefulness of direct application of this paradigm from the psychological laboratory to designs for research on teaching must be reexamined.

Implications for Teacher Training

Practice, by itself, did not enable teachers to increase student achievement. This finding indicates that teachers might profit from a process that would enable them to observe more systematically the effects of their teaching on students--i.e., a training program that would help teachers become researchers on their own teaching effectiveness. Such a program would capitalize on the fact that every day or hour of teaching is an opportunity for a teacher to try new combinations of teaching skills and strategies, observe the effects, and adjust instructional performance to suit the particular students, situation, and subject matter being dealt with. Improvements in teaching effectiveness will be achieved only after teachers themselves learn to define and solve instructional problems in terms of the uniqueness of the complex teaching situations they face alone.

References

Clark, C. M. A model teacher training system: Questioning, explaining, and listening skills in tutoring. (Stanford Center for Research and Development in Teaching, R&D Memorandum No. 96) Stanford University, 1972. (ED 068 451)

Clark, C. M. The effects of teacher practice on student learning and attitudes in small group instruction. (Stanford Center for Research and Development in Teaching, Technical Report) Stanford University, forthcoming.

Harlow, H. F. The formation of learning sets. Psychological Review, 1949, 56, 51-65.

Harlow, H. F. Learning set and error factor theory. In S. Koch (Ed.), Psychology: A study of a science, Vol. II. New York: McGraw-Hill, 1959, pp. 492-537.

Medley, D . Measuring dimensions of teaching style: A study of the measurement properties of Observation Schedule and Record-Form 4V. Unpublished paper, University of Virginia, 1970.

Reese, H. W. Discrimination learning sets in children. In L. Lipsett and C. Spiker (Eds.), Advances in child development and behavior, Vol. 1. New York: Academic Press, 1964, pp. 116-145.