

## THE EFFECT OF A BRIEF TEACHER TRAINING ON STUDENT ACHIEVEMENT

PIETER C. VAN DER SIJDE

University of Twente, The Netherlands

**Abstract** — On the basis of previously determined correlations between teacher behavior and student achievement, a training program for mathematics teachers was developed. Previous research had established that this training did change the behavior of teachers. This article reports the effects of the training program on teacher behavior, student achievement, and student attitudes towards mathematics. From this study it appeared that the timing of the training was important for its effect on student achievement. There was no effect on student attitude.

Teaching is the central process in education, and its goal is to foster knowledge, desirable attitudes, and skills in students. A teacher can use different kinds of models and strategies (see, e.g., Joyce & Weil, 1972). Studying the approaches mentioned by Joyce and Weil makes it clear that most models and strategies are based on educational views and theories that have not been investigated for classroom use. On the assumption that students in a classroom react the same as animals and human subjects in a laboratory, some of these theories of learning are offered for the classroom without further research. What actually happens in a classroom and how learning and teaching in such a context take place are hardly known.

It is the merit of the approach that Rosenshine and Furst (1973) called the descriptive–correlation–experimental loop that what actually happens in a classroom becomes the starting point of further research. A shortcoming of this approach (or maybe also an advantage?) is that it can be theoryless. This means that one does not have to adhere to a specific theory to justify its use. Some writers reject this approach because of its apparent lack of a theoretical basis. Teaching is regarded as a complex of teaching practices, and by observing and cor-

relating teaching practices with student achievement, teaching becomes divisible into relatively specific practices. Training teachers in these practices (as in microteaching) is considered objectionable by some writers. This view is questionable, however, because such training (as part of research in the experimental part of the descriptive–correlation–experimental loop) does not reduce teaching to a mere set of discrete units of behavior. Nor does it give insight into specific cause-to-effect relationships of specific teaching practices. Teaching remains a complex skill, but, as a number of training studies have already indicated, it can be improved (Gage, 1985a, b).

The training programs, such as those developed by Van der Sijde (1987) and Good and Grouws (1979) need not be mere collections of the implications of process–product correlations. They can instead be more or less coherent instructional systems, which can be supported on the basis of evidence. One can regard the training program as a collection of discrete cause-and-effect relationships, but this is not what is researched; in the research the collection of discrete relationships is considered as a whole, as an instructional system. This study investigated the effect of such research-based

training, regarded as a whole (i.e., as an instructional system), on student achievement and attitudes.

### Method

#### *Subjects*

Thirty-three teachers volunteered to participate in the experiment. All of these teachers taught mathematics in the eighth grade using a textbook entitled *Sigma*. The schools the teachers worked at were randomly assigned to four conditions, so teachers teaching at the same school were assigned to the same condition. The teachers in Condition 1 ( $N = 13$ ) and Condition 2 ( $N = 8$ ) participated in the experiment from September to March, while the teachers in Condition 3 ( $N = 6$ ) and Condition 4 ( $N = 6$ , one of whom dropped out) participated only in the period from December to March.

#### *Teacher Training*

On the basis of the results of the Dutch Classroom Environment Study Correlation Phase and on the implications of other theory and research a teacher training program was developed (Van der Sijde, 1985, 1987). The content of the training was organized around three topics: the teaching script (lesson format), management of the classroom, and instruction. The content of the training was put into a 40-page manual entitled "Management en Instructie in Wiskundelessen" ("Management and Instruction in Mathematics Lessons"). A short elaboration of the content of the manual follows:

*The teaching script.* The leading idea behind the teacher training was the teaching script (Schank & Abelson, 1977; Abelson, 1982). This script consisted of five different scenes (lesson phases): homework check, presentation, monitored practice, guided practice, and homework/tutoring (Leinhardt & Greeno, 1986; Leinhardt & Smith, 1985).

Many combinations were possible, but in the training we proposed, after extended discussions with teachers and teacher trainers, there was one specific formulation of a teaching

script. Every lesson should start with a review of homework. This gives teachers the opportunity to check whether or not there are problems concerning the subject matter. This is necessary to know especially when conducting the presentation. When teachers know where the problems lie, they can take this into account. After the presentation they have to check whether or not their presentations were clear by giving the students individual assignments. After the individual assignments (monitored practice) teachers discuss more extended and difficult problems with the students (guided practice). The last phase of a lesson is giving the students homework and having them start it in the class, so that, in the meantime, teachers can tutor some students with (minor) problems. For each lesson phase or scene an indication of time to engage in each was given (see Table 1).

The empirical basis of this lessons schema was the correlation which was found between time-on-task and learning outcomes of students ( $r = .65$ ) for the users of the textbook *Sigma*. This correlation was calculated using a subset of the data of Tomic (1985). Executing this particular script seemed to enable teachers to maximize time-on-task.

*Classroom management.* It appeared from the correlational study, for the users of the textbook *Sigma*, that there were several interesting correlations between the management behavior of the teacher and learning outcomes (see Tomic, 1985). As it appeared from Tomic's study, the main trend in the correlations was a negative correlation between management behavior of the teacher and learning outcomes. All teacher management activities seemed to be of a *curative* nature, not of a *preventive* nature. In the model for classroom management we developed (Van der Sijde, 1989), we emphasized the preventive nature of classroom management. This prevention was recommended to the teachers starting from the cognitive interpretation of Kounin's (1970) theory.

Teachers should perform two simultaneous tasks, instruction and management. While instructing the class, teachers should constantly watch for activities which might interfere with the lesson (instruction). The simultaneous occurrence of instructional and management tasks is what we labelled "overlapping". When

teachers identify elements which might disturb the instruction they have to act. Through their acts they show they are "withit". To show withitness teachers can use the following techniques:

- look a student in the face or just mention his/her name;
- rebuke a student (N.B. it is important to rebuke the right student at the right moment);
- propose alternative behavior to a student (when a student does pay attention the teacher, for example, can call the student to the blackboard);
- let the student describe the desired behavior (when a student does something he or she is not supposed to do, the teacher asks the student what he or she is doing and what he or she is supposed to do).

It is important that teachers are focussed on the group and not on individual students. They have to take care that the attention of the group is with them. When this is not the case they can, for example, give turns, or give pay-attention signals to the group ("This is important", "This will be part of the test next week").

Further it is important that the students become aware of their own responsibility for what goes on in a lesson. It concerns their learning process. Teachers should stimulate this through appealing to student accountability. They can do this through:

- goal-directed questions ("Jim, how far are you?");
- checking of work ("Who has not finished assignment 5?");
- having students answer questions of other students.

In executing a lesson it is important that there is a continuous signal, which means a continuous stream of activities in the class; no abrupt changes and transitions, a more or less constant pace.

*Instruction.* Several recommendations for teachers concerning instruction were formulated in relation to the specific scenes in the teaching script.

1. Presentation. The presentation of subject matter should be preceded by a short indication of relevant concepts which are already familiar to the students (refresh prior knowledge), or a summary of relevant previously dealt with sub-

ject matter. When introducing new subject matter the teacher should use examples, explicit and concrete examples from everyday life. Further the teacher should use the textbook efficiently, for example, by indicating the relevant parts and especially the irrelevant parts. At the end of the presentation the teacher summarizes the main points.

2. Monitored practice. The presentation is followed by a short, clear, individual assignment which enables the teacher to check whether the presentation has been effective. This assignment has to be checked, and it is important that the teacher reacts positively to the students. During the guided practice the teacher walks through the class regularly checking students. After completion of the assignment the work of all students has to be checked (e.g., by asking for the solution, and checking who had different answers). The teacher has to react positively and be task-oriented.

3. Guided practice. After the individual assignment the teacher discusses a more complex problem with the class. The solution is not given by one individual student, but it should be the result of a group process. The teacher gives probes, uses redirections of questions, etcetera.

4. Homework assignments/tutoring. The final phase of a lesson is the homework assignment. Students must be able to start on these in the class and finish them at home. While most of the students work on their homework assignments, the teacher has time to help students individually or in small groups with particular problems identified in earlier phases of the lesson.

5. Homework check. Although last in this enumeration, this homework checking is the first part of a lesson. It is important to check homework and a teacher has to do this. All homework assignments must be checked in one way or other. This means that the amount of homework assignments must be limited to what can be checked in the next lesson.

#### *Observation Instrument: TOOL-W*

Nine graduate students (in Education and Science) were trained for about 40 hours to master the observation instrument TOOL-W implemented within the TOOL-system (Van der Sijde & Dirksen, 1987), which is a series of con-

Table 1  
*A Summary of Instructional and Managerial Rules Included in the Training*

---

<i>Lesson-phase model</i>	Phase 1. Review of homework	(5 to 10 min)
	Phase 2. Presentation	(10 to 15 min)
	Phase 3. Monitored practice	(about 5 min)
	Phase 4. Guided practice	(about 10 min)
	Phase 5. Homework/tutoring	(about 10 min)

*Classroom management*

- Use a beginning-of-the-lesson signal
- Make transitions between two lesson phases as smooth as possible
- Make sure that there is a continuous stream of activities in the classroom
- Be withit
- Maintain group focus
- Let students be responsible for their work

*Instruction*

- Discuss all homework assignments
  - Start presentation with a review of relevant concepts
  - Refresh foreknowledge by means of questioning
  - Use concrete examples during presentation
  - Indicate (un)important passages in the textbook
  - Indicate what is important (oral/blackboard)
  - Give clear assignments for monitored practice
  - React positively to students
  - Check the assignments during monitored practice
  - Redirect questions and use probes
  - At the end of the lesson give time to students to work on their homework assignments
  - Tutor at the end of the lesson students who encountered problems
- 

nected programs used on a portable microcomputer (CANON X-07) for the specification and use of observation instruments. The TOOL-W observation instrument was developed for use in observing behavior in the classroom. It contained 33 low-inference categories: 4 codes for the direction-of-gaze of the teacher, 6 codes for the lesson phase, 8 codes for classroom management, and 15 codes for instruction. Every code was a low-inference operationalization of a recommendation. The structure of TOOL-W is depicted in Figure 1.

*Training of Observers*

The training of observers started with an explanation and discussion of the observation codes in the TOOL-W instrument. Further training consisted of practical exercises using videotaped and audiotaped situations. The final part of the training consisted of making three observations in classroom situations. During the training no formal inter-observer agreement was calculated, but a hard copy of the data was made and the results of the observation were discussed by the observers and the author. At

the end of the training the agreement between the observers was judged to be satisfactory.

*Attitude Test*

The attitude test was composed of items of the student questionnaire developed for the Second Mathematics Study of the International Association for the Evaluation of Educational Achievement (IEA). The original questionnaire (Pelgrum, Eggen, & Plomp, 1983) contained 84 items on six subjects: (a) mathematics at school, (b) mathematics as process, (c) mathematics and I, (d) mathematics and society, (e) male-female and mathematics, and (f) computer and calculators. The data were reanalyzed, using only the data for the eighth grade, in order to construct an attitude test containing only the first three of the above subjects. It appeared from factor analyses on the data of 1500 students that the items of the subjects "mathematics at school", "mathematics as process", and "mathematics and I" loaded on two factors, which could be labelled as "enjoying mathematics" and "self-image". By means of item analysis two scales were constructed: en-

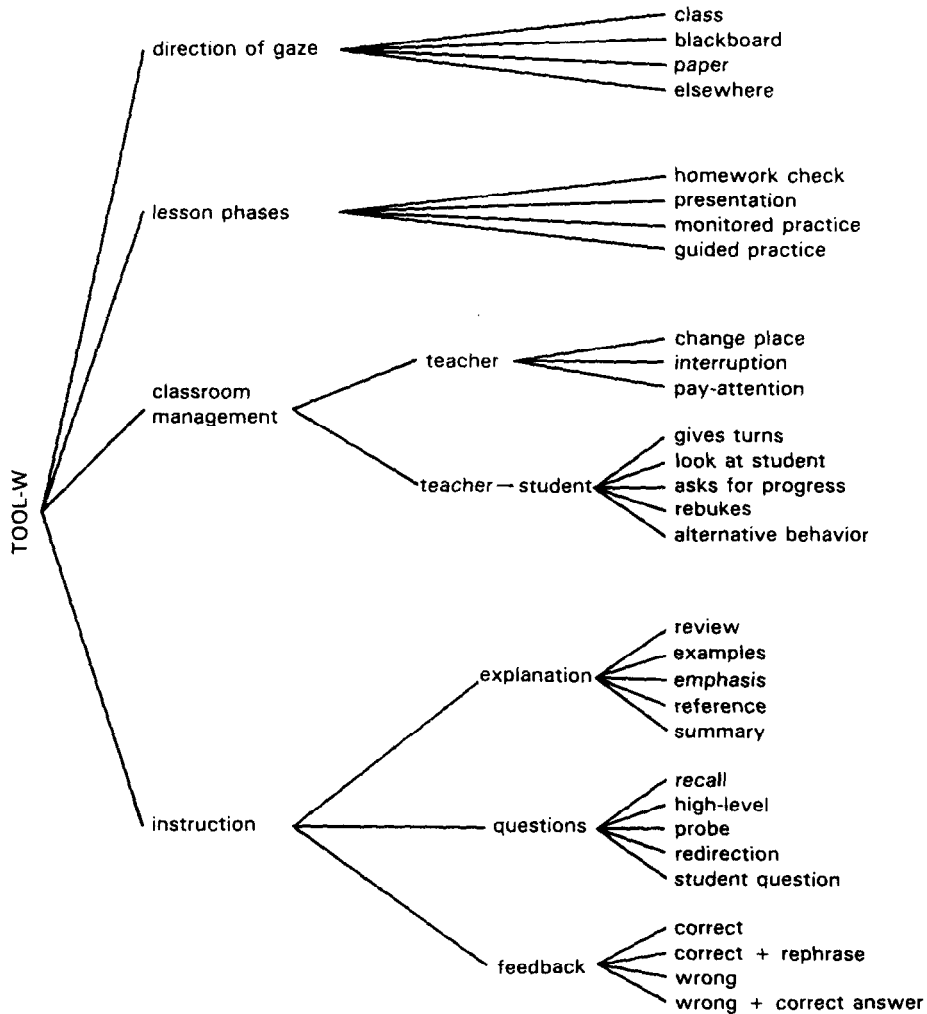


Figure 1. Structure of the observation instrument TOOL-W.

joying mathematics (number of items was 13, Cronbach's alpha was .85, and examples of items are: "I like mathematics", "I want more mathematics") and self-image (number of items was 5, Cronbach's alpha was .85, and examples of items are: "I'll never be good at mathematics", "I have more trouble with mathematics than others"). Each item had to be rated on a five-point-scale, and for each of the two sub-scales a score was calculated.

#### *Achievement Test*

Each achievement test contained 30 items.

All of these items were constructed by the Dutch Central Institute for Test Development (CITO). The pretest (with items on seventh-grade mathematics) was the same for all teachers and their classes. The teachers were given the test to check for opportunity-to-learn. If the test contained items on topics not dealt with then these were excluded from the analysis. The posttests differed among teachers, because every posttest was intended to correspond with the content covered by the teachers in the given time-period. Theoretically this could mean that there would be 33 different tests. Practically it meant that there were 5 "dif-

ferent" tests having 20–25 items in common (see also "design and data").

### Procedure

The experiment contained four conditions:

*Condition 1.* The teachers in this condition were enrolled in our institute and participated in a one-day training course. After the training an achievement test and attitude test were administered to the students of those teachers (pretest). Then 8 to 10 lessons of those teachers during a 2.5-month period from September to December were systematically observed by trained observers using the TOOL-W instrument. At the end of the 2.5-month period the same attitude test and a different achievement test were administered (first posttest) and after another 2.5 months from January to March, the same attitude test was administered for the third time, together with a third achievement test (second posttest).

*Condition 2.* This condition resembled Condition 1 except for the fact that the teachers received the training not before the observation period (September to December) but in January after the observations of the lessons.

*Condition 3.* The teachers in this condition received the training manual in January, and in their classes only the first and second posttest, each con-

sisting of the attitude test and the achievement test, were administered in December and March–April, respectively.

*Condition 4.* In the classes of the teachers in this condition, the first and second posttest were administered in December and March–April, and the teachers did not receive the training manual.

### Design and Data

The design used in this study is depicted in Table 2. In order to study the implementation of the lesson-phase model set forth in the training manual, three variables were constructed. The first is the  $L$ -score, where  $L = \{5 - \text{OLP}\}$  (OLP is the number of observed lesson phases); in the "ideal" lesson  $L = 0$ . The second variable is  $F$ , where  $F = 0$  when a lesson starts with phase one, and otherwise  $F = 1$ . For every lesson an  $S$ -score was calculated; the  $S$ -scores express the violations of the lesson-phase model. An ideal lesson (according to the lesson-phase model) starts with phase one, followed by phases two, three, four, and five. Violations of the model would occur, for example, when phase one is followed by phase four (forward violation) or

Table 2  
*Design of the Study*

Treatment condition	Time of year				
	September	September to December	December	January	March/April
Condition 1	Training Manual Achievement test Attitude test	Observation of 8–10 lessons	Achievement test Attitude test	No training No manual	Achievement test Attitude test
Condition 2	No training No manual Achievement test Attitude test	Observation of 8–10 lessons	Achievement test Attitude test	Training Manual	Achievement test Attitude test
Condition 3	No training No manual No achievement test No attitude test	No observation	Achievement test Attitude test	No training Manual	Achievement test Attitude test
Condition 4	No training No manual No achievement test No attitude test	No observation	Achievement test Attitude test	No training No manual	Achievement test Attitude test

phase three is followed by phase one (backward violation). The ideal lesson has an *S*-score of 0. Some violations were considered minor (going from phase four to phase three, or from phase three to phase five: a 1 point violation), while others are considered major (going from phase five to phase two: a 2 points violation). *S* was the sum of the violation points and was used to establish the degree of implementation of the lesson-phase model. Data on these variables were available only for teachers in Conditions 1 and 2.

Furthermore, in TOOL-W a number of Kounin's (1970) variables were operationalized. In the section on results, only those for several Kounin variables are reported: *Withitness* was operationalized in terms of low-inference variables: (a) The teacher watches the student (direction of gaze is towards the student); (b) the teacher changes place (takes another position in the classroom); (c) the teacher looks at a particular student to indicate he or she has seen the student's inappropriate behavior; (d) the teacher rebukes a student; and (e) the teacher suggests alternative behavior to a student who shows appropriate behavior. The number of times the teacher performed each of these behaviors was observed. *Withitness* is the standardized sum of the five mentioned low-inference behaviors. *Group alerting* was operationalized by determining the number of times the teacher alerted the class, for example with "This is important" or "Pay attention". *Learner accountability* was operationalized by determining the number of times a teacher asked for progress made by the pupils (e.g., "Who has completed . . .?", "Who is not ready by now?").

The difference as the result of the training on management and instruction behavior is expressed as the effect size, which is defined as  $E = ((X_A - X_B) / SD_B)$ , where  $X_A$  stands for the mean score for the teachers in Condition 1, and  $X_B$  is the standard deviation of the scores in Condition 2.

In order to compare the increases in scoring ( $\Delta$ (posttest 1–pretest);  $\Delta$ (posttest 2–pretest);  $\Delta$ (posttest 2–posttest 1)) on the tests over a period of time, the mean class score of each teacher was converted to a standard score using

the overall mean of all teachers in the Conditions to be compared and its standard deviation:  $((\text{overall mean}) - (\text{mean class score individual teacher})) / (\text{standard deviation (overall mean)})$ .

## Results

### *Inter-observer Agreement*

To establish the agreement between the observers, a number of lessons ( $N = 15$ ) were observed by two observers at the same time. Using Cohen's kappa (Cohen, 1960) we calculated the inter-observer agreement.

The kappa for the direction of gaze was: 0.48; for the lesson phases: 0.98; for classroom management: 0.84; and, for instruction: 0.59.

*Effects of the Training on Teacher Behavior Lesson-phase model.* The teachers in Condition 1 spent significantly more time on monitored practice ( $t_{19} = 1.60, p < .10$ ) and on guided practice ( $t_{19} = 1.30, p < .10$ ) (see Table 3) than the teachers in Condition 2.

The training emphasized the importance of including each of the five lesson phases in every lesson. But not every observed lesson included these five phases. Nonetheless there was a significant difference ( $p < .001$ ) between the teachers in Condition 1 and Condition 2 with respect to the *L*-variable (see Table 4).

Further, according to the training program, every lesson should start with a review of homework. With respect to the *F*-variable there was no difference between these two conditions (Chi square = 0.03,  $p > .10$ ). With respect to the *S*-variable, there was a difference between the conditions ( $t_{169} = -1.49, p < .10$ ). The mean score for teachers in Condition 1 was 3.12, with a standard deviation of 3.40; for teachers in Condition 2 the mean score was 4.00, with a standard deviation of 4.69.

*Teacher behavior: Management.*<sup>1</sup> The mean withitness score of the teachers in Condition 1 was 0.26 (standard deviation 2.85), and of the teachers in Condition 2 was -0.42 (standard deviation 2.00). The difference for withitness was not statistically significant (Mann-Whitney  $U =$

<sup>1</sup>The mean scores are based on a mean score per teacher.

Table 3  
*Time Engaged in the Five Phases of the Lesson Model for the Teachers in Conditions 1 and 2*

Lesson phases	Number of Minutes Actually Spent*		
	Recommended number of minutes per lesson	Condition 1	Condition 2
Review of homework	5–10	16.50	19.40
Presentation	10–15	9.76	10.28
Monitored practice	5	8.90	5.10**
Guided practice	10	3.50	1.76**
Homework/tutoring	10	5.20	6.33

\*The time is generalized to a complete lesson of 50 minutes (only 25 minutes were systematically observed by observers).

\*\* $p < .10$ .

Table 4  
*The Percentage of Lessons with the Number of Lesson Phases Present in Conditions 1 ( $n = 103$ ) and 2 ( $n = 68$ )*

Condition	Number of Lesson Phases Present					
	5	4	3	2	1	0
Condition 1	10.3%	33.3%	32.5%	18.8%	4.3%	0.9%
Condition 2	2.7%	28.4%	32.4%	29.7%	6.8%	0.0%

Chi-square = 27.16,  $df = 5$ ,  $p < .001$ .

46,  $p > .10$ ); the effect size was 0.34. The mean group-alerting score of the teachers in Condition 1 was  $-0.05$  (standard deviation 1.02), and of the teachers in Condition 2 was 0.31 (standard deviation 0.98). The difference was not significant (Mann–Whitney  $U = 71$ ,  $p > .10$ ); the effect size was 0.37. The mean score for learner accountability of the teachers in Condition 1 was 0.30 (standard deviation 1.08), and of the teachers in Condition 2 was  $-0.037$  (standard deviation 0.73). The difference between the conditions was significant (Mann–Whitney  $U$ -test value 26.5,  $p < .10$ ); the effect size was 0.92.

*Teacher behavior: Instruction.*<sup>2</sup> Instruction is divided into three components: explanation (presentation of subject matter), questioning, and feedback. The means, standard deviations, and effect sizes presented in Table 5 show that the largest effect sizes were for giving a review (at the start of presentation) ( $E = 1.00$ ), for use of the text book ( $E = 0.63$ ), and for providing

positive feedback ( $E = 0.60$ ), while a negative effect size ( $E = -0.43$ ) was obtained for giving a summary (at the end of presentation), all of these subjects were statistically significant at the .01 level.

#### *Effects of the Training on Student Achievement and Attitudes*

The experiment involved four groups of teachers and their classes. The results on the *pretests* of achievement and attitudes indicate that the results of teachers in Condition 1 are comparable with the results of the teachers in Condition 2 (see Table 6). For attitude scores on enjoying mathematics, the Mann–Whitney  $U$  was 50 ( $p > .10$ ); for attitude scores on self-image the Mann–Whitney  $U$  also was 50 ( $p > .10$ ); for scores on the achievement test the Mann–Whitney  $U$  was 47 ( $p > .10$ ).

When the results on the first posttests by classes in the two experimental conditions (Condi-

<sup>2</sup>The unit of analysis was the lesson. The presented mean scores refer to the mean number of intervals of 5 seconds per 5 minutes, per lesson.



Table 5

*The Mean Number of Intervals of 5 Seconds per 5 Minutes per Lesson and Standard Deviations Engaged in a Particular Instructional Behavior of the Teachers in Experimental Conditions 1 and 2 as well as the Effect size of the Training on the Instructional Behavior and the t-value of the Differences*

	Condition 1		Condition 2		Effect size	t-value
	M	SD	M	SD		
<i>Explanation</i>						
Review	0.55	0.50	0.25	0.30	1.00	4.42*
Questions on Prerequisites	2.75	1.95	2.55	1.60	0.13	0.70
Use of the textbook	0.55	0.45	0.30	0.40	0.63	3.69*
Indicate importance	0.65	0.30	0.60	0.55	0.09	0.76
Use of examples	0.25	0.25	0.20	0.25	0.20	1.27
Summary	0.30	0.35	0.60	0.70	-0.43	-3.68*
<i>Questions</i>						
Lower order questions	10.30	4.30	9.65	4.20	0.15	0.97
Higher order questions	0.55	1.00	0.50	1.10	0.05	0.31
<i>Feedback</i>						
Positive feedback	1.04	0.65	0.76	0.47	0.60	3.04*
Negative feedback	0.40	0.25	0.49	0.39	0.23	-1.83**

\*  $p < .01$ . \*\*  $p < .05$ .

tions 1 and 2) were compared with those classes in Conditions 3 and 4 no significant differences were found (attitude scores for enjoying mathematics: chi square = 0.65 ( $df = 2$ ); attitude score for self-image: chi square = 3.36 ( $df = 2$ ); scores on achievement: chi square = 0.45 ( $df = 2$ ). The differences were assessed using the Kruskal-Wallis test.

There were also no significant differences on the first posttest between the students of the teachers in Condition 3 and Condition 4 on attitude scores for enjoying mathematics (Mann-Whitney  $U = 13$ ,  $p > .10$ ), for self-image (Mann-Whitney  $U = 16$ ,  $p > .10$ ), and for achievement (Mann-Whitney  $U = 13$ ,  $p > .10$ ). The differences among the four conditions on the results on the second posttest were also non-significant (attitude score for enjoying mathematics: chi square = 1.48 ( $df = 2$ ); for self-image; chi square = 1.49 ( $df = 2$ ); and for achievement: chi square = 2.81 ( $df = 2$ ). These differences also were assessed using the Kruskal-Wallis test.

To compare the increase in score from the pretest to the first posttest, both scores were standardized on the overall mean of the groups concerned as described in "Design and Data" above. The differences in increase scores between Conditions 1 and 2 were not significant for the attitude scores (enjoying mathematics: Mann-Whitney  $U = 39.5$ ; self-image: Mann-

Whitney  $U = 46.5$ ), or for the achievement scores (Mann-Whitney  $U = 40$ ).

The increases in scores from the first posttest to the second posttest did not significantly differ for the teachers in the treatment conditions (Conditions 1 and 2 combined) and the teachers in the control conditions (Conditions 3 and 4 combined): for enjoying mathematics: Mann-Whitney  $U = 109.5$ , for self-image: Mann-Whitney  $U = 89$ , for achievement scores: Mann-Whitney  $U = 111$ . Nor was there any significant increase in scores among the four conditions: for enjoying mathematics: chi square = 1.48 ( $df = 3$ ); for self-image: chi square = 2.23 ( $df = 3$ ); for achievement scores: chi square = 0.28 ( $df = 3$ ). All were assessed using the Kruskal-Wallis test. When the increase in scores on the attitude test from the pretest to the second posttest were compared, there was no significant difference (for enjoying mathematics: Mann-Whitney  $U = 40$ ; for self-image: Mann-Whitney  $U = 40$ ). On the achievement test, however, there was a significant increase (Mann-Whitney  $U = 27$ ,  $p < .10$ ) in favour of Condition 2.

## Discussion

It appears, from the results of the inter-ob-

Table 6  
*Mean Class Scores and Standard Deviations on the Achievement and Attitude Test Administered during the Schoolyear*

	Time of year					
	September (pretest)		December (posttest 1)		April (posttest 2)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Achievement</i>						
Condition 1	.622	0.07	.524	0.16	.466	0.08
Condition 2	.609	0.08	.595	0.19	.519	0.16
Condition 3	—	—	.493	0.17	.420	0.11
Condition 4	—	—	.609	0.15	.532	0.14
<i>Attitude: Enjoying mathematics</i>						
Condition 1	21.36	1.8	21.05	2.0	20.93	1.8
Condition 2	21.43	1.3	21.60	1.1	21.67	1.2
Condition 3	—	—	21.21	1.8	20.20	1.8
Condition 4	—	—	21.97	1.3	21.82	1.4
<i>Attitude: Self-image</i>						
Condition 1	17.93	1.7	18.05	1.7	18.85	2.4
Condition 2	18.04	0.9	18.34	0.8	19.04	1.5
Condition 3	—	—	18.21	1.7	17.78	2.1
Condition 4	—	—	18.75	1.3	18.77	1.0

server comparisons, that there was a reasonable level of agreement between the observers. The test results for the students in Conditions 1 and 2 were similar for attitudes as well as for achievement in that they did not differ significantly. This similarity held for the pretests as well as for the first and second posttests. There were also no significant differences among all four conditions for each of the two posttests. From these results, it can be deduced that the observations of lessons of teachers in Condition 1 and 2 did not affect the posttest scores; in this sense there was no evidence of a Hawthorne effect.

To assess the effectiveness of the training, not only the results on the posttests are important. Rather, the increase in scores from pretest to posttest(s) should also be considered. In comparing the differences in increase in scores for the teachers in Condition 1 with those in Condition 2, we found that there was no significant difference in increase scores from the pretest to the first posttest (i.e., in the period September–December). This is a remarkable result, because such a result has not been reported so far in other training studies. As in all other training studies, the first part of the experiment can be regarded as a nonequivalent control group design and although no effect of the training is evident in an increase in the achievement

and attitude scores, the training did have an effect on the teaching behavior. As it appears from the results of the classroom observations, the mean effect size of training (calculated as the combined mean for all behaviors) was approximately 0.33. Further, the teachers in the first condition implemented the teaching script (the lesson phase model), a finding which appears in their lower mean score on the *S*- and *L*-variables. There was no significant difference on the *F*-variable, indicating that most teachers (irrespective of the experimental conditions) started their lessons with a check of homework. The teachers in the first condition spent more than the recommended time on monitored practice (from Table 3 it appears that this monitored practice went on for approximately 9 minutes per lesson), while the teachers in the second condition spent almost exactly the recommended time on this phase (about 5 minutes per lesson). A significant difference was also found for guided practice (Condition 1: about 3.5 minutes, and Condition 2: about 2 minutes), with neither condition approaching the recommended 10 minutes. It appeared from personal communications with the participating teachers that in practice the discrimination between guided and monitored practice, although in theory rather clear, is hard to make. If these two

forms of practice are considered as one, the teachers in the first condition spent approximately 12.5 minutes per lesson, and the teachers in the second condition spent approximately 7 minutes, while the lesson phase model recommended approximately 15 minutes.

The teachers in experimental Condition 1 behaved according to the recommendations in the training more than the teachers in experimental Condition 2 (they were unaware of the content of the training). Although the teachers in the first experimental condition behaved more according to the recommendations in the training, the changes in their actual teaching behavior did not have an effect on increases in scores on the tests on the period September–December. Possibly a period of approximately 2.5 months is too short. The teachers in the first experimental condition had to find out for themselves the degree to which they already behaved according to the new teaching script. The difference in the increase in scores on the achievement test in the September–April period was, however, significant but in favor of the second experimental condition. No effect was found on the increase in scores on the attitude tests.

These results indicate that the most effective condition was the one with the training after the observation period. There is a possible explanation for this effect. The teachers in the first condition received their training before their classroom observation data were available. The teachers in the second condition, however, had all their observation data available and were confronted with the results of the classroom observations during their training. In other words, concrete indications for change based on how they actually behaved in class and what the training recommended, could only be given to the teachers in the second condition. The teachers in Condition 1 got indications for change based only on how the “average” teacher behaved in class. The implementation process of the teachers in Condition 2 could, for that reason, be more goal directed.

Probably the most effective design is the design in which training is given on two occasions: a first training before a period of observations, and a second training after a number of observations of lessons in which the same training, but tailored, is given to the teachers after two or three months. A number of training studies

have had a second training session, but not in the above sense. In the training studies by Anderson, Evertson, and Brophy (1979), and Good and Grouws (1979) the period of time between the first and second training session was only one week.

When the teachers in the third and fourth conditions were compared on the increase in scores on the tests in the December–April period, there were no significant differences. This means that sending the training manual by mail did not affect student achievement or attitudes. The same results were obtained by Coladarci and Gage (1984). The design used by them resembles the design used in this study (for the period January–April) except for the fact that the length of time between the tests in their study was about one school year, and in the present study it was 2.5 months. In the study performed by Coladarci and Gage no significant differences were found either. There is another difference between the present study and that of Coladarci and Gage (1984). In the latter study the teachers were asked to participate in the study and upon their agreement teachers in the experimental group were sent the manual(s) by mail. In this study the teachers in the third condition received the manual and were asked to read and practice it. There was no check at all on reading the manual. This is called a minimum intervention by Coladarci and Gage. Even when a less minimal intervention is added, using a design with three conditions, one with face-to-face training (the second condition), one with only a training manual (the third condition), and a control condition with neither a manual nor meetings (the fourth condition), there is no significant difference. Assuming that the content of the training was appropriate, it might be that the minimum intervention in this study was too small with respect to the degree of participation and time. A period of 2.5 months (also including holidays) may be too short to have any observable influence on the students’ learning outcomes; however, it may not be too short for noticeable changes in teaching behavior to occur.

This study, like others, demonstrates that even a short teacher training course can successfully effect change in the teachers’ teaching behavior, and student achievement. But, unlike in other studies, it remains unclear how im-

plemented changes in teaching behavior affect student achievement.

## References

- Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, **36**, 715-729.
- Anderson, L. M., Everson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal*, **79**, 193-223.
- Borg, W. (1975). *Instructional handbook for the protocol modules on classroom management*. Logan: Utah University.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37-46.
- Coladarc, T., & Gage, N. L. (1984). Effects of a minimal intervention on teacher behavior and student achievement. *American Educational Research Journal*, **21**, 539-555.
- Gage, N. L. (1985a). *Hard gains in the soft sciences: The case of pedagogy*. Bloomington: Phi Delta Kappa.
- Gage, N. L. (1985b). *What educators should know about research on teaching*. Paper prepared for the Doctor Residency Seminar, College of Education, University of Florida, Tampa.
- Gage, N. L., & Giaconia, R. (1981). Teaching practices and student achievement: Causal connections. *New York University Education Quarterly*, **12**, 2-9.
- Good, T. L., Biddle, B. J., & Brophy, J. E. (1975). *Teachers make a difference*. New York: Holt, Rinehart & Winston.
- Good, T. L., & Grouws, D. A. (1979). The Missouri mathematics effectiveness project: An experimental study in fourth grade classrooms. *Journal of Educational Psychology*, **71**, 355-362.
- Joyce, B., & Weil, M. (1972). *Models of teaching*. London: Prentice-Hall.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.
- Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, **78**, 75-95.
- Leinhardt, G., & Smith, D. (1985). Expertise in mathematics instruction: subject matter knowledge. *Journal of Educational Psychology*, **77**, 247-271.
- Pelgrum, W. J., Eggen, T. J., & Plomp, T. (1983). *Tweede wiskunde project: Analyses van inkomsten leerstofaanbod en resultaten* [Second mathematics study: Analyses of results — subject matter and results]. Enschede: University of Twente.
- Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research of teaching* (p. 00-00) Chicago: Rand McNally.
- Schank R., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale: Lawrence Erlbaum Associates.
- Tomic, W. (1985). *Docentgedrag en leerresultaten* [Teacher behavior and learning outcomes]. Enschede: Febodruk.
- Van der Sijde, P. C. (1985). *The implementation of a research-based teacher training*. Enschede: University of Twente.
- Van der Sijde, P. C. (1987). *Training the teaching script*. Almere: Versluys.
- Van der Sijde, P. C., & Dirksen, A. (1987). The TOOL-system. In Maarsse, F. J. et al. (Eds.), *Computers in psychology: Methods, instrumentation and psychodiagnostics* (pp. 129-134). Lisse: Swets & Zeitlinger.

Received 19 May 1989 □