DOCUMENT RESUME

ED 251 929 EA 017 272

AUTHOR	Gall, Meredith D.; And Others
TITLE	Involving the Principal in Teachers' Staff Development: Effects on the Quality of Mathematics Instruction in Elementary Schools. Final Report, FY 1983.
INSTITUTION	Oregon Univ., Eugene. Center for Educational Policy and Management.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Sep 84
NOTE	220p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	*Administrator Role; Educational Improvement; Educational Research; Elementary Education; Inservice Teacher Education; Instructional Innovation; *Leadership; Mathematics Instruction; Principals; *Teacher Effectiveness
IDENTIFIERS	Active Mathematics Teaching (Good and Grouws)

ABSTRACT

A study to determine whether the instructional leadership shown by principals is a causal factor in the effectiveness of schools demonstrated that involving principals in staff development programs can increase implementation of the programs' objectives. Fifty-three fourth- and fifth-grade teachers were assigned to three treatment conditions: one in which principals participated actively in a staff development program for training teachers in an instructional model for mathematics; a second in which principals did not participate; and a third, the control group, in which neither teachers nor principals participated in the program. The teachers' math lessons were observed immediately before and after the staff development program, and again three months later. Their students were administered curriculum-referenced and nationally standardized mathematics tests before and three months after the program. Students of teachers in both trained groups made slightly greater gains on the tests than did those of control group teachers. Teachers in the involved-principals group showed a higher level of implementation in the delayed lesson, and their students showed greater gains on the curriculum-referenced test than did those of trained teachers with uninvolved principals. Students in the uninvolved-principal group outgained those in the involved group on one subtest of the standarized measure. (PGD)

INVOLVING THE PRINCIPAL IN TEACHERS' STAFF DEVELOPMENT: EFFECTS ON THE QUALITY OF MATHEMATICS INSTRUCTION IN ELEMENTARY SCHOOLS

by
Meredith D. Gall, Glen Fielding,
Del Schalock, W.W. Charters, Jr.,
and Jerzy M. Wilczynski

INVOLVING THE PRINCIPAL IN TEACHERS' STAFF DEVELOPMENT: EFFECTS ON THE QUALITY OF MATHEMATICS INSTRUCTION IN ELEMENTARY SCHOOLS

by
Meredith D. Gall, Glen Fielding,
Del Schalock, W.W. Charters, Jr.,
and Jerzy M. Wilczynski

September 1984

Center for Educational Policy and Management
College of Education
University of Oregon
Eugene, Oregon 97403
(503) 686-5173

The preparation of this final report for FY 83 was made possible through an Institutional Grant awarded by the National Institute of Education to the Center for Educational Policy and Management. The opinions expressed in this report do not necessarily reflect the positions or policies of the NIE or the U.S. Department of Education.

ABSTRACT

Recent research suggests that principals of effective schools exhibit instructional leadership behavior such as participating in teachers' staff development programs and observing and giving feedback on teachers' instruction. Because that research employed descriptive and correlational research designs, however, it is unclear whether improved school effectiveness is *caused* by the principal's instructional leadership. The present experiment was conducted to determine whether systematically manipulating principals' instructional leadership behavior would lead to positive changes in teachers' instruction and students' achievement.

Fifty-three fourth- and fifth-grade teachers were assigned to three treatment conditions. *Active Mathematics Teaching*, a staff development program developed by Good and Grouws, was used as the intervention. This program trains teachers in an instructional model that involves assigning homework, conducting continual review, emphasizing whole group development instruction, and monitoring and checking seatwork.

Teachers in the *principal involvement condition* participated in this program along with their principals. In addition, the principals received instructional leadership training and conducted observations and conferences with their teachers. Teachers in the *regular inservice condition* participated in the program, but their principals were not involved in any way. In the *control condition*, neither the teachers nor their principals participated in the program.

Teachers' math lessons were observed immediately before and after the staff development program, and again three months later. Also, their students were administered curriculum-referenced and nationally standardized mathematics tests before and three months after the program.

Both trained groups of teachers outperformed the untrained group in implementing the program's instructional model in the post lesson ($p=.001$) and delayed lesson ($p=.001$). Both trained groups also had a lower percentage of off-task students in the post lesson ($p=.10$) and delayed lesson ($p=.05$). Students of teachers in both trained groups made slightly greater gains than control group students on the achievement tests.

The two trained groups did not differ in implementation of the instructional model in the post lesson, but the principal involvement group had a higher implementation level ($p=.07$) in the delayed lesson. The two groups did not differ in percentage of off-task behavior in either lesson, but this was not an explicit objective of the staff development program. Students of teachers in the principal involvement group outgained students of teachers in the regular inservice group ($p=.07$) on the curriculum-referenced test. The regular inservice group outgained the principal involvement group ($p=.10$) on the computation subtest of the standardized measure.

Taken together, the results demonstrate that involving principals in a staff development program increases implementation of the program's objectives. If their instructional leadership role is activated, principals can (1) selectively direct teachers' attention to particular instructional improvement objectives, and (2) help teachers maintain the improvement over time. The study also found that district administrators play a low-profile, but central role in staff development program implementation.

TABLE OF CONTENTS

PREFACE

A study of this magnitude requires the effort of many individuals. We wish first to acknowledge the district administrators, principals, and teachers who participated in the experiment. They responded graciously to the many procedures specified in the project design. Their participation was motivated not only by a desire to improve their own school, but to contribute to education generally.

The eight substitute teachers who conducted the classroom observations were largely responsible for the success of this phase of the study. They were delightful colleagues. Although we would like to acknowledge each of them by name, we must refrain so as to preserve the anonymity of the participating ditricts. Frank Gebhard, professor of education, also made helpful contributions to this phase of the study.

Marie Hahn and Mark Schalock spent many hours organizing the data for the computer. Their patience in tracking down fugitive test scores and in "cleaning" the data is much appreciated. Mark Schalock also conducted some of the interviews of principals.

Doug Grouws' consultation to the project in its initial phase proved invaluable. Tom Good also provided helpful consultation and support. We appreciate their careful critique of the final report.

Several staff members of the Center for Educational Policy and Management contributed to our efforts. Ken Duckworth developed a conceptual framework and research agenda that helped us to formulate the particular problem investigated here. Bob Mattson provided continuing dialogue about how our study relates to larger issues of educational policy. Jane Arends contributed efficient and insightful administrative support to the project. Wynn DeBevoise and Jo Ann Mazzarella provided expert editorial assistance in the preparation of the report. Terri Williams and Sissel Lemke typed the report. We thank all of you for your help.

Meredith Gall, professor of education, was one of three project co-directors. He prepared the technical proposal for the project, managed the collection of classroom observation data, and wrote the final report with the exception of Chapter 3. Any inaccuracies in the report are his responsibility alone. Glen Fielding, associate professor of education, was another project co-director. He designed many of the experimental procedures, provided continuing liaison to the schools, implemented the instructional leadership component of the experiment, and wrote most of Chapter 3. Del Schalock, research professor of education and assistant dean, was the third project co-director. He made essential contributions to the design of the experiment and was largely responsible for the recruitment of the district sites. He also co-founded the Valley Education Consortium and developed the approach to curriculum-referenced testing used in the experiment.

Sandy Charters, professor of education, brought order to the data by conceptualizing and executing most of the statistical analyses presented in Chapter 4. In addition, he contributed to the final report by interpreting the meaning and significance of the statistical results. Jerzy Wilczynski, assistant professor of computer science, created the computer data files used in all of the analyses, and he executed many of the district-level analyses reported in the appendix. Miguel Perez, a doctoral student in educational psychology, also contributed to this effort.

CHAPTER 1

RESEARCH PROBLEM AND OBJECTIVES

Statement of Purpose

The purpose of the present study was to determine whether the effectiveness of a staff development program for elementary teachers can be improved by training the teachers' principals in instructional leadership skills related to the program. Although instructional leadership is considered an important characteristic of good school administrators, few attempts have been made to study its effects by experimental manipulation.

A related purpose of the study was to replicate previous research on a specific staff development program called Active Mathematics Teaching. This program was developed and validated by Tom Good and Doug Grouws at the University of Missouri (Good, Grouws, and Ebmeier 1983). Their research demonstrated that the program increased teachers' use of active teaching behaviors and, subsequently, improved students' math achievement.

The experiment reported here was done under the auspices of the Center for Educational Policy and Management (CEPM) at the University of Oregon. One of CEPM's missions is to discover how staff development can be used as an instrument of policy to bring about school improvement (Duckworth 1983). The present study is related to that mission.

The Problem

Staff development continues to be of interest as a strategy for improving the use of human resources (teachers, teacher aides, principals, and others) in schools. Many educators are involved directly or indirectly in staff development (Joyce, Howey and Yarger 1976), and there is substantial investment in it by universities, school systems, and individual educators. For example, a survey of Oregon school districts several years ago (Schalock 1977) found that typically 3 to 5 percent of school district budgets is allocated to staff development although it is usually not identified as a specific line item.

Despite continuing interest in staff development, it remains a poorly understood phenomenon. A previous project by CEPM researchers (Gall, Haisley, Baker, and Perez 1982) found that staff development is prevalent in school districts, but that it lacks many of the characteristics identified by research as effective. Teachers receive almost two weeks of inservice per year but it consists mostly of one-shot activities dispersed over a wide range of unrelated goals.

The use of staff development as an instrument of policy for school improvement has not been clearly articulated. As a start toward formulating such a policy, it is helpful to think of staff development as having four purposes:

1. the personal professional development of teachers;
2. the credentialling of teachers for role positions;
3. the induction of new teachers into the profession;
4. school improvement.

These purposes are not necessarily mutually exclusive. A particular staff development program can serve several purposes. In practice, though, staff development programs are either instruments of policy to promote the development of the individual teacher (the first three purposes) or they are instruments of policy to promote school improvement. This policy distinction is evident in the cyclical Staff Development/School Change model of inservice education developed by Miller and Wolf (1979) and in the separate models of inservice education to meet school system needs and individual teacher needs developed by Campbell (1981).

The present study was intended to develop knowledge about the use of staff development as an instrument of policy to promote school improvement. The primary unit of sampling, intervention, and analysis was the individual school. An experiment was done to determine whether a particular staff development program could increase the capacity of a school's teachers and principal to improve the basic skills achievement of students.

The connections between staff development and school effectiveness are for the most part assumed rather than empirically validated. Our previous research (Gall et al. 1982) suggests that the causal chain from staff development activity to teacher change, and then to student gains, is poorly articulated in practice. There is experimental evidence, however, that this sequence of effects is possible.

This line of research includes a recent set of four experiments reported by Good and Grouws (1979); Anderson, Evertson, and Brophy (1979); Crawford et al. (1978); and Stallings (1980). (These experiments are summarized in Gage and Giaconia 1981, and in Mohlman, Coladarci, and Gage 1982). Each of the experiments was successful in demonstrating positive effects of a staff development program on teacher instructional performance and student achievement in basic skills. The content of the staff development programs involves organized sets of behaviorally-defined instructional skills that were validated against student achievement criteria in prior correlational research. These skills collectively have been called a "direct instruction" model (Rosenshine 1976).

A staff development program that was tested in one of the experiments (Good and Grouws 1979) was used in the present study. This staff development program was selected for several reasons. First, it produced clear, significant effects on student achievement in three experiments (Good and Grouws, 1979; Good et al. 1983). Second, teachers reacted favorably to this program (Keziah 1980; Andros and Freeman 1981). Third, the school districts that were recruited for the present study had identified mathematics -- the subject of the Good and Grouws inservice program -- as a priority for school improvement. And fourth, the direct instruction skills covered in the Good and Grouws program are not tied to a specific mathematics curriculum.

An important problem in research on staff development is whether variations in format of staff development programs affect their outcomes. Relatively little is known about which variations are most effective. The variation of interest in this study was whether involvement of school principals in staff development programs for teachers contributes to the programs' effectiveness.

There is currently much interest in the role of the principal as instructional leader. Recent reviews of research on instructional leadership (Sweeney, 1982; DeBevoise 1984) have concluded that the principal's leadership behavior has a significant effect on students' academic achievement. This research consists entirely of correlational and case studies, however. Experiments are needed to determine the causal properties of principal leadership behaviors and to validate procedures for increasing principal leadership capability.

It seems reasonable to hypothesize that a school principal might enhance or undo the intended effects of a staff development program designed to increase basic skills achievement of students. On the positive side, the principal can engage in such behaviors as demonstrating approval of the program's goals, rewarding teachers for implementation of the program's objectives, securing resources, and facilitating needed organizational change. On the negative side, the principal can weaken the inservice program by such behaviors as refusing to acknowledge its presence, encouraging teachers to pursue other priorities, and failing to provide needed resources. These positive and negative behaviors seem important, but they have been little studied through experimental research. A major purpose of the present study was to determine whether school principals can be trained in positive leadership behaviors specific to staff development programs for teachers, and whether such training enhances the effects of staff development programs on teachers and students.

The following hypotheses were tested in the experiment:

1. Teachers who participate in the staff development program (with or without principal involvement) will use active teaching techniques more frequently than teachers who do not participate in the program.
2. Teachers who participate in the staff development program with principal involvement will use active teaching techniques more frequently than teachers who participate in the program without principal involvement.
3. Students of teachers who participate in the staff development program (with or without principal involvement) will earn higher scores on mathematics achievement tests than students of teachers who do not participate in the program.
4. Students of teachers who participate in the staff development program with principal involvement will earn higher scores on mathematics achievement tests than students of teachers who participate in the program without principal involvement.
5. Students of teachers who participate in the staff development program (with or without principal involvement) will exhibit less off-task behavior during instruction than students of teachers who do not participate in the program.
6. Students of teachers who participate in the staff development program

with principal involvement will exhibit less off-task behavior than students of teachers who participate in the program without principal involvement.

CHAPTER TWO

REVIEW OF LITERATURE

Three areas of research are relevant to the present study: research on the effects of staff development; research on variations in staff development; and research on principals' instructional leadership. Each area will be discussed as it relates to the purpose of the study.

Effects of Staff Development

The intent of staff development (sometimes called "inservice education") is to increase teacher effectiveness. Bruce Joyce and his colleagues have defined inservice education as the "formal and informal provisions for the improvement of educators as people, educated persons, and professionals, as well as in terms of the competence to carry out their assigned roles" (Joyce, Howey and Yarger 1976, p. 6).

Staff development is a large, labor-intensive effort that involves the expenditure of substantial school system resources. In 1975 it was estimated that "...there may be as many as a quarter of a million persons in the United States who engage as instructors in some form of ISTE (inservice teacher education) activity--this is about one instructor for every eight teachers" (Joyce et al. 1976, p. 6). At that time, too, about half of American teachers held a master's degree, representing a considerable investment in inservice education beyond initial certification. But despite this considerable investment, the research basis for the effectiveness of staff development is virtually nonexistent. Most of the research consists of evaluating the immediate effects of staff development on teacher attitudes, knowledge, and behavior (Showers 1982).

An important new line of research is the small but growing number of studies that have investigated, through controlled experimentation, the effects of staff development on teacher effectiveness--defined as capability to bring about improvements in student performance and basic skills achievement. These studies are recent; most of them were completed in the last five years: Anderson, Evertson, and Brophy (1979); Coladarci (1980); Crawford et al. (1978); three experiments by Good, Grouws, and Ebmeier (1983); and Stallings (1980). In each study a group of inservice teachers received the experimental training program while a control group of teachers continued their regular activities. Following the training phase, researchers observed the students of both groups over a period of time to determine training effects on students' academic achievement. Because these experiments are central to the development of a model of effective staff development, they are discussed in detail here.

With the exception of the Coladarci study, each experiment demonstrated that teachers who received the inservice program learned to use particular instructional methods more frequently than teachers who did not receive the training. More importantly, the students of the trained teachers scored higher on achievement tests of basic skills than control group teachers (again, with the exception of the Coladarci study). For example, students of the experimental teachers in Good and Grouws' fourth-grade experiment increased from a percentile of 26.57 to 57.58 on the SRA Mathematics Achievement Test over a two and one-half month period. The control group (the whole-class instruction subsample) made a substantially smaller gain -- from the 25th percentile to the 43rd percentile on the same test.

Table 2-1 summarizes the methods used in the experiments. Examining commonalities in the four successful experiments, we find that each of the inservice programs involved at least two meetings. (The "minimal" group in Crawford

Table 2-1

Instructional Process Used in Basic Skills Inservice Programs

1. Anderson, Evertson, and Brophy (1979)

   Project staff met with teachers to discuss the study. Teachers then read a 33-page manual describing 22 research-validated principles of reading group instruction, and took a short quiz on it. Teachers met once again with project staff to discuss the manual. One subgroup of these teachers was observed for their implementation of the principles throughout the school year. Another subgroup was not observed. (End-of-year student achievement in the two trained groups did not differ from each other).

2. Coladarci (1980)

   Five training packets (the same as in the Crawford and Stallings experiment) were mailed to experimental group teachers. The teachers received one packet per week, and responded to a test and questionnaire on it.

3. Crawford and Stallings (1978)

   The "minimal" training group received one training manual and self-administered test per week for five weeks. The "maximal" group received the same manuals and tests, and also came to a two-hour meeting with project staff each week. In these meetings the teacher discussed, practiced, and studied the techniques; engaged in role-playing exercises; and viewed videotapes of a "model" teacher performing the behaviors. (End-of-year student achievement was higher for "maximal" group).

4. Good and Grouws (1979)

   Teachers attended an introductory 90-minute meeting and then read a 45-page manual of research-validated principles of mathematics instruction. Two weeks later teachers attended another 90-minute meeting in which project staff responded to their questions and concerns.

5. Stallings (1980)

   Each teacher was observed for three days, then given a quantitative summary of the observation as feedback to help change his/her instruction to conform to research-validated specifications. Teachers also attended four two-hour workshops over a two-and-one-half month period.

and Stallings' study did not include any meetings, with resulting lower end-of-year achievement scores relative to the "maximal" group). Another common feature across the studies is the use of brief manuals to describe the desired behaviors.

Teacher behavior was observed and critiqued in two of the studies. Stallings' teachers were observed in their classrooms and given both a qualitative and quantitative summary of the results. Crawford's "maximal" group teachers were observed in role-playing exercises during meetings. Teacher behavior was observed in one of Anderson's trained groups, but the observations were not shared with the teachers.

It appears that extensive skill training of the type used in microteaching programs (e.g., Borg, Kelley, Langer, and Gall 1970) is not necessary in a basic skills inservice program. The critical elements appear to be (a) the opportunity to study manuals that present research-validated principles of instruction, and (b) the opportunity to discuss these principles in meetings with an inservice trainer and other teachers at the same grade level. The value of supplementing these process elements with classroom observation and feedback has not yet been clearly established.

The results of this research suggest that teacher effectiveness in basic skills instruction can be increased by a relatively simple staff development process. It should be noted, though, that these studies extended over a period of no more than a single school year. Also, the training programs were not successful for all teachers. Training processes not used in the five experiments described above may produce more sustained effects, and effects for more teachers, than those observed in the experiments. For example, the school principal's instructional leadership may help some teachers to overcome their resistance to implementing the program's objectives. The effective principal may also be able to provide continuing leadership to help all teachers sustain changes in their classroom behavior over substantial periods of time.

Good and Grouws' Inservice Program

Of the four successful inservice programs, the one chosen for the present experiment was developed by Thomas Good and Douglas Grouws (Good et al. 1983). Reasons for selecting this program were described in Chapter 1. The key instructional behaviors covered in the program's training manual are presented in Table 2-2.

The behaviors were derived from previous correlational research on the relationship between instructional processes and student achievement processes in regular classroom settings (Good and Grouws 1977). Together the behaviors form an instructional system having these characteristics:

1. Instructional activity is initiated and reviewed in the context of meaning.
2. Students are prepared for each lesson stage to enhance involvement and to minimize errors.
3. The principles of distributed and successful practice are built into the system.
4. Active teaching is demanded, especially in the developmental portion of the lesson. (Good and Grouws 1981, p. 4).

These characteristics are similar to those of the other instructional systems tested in the experiments described above.

The effectiveness of the Good and Grouws training program has been tested in three experiments by the developers (summarized in Good et al. 1983) and in another experiment by independent investigators (Slavin and Karweit 1984). In

Table 2-2

Good and Grouws' Techniques for Mathematics Instruction in Elementary School

Daily Review (first 8 minutes except Mondays)
- (a) review the concepts and skills associated with the homework
- (b) collect and deal with homework assignments
- (c) ask several mental computation exercises

Development (about 20 minutes)
- (a) briefly focus on prerequisite skills and concepts
- (b) focus on meaning and promoting student understanding by using lively explanations, demonstrations, process explanations, illustrations, etc.
- (c) assess student comprehension
  - (1) using process/product questions (active interaction)
  - (2) using controlled practice
- (d) repeat and elaborate on the meaning portion as necessary

Seatwork (about 15 minutes)
- (a) provide uninterrupted successful practice
- (b) momentum--keep the ball rolling--get everyone involved, then sustain involvement
- (c) alerting--let students know their work will be checked at end of period
- (d) accountability--check the students' work

Homework Assignment
- (a) assign on a regular basis at the end of each math class except Fridays
- (b) should involve about 15 minutes of work to be done at home
- (c) should include one or two review problems

Special Reviews
- (a) weekly review/maintenance
  - (1) conduct during the first 20 minutes each Monday
  - (2) focus on skills and concepts covered during the previous week
- (b) monthly review/maintenance
  - (1) conduct every fourth Monday
  - (2) focus on skills and concepts covered since the last monthly review

the first experiment, one group of fourth-grade teachers received training by attending two meetings and reading a manual. The control group of teachers did not receive the training. The training closely paralleled the training used in the present experiment: the manual was the same, and there were two meetings.

The second study involved a similar experimental design, but was done with sixth-grade teachers. The training format was similar to that of the first experiment, except that a verbal problem-solving manual was added to the original training manual. This new manual covers techniques for using verbal problems without numbers, writing verbal problems, estimating the answer, reading verbal problems, and writing open sentence problems.

The third experiment maintained the same experimental design as the first two studies, but was done with junior high teachers (eighth grade). Also, the training manual included all of the material from the manual used in the second study, plus two supplements: Teachers' Manual Addendum for Junior High Work, and a new Procedural Summary for the Verbal Problem Solving Manual.

The effects of Good and Grouws' training program on students' mathematics achievement are shown in Table 2-3. Comparisons favored the trained group for each measure in each experiment, although the effect was not always statistically significant.

Effects of the training program on teachers' behaviors were systematically assessed in the fourth-grade experiment. Implementation results are shown in Table 2-4. The trained teachers conformed more closely to the instructional model than did the control teachers. An exception is that the training apparently had little or no effect on the amount and quality of time spent on developmental instruction (variables 2, 7, 8, and 9).

Table 2-4 also shows the correlations between teachers' scores on the

Table 2-3

Summary of Student Achievement Effects Observed in Three Experiments by Good and Grouws (1981)

Achievement Measures (Raw Scores)		Trained Teachers Pre	Trained Teachers Post	Control Teachers Pre	Control Teachers Post	p[1]
Fourth-Grade Experiment						
SRA Math Test	M	11.94	19.95	12.84	17.74	.002
	(SD)	(3.18)	(4.66)	(3.12)	(4.76)	
Sixth-Grade Experiment						
SRA Math Test	M	25.03	28.96	26.80	29.65	NS
	(SD)	(5.0)	(4.8)	(4.1)	(3.7)	
Math Problem-Solving Test[2]	M		14.90		14.71	.10
	(SD)		(2.0)		(1.6)	
Junior High Experiment						
SRA Computation Subtest[3]	Adjusted M		29.84		28.86	NS
SRA Math Problem[3] Solving Test	Adjusted M		21.98		20.83	.03

1. Probability value based on analysis of covariance with appropriate pretest as covariate.
2. This test was administered post-treatment only. SRA pretest was the covariate.
3. These tests were administered post-treatment only. A special math pretest was the covariate.

Table 2-4

Mean Occurrence (in percentages) of Implementation Behaviors in Fourth-Grade Experiment and Correlation of the Variables and Residualized Math Achievement Scores -- Good and Grouws (1981)

Variable	Treatment	Control	p	Correlation	p
1. Did the teacher conduct review?	91%	62%	.0097	.37	.04
2. Did development take place within review?	51%	37%	.16	.10	.57
3. Did the teacher check homework?	79%	20%	.0001	.54	.001
4. Did the teacher work on mental computation?	69%	6%	.001	.48	.005
5. Did the teacher summarize previous day's materials?	28%	25%	.69	.20	.26
6. There was a slow transition from review.	7%	4%	.52	-.02	.91
7. Did the teacher spend at least 5 minutes on development?	45%	51%	.52	-.08	.65
8. Were the students held accountable for controlled practice during the development phase?	33%	20%	.20	.12	.50
9. Did the teacher use demonstrations during presentation?	45%	46%	.87	-.15	.41
10. Did the teacher conduct seatwork?	80%	56%	.004	.27	.13
11. Did the teacher actively engage students in seatwork (first 1 1/2 minutes)?	71%	43%	.003	.32	.07
12. Was the teacher available to provide immediate help to students during seatwork (next 5 minutes)?	68%	47%	.02	.28	.11
13. Were students held accountable for seatwork at the end of seatwork phase?	59%	31%	.01	.35	.05
14. Did seatwork directions take longer than one minute?	18%	23%	.43	-.02	.92
15. Did the teacher make homework assignments?	66%	13%	.001	.49	.004

instructional behaviors and the residualized achievement gains of their students on the SRA Mathematics Test. These results suggest that mental computation (variable 4) and homework (variables 3 and 15) are the most critical behaviors in Good and Grouws' instructional system for promoting student achievement.

Data collected by Good and Grouws in the three experiments indicate that the trained teachers generally were very pleased with the training program. Similarly favorable reactions of trained teachers were also reported by Keziah (1980) and by Andros and Freeman (1981).

Slavin and Karweit (1984) recently completed two experiments involving Good and Grouws' program. In the first experiment teachers in grades 4-6 were randomly assigned to three treatments: Good and Grouws' program; training in Good and Grouws' program but with the request that each teacher ability-group his or her students within the class; and training in an individualized instruction model involving cooperative learning by students. Results revealed no treatment differences on the CTBS subtest of Concepts and Applications. The ability-grouped and individualized treatments, however, were both substantially superior to the regular Good and Grouws' program on the CTBS subtest of Computations.

The second experiment involved the same design, with the addition of an untreated control group. The treatment effects observed in the first experiment were also found in this experiment. Also, it was found that each of the trained groups (including teachers receiving the regular Good and Grouws' program) excelled the control group on the Computations subtest but not on the Concepts and Applications subtest.

Inservice Variations

The literature on inservice education, including thousands of entries in ERIC and a book-length bibliography (Collins et al. 1979), emphasizes its complex nature. Content, training methods, purposes, governance, and organizational context must be considered. The effectiveness of a particular inservice variable in improving teacher and student performance depends on how other aspects of the inservice program have been constituted. For example, variations in inservice training methods may have little effect when the criterion is student achievement gains averaged across teachers for a single school year. Training variations may have a substantial effect, though, when the criterion is student achievement gains for as many teachers as possible in a school site and sustained over a multi-year period.

In a previous study, one of the present researchers and his colleagues (Gall, Haisley, Baker and Perez 1982) developed a staff development model that includes 29 dimensions. It is an elaboration and restructuring of the model developed by the staff of the Inservice Teacher Education Concepts Project (Joyce, Howey and Yarger 1976). The primary source for the elaborated model was the recent conceptual and empirical work on curriculum implementation and school improvement (Hall and Loucks 1980; Berman and McLaughlin 1978). Table 2-5 presents the list of staff development dimensions compiled by Gall and colleagues.

The model proved useful in designing the present study because it suggested features that should be included in a treatment or controlled across treatments. For example, the present experiment did not manipulate presence/absence or type of readiness activities (dimension 13). Research on curriculum implementation however, suggests that readiness activities are important to the success of a new school program (including inservice programs). Thus, appropriate readiness activities were included in each of the experimental treatments so that the absence of this factor did not depress the effectiveness of the staff development program. Similarly, such factors as scheduling and training site con-

Table 2-5

Dimensions for Analyzing Inservice Education

A. Teacher Objectives of an Inservice Activity

1. Target competencies
2. Side- or long-term effects
3. Operationalization
4. Complexity
5. Expected level of competence
6. Measurement of teacher competence
7. Rationale

B. Student Objectives of an Inservice Activity

8. Needs assessment
9. Target objectives
10. Side- or long-term effects
11. Measurement of student objectives
12. Expected level of proficiency

C. Delivery System

13. Readiness activities
14. Instructional process
15. Maintenance/monitoring
16. Training site
17. Relevance of content
18. Trainer
19. Scheduling
20. Cost

D. Organizational Context

21. Purpose for participation
22. Inservice cohorts
23. Concurrent organizational changes

E. Governance

24. Governance structure
25. Teacher participation in governance
25. Participation
27. Incentives
28. Sanctions
29. Rationale

ditions were designed in accordance with the best knowledge available about teacher preferences.

The Good and Grouws inservice program specifies some of the key dimensions of the model. For example, the program specifies: target competencies (no. 1), operationalization of competencies (no. 3), measurement of teacher competence (no. 6), rationale for teacher objectives (no. 7), target objectives (no. 9), and measurement of student objectives (no. 11).

The principal's instructional leadership is subsumed in the model shown in Table 2-5 under "concurrent organizational changes" (dimension 23) that occur during an inservice program. When an inservice program is used for school improvement it seems likely that various organizational conditions (e.g., principal leadership, curriculum revision, staff changes) can support or interfere with the program's objectives. Of these conditions, principal leadership seems the most critical to the success of a staff development program directed toward school improvement. The next section reviews the research evidence in support of this proposition.

Instructional Leadership of Principals

DeBevoise (1984) concluded from her review of the literature that there is a lack of consensus about the meaning of the term "instructional leadership." According to some educators, instructional leadership refers to any function that promotes effective operation of a school. Other educators define instructional leadership in terms of more specific functions such as teacher supervision, management of school change efforts, and clarification of school goals.

The present study used a restricted definition of instructional leadership. Instructional leadership was defined here as the principal's demonstration of skills and behaviors that increase the effectiveness of a staff development program for the school's teachers. In the case of Good and Grouws' program, effectiveness is defined in terms of (a) teachers' use of an active teaching model for the mathematics curriculum, and (b) students' gain in mathematics achievement. According to these definitions, a principal demonstrated instructional leadership in the present study whenever he or she engaged in behaviors that helped teachers to use the active teaching model more skillfully and more often, and whenever he or she engaged in behaviors that helped students to improve their mathematics achievement.

Research on instructional leadership has become widespread in recent years. Most investigations have involved comparisons of more effective and less effective schools, where effectiveness has been defined in terms of student achievement gains aggregated to the school level. Reviews of these investigations have been done by Bossert, et al. (1981), DeBevoise (1984), Pitner (1981), Shoemaker and Fraser (1981), and Sweeney (1982).

Table 2-6 summarizes effective instructional leadership characteristics and behaviors based on available research and on a recent comprehensive review of research (Leithwood and Montgomery 1982). Most of these characteristics and behaviors are stated at a high level of generality. However, they provided useful suggestions for designing the instructional leadership component of the principal involvement treatment in the present experiment. For example, several studies (Kean et al. 1979; New York State Office of Education 1974; Wellisch et al. 1978) found that principals of effective schools regularly observed and gave feedback on teachers' instruction. Also, effective principals attend teachers' inservice sessions (Leithwood and Montgomery 1982). Based on these research findings, principals in the principal involvement treatment were given training and directions to attend Good and Grouws' program and then to observe

Table 2-6

Characteristics and Behaviors of Effective Principals Identified in Correlational Studies

Study	More effective principals:
1. Austin (1978)	- participate more fully in instruction - have high expectations of themselves, their teachers, and their students - are oriented toward cognitive rather than affective goals
2. Brookover & Lezotte (1979)	- are assertive in their instructional leadership - assume responsibility for evaluating the achievement of basic objectives - are disciplinarians
3. Edmonds (1979)	- communicate effectively with their teachers and respond to teacher difficulties
4. Kean et al. (1979)	- make frequent observations of reading classes - are reading specialists
5. Madden et al. (1976)	- have more impact on educational decision making
6. New York State Office of Education (1974)	- engage in positive interaction with teachers - make frequent informal classroom observations
7. Venezky & Winfield (1979)	- have a strong desire for academic achievement and transmit this desire to staff and pupils
8. Weber (1971)	- are clearly identifiable instructional leaders - lead their school's beginning reading program
9. Wellisch et al. (1978)	- express a strong concern about the instructional program - regularly review and discuss teachers' performance - participate more often in instructional decisions
10. Leithwood and Montgomery (1982)[1]	- place the academic achievement and happiness of students first in their priorities - articulate high expectations for teachers - hold regular and frequent staff meetings - attend inservice sessions provided for teachers - support and provide staff development for teachers - monitor student progress closely - facilitate communication between the school and the community

Table 2-6 Continued

11. Sweeney (1982)[1]	- give high priority to activities, instruction, and materials that foster academic success - take part in instructional decision-making - ensure that the school's climate is conducive to learning - monitor student achievement on a regular basis - interrelate course content, sequences of objectives, and materials in all grades - support teachers' participation in inservice programs

---

1 A review of research on characteristics of effective principals.

their teachers as they worked on implementing the active teaching system in their classrooms.

The research on instructional leadership of principals summarized in Table 2-6 was based on descriptive and correlational designs. No studies have been reported in which instructional leadership behaviors were experimentally manipulated to determine their effect on teachers' classroom performance and on student achievement. The present study makes a departure from previous research because it includes manipulation of instructional leadership behavior as a primary feature of the experimental design.

CHAPTER THREE

RESEARCH METHODOLOGY

Research Design

The research hypotheses stated in Chapter 1 were tested using a pre-post experimental-control group design. Schools were nonrandomly assigned to three treatment groups. The research design is shown as follows:

$O_1, O_2, O_3 \quad X_1 \quad O_1, O_2, O_3$

$O_1, O_2, O_3 \quad X_2 \quad O_1, O_2, O_3$

$O_1, O_2, O_3 \quad X_3 \quad O_1, O_2, O_3$

where:

O_1 = observational measure of teacher behavior and student off-task behavior

O_2 = curriculum-referenced test of student math achievement

O_3 = nationally standardized test of student math achievement

X_1 = Good and Grouws' inservice program without principal involvement

X_2 = Good and Grouws' inservice program with principal involvement

X_3 = control condition involving the absence of training in Good and Grouws' program

Fifteen schools were assigned to the three treatment conditions before pretraining observations began. All participating classrooms in a school were assigned to the same treatment condition. Teachers in the inservice program treatments participated in two training sessions scheduled during the period January 31 - February 18, 1983. Principals in the principal involvement treatment attended the same two sessions and two additional leadership sessions, one at the end of February and the other at the end of April. Teachers in the principal involvement treatment participated in follow-up activities with their building principals spaced over several months after the two training sessions.

Classroom observation of teacher behavior and student off-task behavior in a math lesson occurred at three points in time. The first lesson was observed during the week of January 24, 1983, which was prior to the staff development program; the next lesson was observed during the two week period of March 8 - March 18, 1983, following the staff development program; and the final lesson was observed during the two-week period of May 3 - May 13, 1983. A few teachers who were unavailable within the two-week period were observed the week following.

The curriculum-referenced test was administered in January 1983, and then again in May 1983. The nationally standardized test was administered in one district in October 1982 and then again in May 1983. In the other two districts the test was administered in May 1982 and then again in May 1983.

Recruitment of Sample

Recruitment of Districts

Three districts in western Oregon participated in the experiment. Each of the participating districts was selected because of the commitment expressed by district administrators to increasing student achievement in elementary mathematics, and because of the long-term program improvement and staff development activities that each had sponsored. The study was viewed by district personnel as an opportunity to extend instructional change efforts already underway.

One participating district (designated District 1 hereafter) was previously engaged in an extensive instructional improvement effort designed to promote teachers' use of a model of instruction called "Elements of Effective Instruction" and commonly referred to as "Instructional Theory into Practice" (hereafter the model will be referred to as ITIP) (Hunter 1976).

The district had organized a series of courses over several years for teachers in using the model. Courses for principals also had been offered to enable them to use the model as a basis for instructional supervision.

Administrators in this district regarded the Good and Grouws' model as complementary to the ITIP approach. The models have several important features in common. For example, both emphasize the importance of having the teacher provide structured, carefully monitored practice in using a skill before permitting students to practice it independently. Active teaching, however, focuses more specifically on mathematics instruction, whereas ITIP applies to teaching across subject areas. District personnel perceived the study as an opportunity for teachers to apply the more general training in instructional skills that the district had provided previously to a specific curriculum area.

During 1982-83 two of the districts (designated Districts 2 and 3 hereafter) were field testing an elementary mathematics program that teachers from the districts had helped to develop. This program, which is described in greater detail later in the chapter, includes learning goals for students, test items for monitoring student progress toward goal attainment, and guidelines for administering the program. The program does not provide, however, a structure for organizing and managing daily lessons at the level of specificity in Good and Grouws' "active teaching" model. District staff wished to determine whether training in the instructional model would enhance teachers' capacity to foster students' achievement of the learning goals specified in the districts' new mathematics program.

Recruitment of Schools Within Districts

All elementary schools in the three districts that contain at least one fourth grade and one fifth grade classroom were included in the sample.

Several schools in two of the districts contained classrooms with children at many different grade levels. These were small schools in isolated areas within the districts, and appeared to be unrepresentative of elementary schools in general. They were excluded from the sample. The number of schools that were selected from each district by treatment is shown in Table 3-1.

Recruitment of Teachers Within Schools

The schools in the sample varied according to the number of teachers assigned to grades four and five. Of the 15 sample schools, eight had two fourth and two fifth grade teachers; four had three or more teachers at a grade; and two had a single teacher at each grade.

To make the sample of teachers within each school as similar as possible, two teachers were selected from each grade in those schools containing two or more fourth and fifth grade classrooms. The principal in the principal involvement treatment in District 2 requested that all three fifth grade teachers in his building participate in the study. This request was honored. One District 3 teacher in the regular inservice group went on maternity leave shortly after receiving training and was eliminated from the sample. The number of teachers by grade per school is indicated in Table 3-2. A detailed presentation of the district, school, grade level, sex, and treatment group of each participating teacher is shown in Appendix A.

Recruitment of Principals Within Districts

A total of five principals from the three districts were selected to serve in the principal involvement treatment. Three principals from the large district participated in this treatment condition; one principal from each of the smaller districts participated.

Table 3-1

Treatment Group Composition

		Schools N	Teachers N
Principal Involvement Group	District 1	3	9
	District 2	1	5
	District 3	1	4
Regular Inservice Group	District 1	3	10
	District 2	1	4
	District 3	1	3
Control Group	District 1	3	10
	District 2	1	4
	District 3	1	4
Total		15 Schools	53 Teachers

Table 3-2

Grade Level and Sex Composition of Teachers in Treatment Groups

		Principal Involvement Group N	Regular Inservice Group N	Control Group N
Grade	4	8	8	10
	5	10	9	8
Sex	M	7	8	7
	F	11	9	11

35

The same procedure for recruiting principals was used in all three districts. Project staff presented an overview of the study, including a general description of the principal involvement treatment, to central office staff and building principals in each district during December 1982. The administrators were told that selection of principals for participation in the principal involvement treatment would be left to the discretion of each district. Project staff indicated that it would be desirable from a research standpoint to choose principals on a random basis, but the use of random selection procedures was not a condition for the district's participation in the study.

As it turned out, none of the three districts randomly assigned principals to the principal involvement treatment. Rather, principals were selected on the basis of their interest in the study, their reputation as effective leaders, or on their current work load. Work load was a factor because some principals had been assigned leadership roles in other projects and had no time available for the leadership role called for in the present study.

According to administrators in each district, principals were assigned to the regular inservice or control treatment based primarily on their expressed interest in the experiment. Principals indicating a desire to learn about the Good and Grouws model were assigned to the regular inservice group. Principals who did not express a particular interest in the model were assigned to the control group.

Financial Incentives for Teachers and Principals

Each of the ten schools assigned to the control or regular inservice treatments received $100 as an incentive for the participating teachers. The control group teachers were viewed as "participants" because even though they did not attend training sessions, their participation was required for the research observations and testing.

Each school assigned to the principal involvement treatment received $250 as an incentive. Schools assigned to this treatment received more money in recognition of the extra demands that the treatment placed on principals. The principals were asked to attend training sessions, to make two formal evaluations of each participating teacher's lessons, and to complete several other tasks.

The money was paid to a general fund in each of the 15 participating schools, to be used in accordance with school priorities and policies.

Description of Sample

Description of Districts

All three districts were located in Oregon's mid-Willamette Valley. District 1 was situated in a large, suburban community and served about 7,000 students. The other two districts, separated by only about ten miles, served small communities. Each had a student population of approximately 2,400. These small districts had been engaged in collaborative school improvement projects for over ten years and were both using the same elementary mathematics program.

Description of Schools

Information was collected on the size of schools in the sample, the grade levels they served, and the policies governing student classroom assignment for mathematics instruction at grades 4 and 5.

The student population of the elementary schools in District 1 ranged from 147 to 740. Each of these nine schools had classrooms from kindergarten or grade one through grade six.

The student population of schools in the sample from the small districts ranged from 254 to 465. Five of the six schools in these districts had classrooms at grade one through six. One school included only grades four through six.

The dominant policy guiding classroom placement of children for mathematics instruction across all three districts was to maintain heterogeneous classrooms, that is, classrooms containing children with widely different backgrounds, abilities, and learning characteristics. However, three schools in District 1 and one school in the small districts practiced within-grade ability grouping for mathematics instruction. For example, one fourth grade teacher might teach math to all high-performing students at that grade level, while the other provided instruction to all moderate-to-low-performing students.

Teachers' Inservice Background

The project staff collected information on district-sponsored inservice programs for teachers offered during the past several years. Virtually all teachers in District 1 had taken at least one course in ITIP conducted by the director of staff development for the district. The director is a highly respected ITIP trainer, and she conducted the training sessions in District 1 in the present experiment.

Teachers in the small districts did not share this background of experience, nor for that matter any common background of inservice work. However, they did have a common experience in Valley Education Consortium (VEC) mathematics program (described later), and in working with building colleagues to identify and resolve problems encountered in implementing this program.

Description of Principals

All of the 15 principals of the sample schools were interviewed prior to the intervention (January 1983) to obtain background information on school policies and procedures and on various instruction-related practices generally followed in their school. A copy of the interview schedule that was used in these meetings is in Appendix B.

Project staff wished to determine the extent to which the practices called for in the principal involvement treatment were consistent with principal's pretreatment practices, or differed from them. This treatment called upon principals to attend the training sessions for teachers on the Good and Grouws model and to observe teachers as they worked with the model in their classrooms. The principals were to engage in several other practices, but participation in training and observation of classroom practice represented the heart of the treatment.

Several important findings emerged from these interviews. Perhaps the most important was that principals assigned to the principal involvement groups and the regular inservice groups had participated in inservice training sessions for teachers more often than had principals in the control group. This was the same in all three districts. Virtually all principals assigned to the treatment groups involving training indicated that they tried to participate in all inservice programs offered in their buildings. These included inservice programs sponsored by the publisher of a recently adopted textbook series, workshops on computer use led by a member of the school staff, and a presentation of new techniques in classroom management. By contrast, principals in the control schools made comments like: "I don't have much involvement in teachers' inservice," or "I participate very little in teachers' inservice." Principals in control schools generated fewer examples of inservice training sessions they had attended than did principals in treatment schools.

A second finding was that District 1 principals in the schools assigned to receive training were more likely to use ITIP as the primary basis for instructional supervision than were principals in control schools in that district. Principals in the small districts, regardless of the experimental group to which they were assigned, did not use any particular model of instructional effectiveness as a guide for supervision. Information on key principal practices is summarized in Table 3-3.

Description of District Mathematics Programs

The elementary mathematics program in District 1 contains general learning goals for each grade level, specific i tructional objectives, and illustrative teaching and assessment procedures that relate to the stated goals and objectives. This information is published in a "Curriculum Guide Notebook" for teachers. The district's textbook series in mathematics appears to correspond to a large degree with the published curriculum.

The elementary mathematics program in the two small districts was developed through the Valley Education Consortium (VEC), an organization of small school districts and educational agencies in a three county area in Western Oregon. The major purpose of VEC is to foster long-term school improvement. (A description of the Consortium and its work i. presented in Appendix C.) Staff from Districts 2 and 3 had extensive involvement in the development of VEC's mathematics program. The elements of this program are listed in Table 3-4. It was developed on the assumption that a district would adopt a textbook series and obtain related materials that teachers need to promote students' achievement of designated learning goals. Informal discussions with teachers in the two districts implementing the program

Table 3-3

Key Principal Practices by District and Treatment Prior to the Experiment

	Principal Involvement and Teacher Inservice Groups	Control Groups
District 1 (large district)	. frequent participation in inservice training for teachers	. infrequent participation in inservice training for teachers
	. extensive use of the ITIP model in the supervision process	. more limited use of the ITIP model in the supervision process
Districts 2 & 3	. frequent participation in inservice training for teachers	. infrequent participation in inservice training for teachers
	. supervision not guided by any particular instructional model	. supervision not guided by any particular instructional model

Table 3-4

Major Elements of the
VEC Mathematics Program
Grade 1-8

- learning goals organized by grade level
- test item pools and other assessment procedures that teachers and administrators can use to monitor student progress toward goal attainment
- computer programs for scoring, analyzing, and reporting test information
- grade-level handbooks for teachers that provide guidelines for goal-based planning, assessment, and decision-making
- guidelines for assessing and reporting students' progress toward graduation competence requirements
- goal-based program evaluation and reporting procedures
- guidelines for choosing and using program-related standardized tests

indicated that the texts in use relate reasonably well to the VEC curriculum, although several teachers suggested that supplementary material is needed to accommodate the full range of specified learning goals.

The main difference between the VEC program used by the two small districts and the program in place in the large district lies in the area of assessment. The VEC program contains extensive test item pools that teachers can use on a day-by-day basis to assess student progress toward goal attainment. Administrators are able to use the same test item pool and to monitor program effectiveness through district-managed mid-year and end-of-year tests.

The textbook series adopted by the large district contains well-developed tests and assessment procedures, although these assessment resources are not as extensive as those available in the VEC program. Guidelines for using test information also are less well developed. This was the case for both teachers and administrators.

Training of Teachers in the Three Treatments

Good and Grouws' Version of Training Program

The training program used by Good and Grouws in their experiments was described by them as being a "treatment [that] consisted of two 90-minute training sessions and a 45-page manual that detailed the treatment and provided a base for teacher reference as necessary" (Good and Grouws 1979, p. 356). In the first session teachers were given a brief orientation to the experiment, an explanation of the program, and a 45-page manual (reproduced in Good and Grouws 1981) to read after the session. Two weeks later a second session was held "...to respond to questions that teachers had about the meaning of certain teaching behaviors and to react to any difficulties that the teachers might have encountered" (Good and Grouws 1979, p. 356).

Professor Grouws met with the research staff to discuss the training sessions so that they could be repeated in the present experiment. A manual describing the training sessions was not available, so it was helpful to meet with Professor Grouws, who had conducted the training sessions in several of the experiments involving the Good and Grouws' program.

Everyone agreed in these meetings that it would be advantageous to double each session from ninety minutes to three hours in order to strengthen the training. For example, the additional time made it possible to show teachers two videotapes illustrating Good and Grouws' instructional model. These videotapes were developed subsequent to Good and Grouws' experiments. Also, the extended three-hour sessions gave the trainer additional time to address teacher questions and concerns.

The training program used in the present program is described below. It should be noted that this training program was identical for teachers in both the principal involvement and regular inservice treatments. The one dimension on which the two treatments differed--namely, principal involvement--is described in a subsequent section titled, "Principals' Role in Principal Involvement Treatment." Following that section is a description of the control treatment.

It is important to note that (a) teachers in the regular inservice treatment, (b) teachers in the principal involvement treatment, and (c) principals of teachers in the principal involvement treatment attended the training sessions together. Thus, regular inservice teachers had the opportunity to hear reactions from principals from other schools to Good and Grouws' model and to observe teachers interacting with principals about the model. Nevertheless, because the teachers and principals in the principal involvement group were from other schools in the district, the influence of these participants on the regular inservice teachers was at best indirect:

the principals had no authority over them, and communication between teachers from different schools is infrequent. Follow-up interviews indicated that there was no cross-school discussion of Good and Grouws' model for the duration of the experiment, with the exception of the two training sessions.

Orientation Session

All teachers and principals in the three treatment groups attended a half-hour orientation session the week of January 10, 1983. A separate session was held in each district. The following topics were covered: purpose and design of the experiment; general nature of the training program; data collection procedures; procedures for preserving data confidentiality; the need to avoid treatment contamination; scheduling; and project staff.

Training Sessions

The training was provided separately for each school district, and an indigenous trainer was employed in each district to conduct that district's two sessions. This procedure was used so that participants in each session (teachers, principals, and trainer) could address the relationship between Good and Grouws' instructional model and local school district conditions. It was thought that implementation of the model would be facilitated if participants did not have to share "talk time" in the sessions with participants from other districts.

The trainer in District 1 was the district director of staff development. She had worked as a teacher in the district and, in recent years, had been responsible for ITIP training (Madeline Hunter's "Elements of Effective Instruction" program referred to as "Instructional Theory in Practice") in the district. In District 2 the trainer was a district specialist in staff development and teacher supervision. The trainer in

District 3 was an elementary teacher who had been involved in many inservice activities with the district's teachers. All trainers had been with their districts for many years.

The three district trainers were oriented to their role in the staff development program by Dr. Grouws during his visit. A member of the research staff provided additional training for them and gave them all necessary materials. In addition, he attended each training session in each district as an informal technical resource to the trainer. On occasion he would field questions that were addressed to him by a participant or that the trainer wished him to answer.

The first training session was held in January 31, 1983 (District 1), February 1 (District 2), and February 13 (District 3). The second training session was held about two weeks later on February 16 (District 2), February 17 (District 1), and February 18 (District 3). Substitute teachers were paid by the research project so that teachers could attend the sessions during regular working hours. All participating teachers and principals were able to attend each session.

The three district trainers were given an outline to follow. A copy of the outline is presented in Appendix D. It is generally self-explanatory, with a few exceptions. The ASCD (Association for Supervision and Curriculum Development) videotape used in the first session was developed by Good and Grouws. The videotape is titled Teaching Mathematics Effectively and is sold by ASCD for use by teacher educators. It features Professor Good explaining the instructional model, and also includes brief segments of classroom teaching that illustrate key elements of the model.

The videotape used in the second session shows a complete math lesson (approximately 40 minutes) taught by an elementary teacher who uses each element of the model. The tape is not narrated. Brief segments of it were used in the ASCD videotape. It was lent to the project by Professor Grouws.

Whole group versus individualized instruction is discussed on page 4 of the outline. This material was written for the trainer's use because some of the participating teachers emphasized individualized instruction, and it was anticipated that they might raise objections to the whole-group approach in Good and Grouws' model.

The questions and answers on pages 8-9 were provided by Professor Grouws, who recalled typical questions from his training experience. He also provided the overhead transparencies mentioned on page 2 of the outline. The lesson plan form mentioned on page 2 and again on page 6 is reproduced on page 10 of the outline.

During the first session all participating teachers and principals were given a personal copy of the manual (approximately 50 pages) used in Good and Grouws' fourth-grade experiment.

The trainers generally followed the outline for the two sessions, although there were variations across districts. The trainer in District 1 emphasized the relationship between the ITIP instructional model and Good and Grouws' instructional model. Trainers in Districts 2 and 3 related the model to local teaching conditions. The lesson plan form was used by only two trainers in just one of the sessions.

The member of the research staff who attended the training sessions took notes on each session. A summary of his notes are presented in Appendix E. A CEPM staff member attended the second training session in District 2. A brief publication based on her impressions is included in the same appendix.

Debriefing Session

Teachers and principals in the two trained groups in Districts 2 and 3 voluntarily requested a third session at the conclusion of the experiment.

This session was held June 10, 1983 in each district separately with most teachers and both principals attending. The session was used by participants to share experiences and to discuss their current views of Good and Grouws' instructional model and training format. A member of the research staff attended both sessions and took notes. A summary of these notes is presented in Appendix E.

Control Group Teachers

Control group teachers and principals attended the orientation session described above. They were told that they were participating in an experiment on improving mathematics instruction. The control group teachers were asked to continue their regular math instruction for the duration of the experiment. They were told that they could receive the inservice program used in the other treatments after the conclusion of the experiment. No information about the Good and Grouws instructional model was provided for them.

The same observation and test data were collected from the control group as from the two trained groups. No feedback based on these data was provided to the control group, however, whereas the two trained groups did receive summarized results of the post and delayed observations and pre-achievement test administration.

Principal Involvement Treatment

Design

The principals in the principal involvement treatment were expected to attend the training sessions for teachers on the Good and Grouws model, and to support teachers' use of the model in their daily mathematics instruction.

The primary form of support expected of principals was to make observations of teachers' math lessons (complete lessons were to be observed), and to provide feedback to teachers on their instructional performance. Their observations were to take place independently of those conducted by members of the research team. The principals were only expected to conduct observations of the fourth- and fifth-grade teachers who were participating in the experiment.

In addition to observing teachers' classroom instruction, the principals were expected to organize and moderate a meeting among participating teachers in their building to clarify and resolve problems encountered in using the model.

The principals' responsibilities for instructional leadership in the principal involvement treatment, and the schedule for carrying them out are outlined in Table 3-5.

Preparation and Support for Principals

Shortly after the last inservice training session for teachers was completed (mid-February 1983), training sessions for principals in the principal involvement group were held. The training sessions were intended to clarify the nature of the tasks that principals were to perform and to offer the support and training that principals needed to complete these tasks.

One training session was held for the two principals from the small districts, and one was organized for the three principals from the larger district. Project staff and the district trainers served as facilitators of these sessions.

The sessions had several specific purposes. One was to discuss with the principals the responsibilities outlined on the task-timeline sheet

Table 3-5

Leadership Responsibilities of Principals in the Principal Involvement Treatment

Task	Timeline
Principals were expected to:	
1. Attend the two-and-one-half-day training sessions for teachers;	by February 17
2. Observe one complete math lesson for each teacher in their buildings who received training (to include pre- and post-conferences);	by April 1
3. Arrange and facilitate a 30 to 40 minute meeting with participating teachers in their buildings to discuss and resolve any problems that may have developed around using the model.	by April 15
4. Make another observation of each participating teacher's math lesson including pre- and post-conferences.	by May 15

(Table 3-5). Also, the principals were given an opportunity to review and refine the tasks in light of their particular circumstances. As it turned out, all principals indicated that they were comfortable with the tasks and timelines outlined.

Another purpose of the training sessions was to introduce principals to different options for observing teachers' math lessons. Prior to the sessions, project staff and district trainers had identified three different procedures principals might use for conducting observations: (1) using a checklist that identified the key instructional practices called for in Good and Grouws' model (Appendix D); (2) using the more detailed coding system that members of the research team were employing for their observations; or (3) making a verbatim record of teachers' statements. The verbatim record was included as an option because the principals from the large district had received training in recording verbatim teacher statements and were accustomed to using this approach.

Each of these options was discussed with the principals. No one approach was presented as preferable to the others. It was intended that each principal would choose the procedure that he or she found most appropriate and workable.

The training session in the large district also focused for a brief period of time on the format used to report results on the criterion-referenced pretest that was administered in January as part of the study. Administrators in the large district were unfamiliar with this format, so time was set aside to answer questions the principals had about it.

The principal involvement treatment did not require principals to make any particular use of the pretest results. The reporting format was discussed with principals in the event that teachers in their building, who received pretest results for students in their classes, had questions about the format.

At the end of April, after principals had had an opportunity to observe teachers' math lessons and to discuss with teachers the implementation of the Good and Grouws model, a second meeting of principals was convened. As with the earlier training session, one meeting was organized for the three principals from the large district, and one for the two principals from the small districts.

These meetings were informal and lasted about 40 minutes. Their purpose was to provide an opportunity for principals to report progress and share concerns about their roles as instructional leaders. It was thought that principals might learn from each other's experience in the project. The meetings also provided a chance for project staff to check whether the tasks assigned to the principals were being carried out.

Implementation of the Treatment

As stated above, principals in the principal involvement treatment had three responsibilities: (a) to attend the two training sessions for teachers; (2) to make two observations of each participating teacher's math lessons, including pre- and post-observation conferences; and (3) to facilitate meetings among participating teachers concerning issues encountered in using the Good and Grouws model. Information is provided in this section on the degree to which these responsibilities were fulfilled.

Attendance at inservice training sessions. All principals assigned to the principal involvement treatment attended teacher training sessions. Principals were not given any special instructions for participating in these sessions, though in one district the principal made several strong comments about the importance of using various practices identified in the model. For example, he stressed the need to assign homework in accordance with the policies outlined in the model even though several teachers expressed reservations about students' willingness or capacity to complete homework assignments on a regular basis.

Conducting observations of teachers' math lessons. Information on frequency of observations was obtained through interviews with the principals in late May and early June 1983.

Two of the principals in District 1 recorded teacher statements verbatim as this was the approach to which they were accustomed. One of the principals stated that "it was helpful to have options to choose from" in developing an observational approach, but found the clinical supervision practices she generally used were applicable to the present project. Her approach was to "identify something that teachers do well and something they can improve." According to this principal, the behaviors called for in the Good and Grouws model were essentially the same as those called for in "Instructional Theory Into Practice," so little adjustment was needed to carry out the supervision process, other than learning a new set of labels for instructional behaviors.

The third principal in District 1 and the two principals in District 2 and 3 used the checklist developed by project staff for structuring classroom observations. Each found the checklist to be a simple method for focusing and recording observations.

One principal in District 1 and one in District 2 made the two observations of each participating teachers' math lessons that were specified in the treatment. One principal in the large district made three observations of each participating teacher. Another principal in this district made two observations of three of the trained teachers, but only one of the fourth. One principal in the small districts made only a single observation of each teacher. The number of observations completed by each principal is summarized in Table 3-6.

Conferences. All principals in the principal involvement treatment held brief pre-observation conferences with teachers to prepare for each classroom observation. The preconferences were arranged in order to establish a specific time to carry out the observations and to clarify or confirm the observational procedures to be used.

Each principal also held post-observation conferences with teachers. These conferences were used to share with teachers information collected through the observations and to discuss any problems that may have arisen in implementing the model.

All participants reported that teachers were implementing most of the practices identified in the active teaching model, and the post-observation conferences served to reinforce these practices. One principal submitted to project staff the checklists he had used to guide and record observations of three different classrooms, though these checklists had not been requested.

The checklists showed that the clear majority of desired practices were being followed in each classroom. The records indicated, however, that one teacher had not conducted mental computation exercises, and another teacher had spent less than 20 minutes on the development phase of the lesson. The principal in this school said that he had discussed these omissions and deviations from the model with the teachers and came to the conclusion that there were valid reasons for not doing mental computation during every lesson or for adhering precisely to the recommended time allocations. "Kids get tired of mental computation if they have to do it each day," a teacher had told him. Another teacher indicated that she needed to spend more time on review that the recommended eight minutes when students' homework revealed that they had problems with a previous lesson. When review was lengthened,

Table 3-6

Frequency of Observations Conducted by Principals in the Principal Involvement Treatment

Principals	Observations
District 1	
Principal A	2 observations per teacher
B	2 observations of 3 teachers; a single observation of the fourth teacher
C	3 observations per teacher
Districts 2 and 3	
Principal D	2 observations per teacher
E	1 observation per teacher

development was shortened, she explained.

A principal in District 1 used conferences to "problem solve" with two teachers concerning ways of adapting the model to heterogeneous classes, like preparing a separate development section for each of two or three groups rather than making one presentation to the class as a whole. This necessitated shortening the development section, the principal noted. But in her view the benefits of tailoring the presentation of new material to students' learning backgrounds and abilities in some cases outweighed the costs of shortened development sections.

A principal in District 2 highlighted in conferences the importance of "manipulatives" in the development phase of the lesson. He observed that teachers generally did not use manipulatives. He concluded from the conferences that he needed to take steps to assure that high-quality manipulatives were more readily available to teachers.

Facilitating teacher use of the model. None of the principals convened a special meeting of teachers to discuss the use of the Good and Grouws model. In all schools in this treatment, however, teachers met among themselves informally or during regular staff meetings to exchange views about the model. In no case was this interaction explicitly orchestrated by the principal.

The failure of principals to organize a separate meeting for teachers in April reflected the principals' considered judgment rather than an unwillingness to comply with treatment specifications. All principals in the treatment had conversed informally with teachers about the model, had observed formally at least one math lesson taught by each teacher, and were well aware that teachers had talked about the model among themselves and shared techniques for adapting the model to their classrooms. It seemed unnecessary to organize a separate meeting in April to discuss matters that already had been discussed.

Principals' Reactions to their Participation in Experiment

Reactions to the model. Principals in the principal involvement treatment had generally positive reactions to the model. Representative comments made by principals in the interviews conducted in late May and early June follow:

District 1

Principal A: "The model was complete and well structured. It was easily related to ITIP."

Principal B: "This active teaching approach seems to be an effective way of teaching, not just math, but in other areas, too. Our new reading series includes more teacher-directed activities than in the past. It makes sense to have kids spending a large amount of time interacting with the teacher.... Also, the 15-minute limit on homework was helpful, along with the notion that homework should be regularly assigned."

Districts 2 and 3

Principal C: "I liked this method. It helped teachers organize lessons better. There were clear checkpoints to make sure things were going as they should."

Principal D: "This is a solid approach. It helps teachers keep students on task. It helps them to be more consistent and organized in their teaching..."

One principal raised doubts about certain features of the model. He saw value in the overall structure of the model, and reported that the recommendations on homework appeared to be well received by both teachers and students. He was concerned, however, about what he perceived to be the overly prescriptive time allocations for various phases of a lesson and the "bias" toward whole-group instruction embedded in the model. He also criticized the training sessions for failing to discuss the theoretical underpinnings of the model.

All the principals in the principal involvement treatment commented on the difficulty that some teachers had in using the model with heterogeneous classes. One principal indicated that the project "stimulated our thinking about ability grouping." Another principal suggested that teachers who had practiced some form of individualized math instruction for the first semester were hard pressed to alter their practices in February simply on the basis of two brief training sessions on the Good and Grouws model. This principal recommended that the model should be introduced at the beginning of the year if it is to be introduced at all.

Reactions to their role as instructional leaders. All of the principals indicated that their attendance at the training sessions provided a common basis for subsequent observations and discussions with teachers. Since they regularly attended instruction-related inservice activities, participation in the training on the Good and Grouws model was not perceived as burdensome or unusual in any way.

One of the principals, however, perceived that his district was uncertain regarding teachers' implementation of the model. He wondered whether he was supposed to require teachers to use the model, even if they had misgivings about it, or whether teachers were free to use, adapt, or reject the model as they saw fit. He would have preferred the district to convey a clear message on this issue at the training sessions.

A second principal in the large district echoed this concern about the districts' intentions. He stated that he "would have liked more explicit direction from the district."

With respect to conducting classroom observations, all the principals indicated that the treatment called for more frequent formal observations of math instruction than they ordinarily would carry out in a comparable time period. None of the principals, however, suggested that the observations called for were an unreasonable burden. The principals found the observations to be of value in assessing the strengths and limitations of the model and in focusing teachers' attention on specific elements of the model that they were using or not using.

Principals' Roles in Other Treatments

Regular inservice treatment. Principals in the regular inservice treatment were requested in the orientation sessions (described earlier) not to participate in the inservice training sessions and not to attempt to observe or evaluate teachers' use of the Good and Grouws model. They were asked to permit project staff to collect data in their building, but were not expected to participate in the study in any other way.

To verify that principals in the regular inservice treatment refrained from involvement in the study, interviews with each of these principals were conducted in May, 1983. The interviews indicated that none of the principals had attended the training sessions, and none had asked teachers about the Good and Grouws model, or had tried to observe teachers' use of the model. Several of the principals in this treatment stated that they had heard informally from participating teachers that it was difficult to use the Good and Grouws approach in heterogeneous classrooms. These same principals also indicated that they had learned through informal comments that the model called for regular homework, a practice the principals approved. However, none of the principals in the regular inservice treatment took any steps to enhance teachers' use of the active teaching approach.

One principal in District 1 who was assigned to the regular inservice treatment also taught fifth grade mathematics. This principal worked in a small school in which the building administrator customarily assumed some teaching duties. The principal participated in the inservice training sessions, as did the principals in the principal involvement treatment, but he participated in the role of a teacher, not an "instructional leader," and did not make formal observations of other teachers' mathematics lessons.

Control treatment. Interviews also were conducted in late May, 1983 with principals in the control groups. None of these principals reported having any information about the Good and Grouws model, or about the tasks that principals in the principal involvement treatment were expected to perform. None indicated that they had made any attempt to observe teachers' use of the model.

Observational Measures of Classroom Instruction

A classroom observation instrument was developed and used in the experiment. One purpose of the measure was to determine whether teachers who received training in the experiment made changes in their instructional behavior in conformance with Good and Grouws' active teaching model. A related purpose was to determine whether students of the trained teachers increased their on-task behavior during math instruction. Together, the observation instrument and the student achievement tests described below provide a set of dependent variables that were hypothesized to be affected by the principal involvement treatment and the regular inservice treatment.

Another use of the observation instrument was to measure fidelity of treatment implementation. The major outcome of the experiment was gain in student mathematics achievement. If gains were not observed, a possible explanation would be that teachers did not implement the instructional model in the two treatments that involved training (principal involvement and regular inservice). Lack of fidelity of treatment implementation is a persistent problem in field experiments in education (Charters and Jones 1973). The classroom observation instrument provides a direct measure of fidelity of treatment implementation in this experiment.

As stated in the section on experimental design, this instrument was used to observe teachers' math instruction at three points in time. It was used shortly before the training period, again about three weeks after training, and finally about ten weeks after training.

The Observation Instrument

Initial consideration was given to using or adapting the checklist format developed by Good and Grouws for assessing instructional effects of their training program. An example of their checklist format is shown in Appendix F. The type of data yielded by this instrument is shown in Table 2-4, which presents classroom instruction effects observed by Good and Grouws in their fourth-grade study.

The decision was made to retain most of the classroom variables in Table 2-4 but to measure them more precisely. Instead of relying on time estimates, observers recorded the exact amount of time that each observational behavior occurred.

The observation instrument used in the experiment is shown in Appendix G. The first page of the instrument is used to record identifying information and answers to six questions about the planned lesson. These questions are addressed to the teacher before the lesson begins to help the observer become oriented to the lesson and to help the observer decide which observation categories to assign to behaviors occurring during the lesson.

The next four pages make up the main part of the instrument and are identical to each other with one exception. The design of this part of the instrument can be understood by examining the top half of page two. The first row contains thirty-second intervals. The next 13 rows contain observation variables. With the exception of the row labeled "No. of students off-task," all of the other rows refers to observation of teacher behavior. Each of these rows refers to a single observation variable with the exception of the row labeled "Seatwork (monitored +)." This row is used to record two observation variables: monitored seatwork (+) and unmonitored seatwork (-).

In using the instrument, observers use a stopwatch to determine 30-second intervals. At the end of each interval, the observer circles the behavior that occurred during that interval. A 30-second interval was used because pilot work indicated that teachers seldom engage in more than one of the behavior categories during that period of time. If a teacher should use two of the behavior categories within a 30-second interval, the observer would circle both categories. Each category would be assigned a value of 15 seconds for the purpose of this data analysis.

Another feature of the instrument is that observers switch from recording teacher behavior to recording student off task behavior after every five minutes of the lesson. At these points in time the observer is given 30 seconds in which to look at each student in the class and determine whether he/she is on or off task. The observer counts the number of off task students and enters this total in the row labeled "No. of students off task." The "5:00" column in the top half of page 2 contains only a line on which to record the off-task student count. This feature, which is repeated at five-minute intervals, is designed to cue observers to switch from recording teacher behavior to recording student behavior.

The bottom set of rows on page 2 and the set of rows on pages 3-5 are used for continuous recording of behavior as the lesson progresses. The instrument contains space for recording lessons up to 60 minutes in length. The instrument can be modified easily by adding pages to record lessons that last beyond 60 minutes.

A manual was developed to assist observers in recording behavior on the instrument. A copy of the manual is presented in Appendix H. The manual contains a definition of each behavior and a set of decision rules to help observers record the behaviors reliably. These rules were developed based on experience during observer training and initial data collection.

The last page of the instrument (Appendix G) is used for high inference ratings of student comprehension (questions 1 and 2), and a rating of verbal problem-solving occurrence (question 3). Item 4 is used by observers to record information about classroom organization and unusual occurrences.

Content validity was the major validity concern in designing this instrument. The observation variables and manual needed to be inclusive of, and defined in the same way as, the instructional behaviors presented in Good and Grouws' Teachers Manual (Good, Grouws, and Ebmeier 1983) used in the present experiment. Dr. Grouws made a visit to Oregon to assist the research team in meeting this goal. He reviewed the philosophy of the instructional model, and showed several videotapes of teachers using the model. While showing the videotapes, he and the research team discussed the meanings of each instructional behavior and distinctions between similar-appearing behaviors. He also discussed his experiences with classroom observation using his own instruments.

In addition to this meeting with Dr. Grouws, the research team continually consulted Good and Grouws' Teachers Manual to be certain that the definitions in the observation manual (Appendix H) were consistent with the definitions in the Teachers Manual. Also, the summary of the instructional model (see Table 2-2) was consulted to check that the observation instrument included all of the essential elements of the model. The instrument and observation manual were reviewed independently by members of the research team until everyone was satisfied that they were content valid with respect to the Good and Grouws' Teachers Manual.

Data Collection

Eight substitute teachers were employed by the research team to collect the observational data. Substitute teachers were selected because they were knowledgeable about classroom instruction; because they were available at the times required; and because they were likely to be accepted by teachers as observers in their classrooms. All the observers possessed a teaching certificate, and all but one had experience teaching at the elementary school level. Also, all but one of the observers lived in or around Districts 2 and 3. The reason for this observer characteristic is that observer training occurred in these two districts, and it was convenient to select observers from the vicinity in order to minimize driving time. District 1 was approximately an hour's driving time from the main site used for observer training.

All eight of the observers completed the training period satisfactorily and participated in the first round of data collection. Two of the observers were unable to participate in subsequent rounds because they obtained longer-term employment in the interim.

The observers received five half-days of training to learn how to use the observation instrument reliably. Training involved the scoring of videotaped math lessons and also scoring of "live" math lessons in control group classrooms in District 3. Observers also received three half-days of refresher training just prior to the second round of observation and again just prior to the third round of observation. They were given very little information about the nature of the experiment. They were never told the treatment group assignment of the teachers whom they observed.

Classroom observations were scheduled at the teachers' convenience. The scheduling was arranged so that there would be no bias in assignment of particular observers to treatment groups or in time of week that a treatment group was observed. If a teacher was ill or unavailable when an observation was scheduled, a make-up date was arranged. Observations were not scheduled on days when the teacher was giving a long test, when a student teacher was instructing the class, or when an unusual event was scheduled. Also, no observations were scheduled on Mondays, because in Good and Grouws' instructional model, weekly and monthly reviews are scheduled for that day (see Table 2-2). It was possible to observe the elements of daily review, development, seatwork, and homework shown in Table 2-2 on the other four school days, except for Friday, when no homework was assigned; if an observation occurred on that day, the observation instrument manual (Appendix H) specifies a procedure to be followed.

Observer Reliability

Twenty-five of the 159 observed lessons (53 teachers x 3 times of observation) were scored by two observers. The 25 lessons were used to derive estimates of inter-observer reliability. The lessons spanned the three districts, the three times of observation, and the three treatments. Also, many different pairs of observers from the set of eight observers are represented in the 25 lessons.

Table 3-7 presents between-observer correlation coefficients for total time measures of the 13 observation categories, for lesson length, and for average number of off-task students. The coefficients are generally high.

Table 3-7

Inter-observer Reliability for Observation Categories

Observation Variable	r
Transitions	.83
Interruptions	.95
Review previous work	.61
Check homework	.95
Assign homework	.80
Quiz	.62
Mental computation	.98
Development	.96
Controlled practice	.46
Seatwork directions	.85
Monitored seatwork	.73
Unmonitored seatwork	.84
Check seatwork	.81
Lesson length	.99
Off-task students	.87

An exception is the reliability coefficient for controlled practice (r=.46). This category occurred infrequently in the sample of 25 lessons. Observers agreed perfectly on time spent in controlled practice in 21 of the lessons, and were within 2.5 minutes of each other in two or more of the lessons. In only two lessons was the disagreement excessive (0 minutes vs. 10.5 minutes, and 0 minutes vs. 7.5 minutes).

<u>Data Analysis Procedure</u>

The first step in scoring the data collected by the classroom observation instrument was to compute the amount of time that each teacher spent using each of the 13 behavior categories. When two observers had collected data on the same teachers at the same point in time, the observers' time totals for each behavior category were averaged to yield a single time score.

The off-task variable was derived by first summing the number of off-task students at each five-minute interval in the lesson. Next the sum was divided by the number of such intervals to yield a score that represented the average number of off-task students for that lesson. When two observers had collected data on the same class, their scores for this variable were averaged to yield a single score.

The teachers' time scores for the 13 behavior categories were analyzed in three ways. The primary statistical analysis was to determine whether the teacher's time use for a behavior category met the criterion established by the researchers for that category. The reason for using a time criterion is that Good and Grouws' instructional model specifies minimal, maximal, or optimal amounts of time that particular behaviors should be used. Therefore, it seemed important to determine the percentage of teachers in each treatment group who used each behavior category for the criterion time period. The rationale for the time criterion is provided in Appendix I.

The second statistical procedure involved computing the percentage of teachers in a treatment group who used a behavior category for any period of time. The third procedure was to compute the actual amount of time that teachers in each treatment group spent on each behavior category. These two analyses of the observational data provide a fuller description of teachers' instructional behavior than would be provided by just computing percentage of teachers meeting a time criterion.

Measures of Students' Mathematics Achievement

Both curriculum-referenced tests and norm-referenced tests were used as pre- and post-treatment measures of student achievement in mathematics. Tests developed by the Valley Education Consortium (VEC) for its instructional program in mathematics served as the curriculum-referenced measures. Various levels of the California Achievement Test, published by McGraw-Hill, served as the norm-referenced measures.

The VEC Tests

The VEC grade 4 and grade 5 mid-year tests in mathematics were administered in January 1983 in each of the three districts to fourth- and fifth-grade students, respectively. These tests were a pre-treatment measure of student achievement. The VEC end-of-year tests in mathematics for grades 4 and 5 were administered in May 1983 in each district as a post-treatment measure of achievement.

The VEC tests were prepared from test item pools included in the Consortium's mathematics program. The item pools had been developed by teachers, working collaboratively with staff from an educational research and development agency.

The VEC tests were designed to assess the knowledge and skills identified in the VEC mathematics program, which the two small districts in the study were implementing. Teachers in the two districts critically reviewed the tests and determined that the final selection of test items matched designated learning goals in the VEC mathematics program. A panel of teachers and administrators in the large district participating in the study analyzed the relationship between the district's curriculum and the VEC curriculum. They found a close match between the two, and so it seemed reasonable that the VEC tests would have curriculum validity in the large district as well as in the two small districts. In fact, one of the reasons administrators in the large district wished to participate in the study was to gain access to and familiarity with the VEC testing program.

Differences between grade 4 and grade 5 tests. The VEC tests at grades 4 and 5 have different items, reflecting the differences in learning outcomes that students are expected to achieve in these grades. For example, students in grade 5 are expected to be able to work with decimals and the radius of a circle, whereas students in grade 4 are not.

Another difference between the fourth and fifth grade tests is in the number of items used to assess learning in various content areas. The number of items by content area at each grade level is shown in Table 3-8. Both mid-year and end-of-year VEC tests for a particular grade level covered the same content areas, included the same number of items in each control area, and had the same total number of items.

Differences between mid-year and end-of-year tests. Differences between the mid-year and end of-year tests at each grade level were slight. The tests covered the same content, but some of the items on the end-of-year test were of better quality than comparable items on the mid-year test. This was because teachers had had several opportunities to critically review the items

Table 3-8

Number of VEC Test Items by Content Area

Content Area	Number of Items	Number of Items
Numeration	5	5
Whole Numbers	30	20
Measurement	10	10
Geometry	10	10
Fractions	10	20
Decimals	--	5
Total	65	70

on the mid-year tests after the results from them had been reported, and to refine items that appeared weak. These changes, however, did not influence in any appreciable way the difficulty level of the items.

The CAT

The mathematics section of the California Achievement Test (CAT) was used as an additional measure of student achievement. This test was used because each of the three districts participating in the study administered the CAT on an annual basis for the purpose of program evaluation. It appeared that data produced from the districts' testing programs would also have value in the research project.

One complicating factor in using the CAT, however, was that the schedule of test administration in District 1 differed from the schedule followed in the small districts. In District 1, the CAT had been administered in October 1982, and the district had not planned to administer it again during the 1982-83 school year. In the small districts, the CAT had been administered in May 1982, and the district planned on administering it in May 1983.

Since the experiment required a post-treatment measure of achievement, administrators in District 1 agreed to administer the mathematics section of the CAT to fourth- and fifth-grade students in May 1983. This was in addition to the October 1982 administration. In the small districts, the May 1982 administration of the CAT served as the pre-treatment measure, and the May 1983 administration served as the posttest measure. Information regarding the use of the CAT tests in the experiment is summarized in Table 3-9. It should be noted that both District 1 and the two small districts administered the same levels of the CAT (14c for the fourth graders and 15c for fifth graders) for the posttest. District 1 and the small districts, however, administered different levels of the CAT and at different times for

Table 3-9

Administration of CAT Tests

Grade 4 Students

		Date of Administration	Level of Administration	No. of Items
Large District	Pretest	October 1982	14c	85
	Posttest	May 1983	14c	85
Small District	Pretest	May 1982	13c	85
	Posttest	May 1983	14c	85

Grade 5 Students

		Date of Administration	Level of Administration	No. of Items
Large District	Pretest	October 1982	15c	85
	Posttest	May 1982	15c	85
Small District	Pretest	May 1982	14c	85
	Posttest	May 1983	15c	85

the pretest. Since the fourth-grade students in the small districts were in the third-grade at the time of the pretest (May 1982), level 13c of the CAT was used. Since fourth-grade students were actually in the fourth grade at the time of the pretest (October 1982), level 14c was used. The same differences occurred in the fifth-grade sample for the same reasons.

Relationship Between the VEC and CAT Tests

The VEC and CAT tests for grades 4 and 5 both cover the broad content areas of numeration, whole numbers, measurement, geometry, and fractions. The CAT tests for grade 4 and 5 and the VEC test for grade 5 also have items on decimals. Decimals are not covered, however, on the VEC test for grade 4.

Within broad content areas, the VEC and CAT test differ in two significant aspects. First, the difficulty level of many items on the CAT is higher than the difficulty level of VEC test items keyed to similar learning objectives. For example, an item on the CAT test from grade 4 requires students to subtract five-digit numbers, whereas the VEC test for grade 4 only requires students to subtract numbers with four or fewer digits. This difference is due to the fundamentally different nature of the two tests. Because the CAT is a norm-referenced test, it contains items at varying levels of difficulty. The VEC tests are curriculum-referenced, however, and so they contain items only at the level of difficulty for the grade level being assessed.

The second difference between the CAT and VEC tests is the particular subskills that are assessed. In the area of whole numbers, for example, the CAT 15c test includes an item on prime numbers, whereas the VEC fifth grade test does not. On the other hand, the VEC tests assess measurement topics not c red in the CAT tests.

CHAPTER FOUR

RESULTS

This chapter presents statistical tests of the six hypotheses stated in Chapter 1. The hypotheses specify effects of the three treatments (principal involvement, regular inservice, and control) on the measured outcomes. The outcomes involve teacher instructional behavior, student mathematics achievement, and student off-task behavior.

The hypotheses were tested using a planned contrast procedure. A priori one-tailed _t_ tests were done comparing (a) the combined trained groups and the control group, and (b) the principal involvement group and regular inservice group. Adjusted post and delayed treatment means were used in these comparisons. The adjustments were made by analysis of covariance, using the corresponding pre mean as the covariate. The homogeneity of regression and homogeneity of variance assumptions of analysis of covariance were checked to determine the soundness of the adjustment procedure.

The analysis of covariance was first done by an SPSS program to determine the regression effect, overall treatment effect, and adjusted means. A separate aralysis of covariance was done, again using an SPSS program, to test the covariate-by-treatment interaction (that is, to test the homogeneity of regression assumption). The homogeneity of variance assumption was checked by the Bartlett-Box test using an SPSS program.

In several instances where the assumptions of analysis of covariance were not satisfied, a gain score analysis was done. The mean gain of each group was determined, and the difference between mean gains was tested using the _t_ test for independent means. This approach does not involve the use of regression estimation. Also, although the _t_ test involves a homogeneity of variance assumption, it is not seriously affected by a violation of this assumption. The effects of violation are minimal when the N's are equal or nearly equal, as in the contrast of the principal involvement and regular inservice treatments; and they are not serious when the N ratio is 2:1, as in the contrast of trained groups and the control group (Shavelson 1981, p. 423).

Some researchers have criticized the validity and reliability of gain scores as a measure of change (e.g., Linn and Slinde 1977). However, a recent analysis of gain scores (also known as difference scores) found that the traditional criticisms are unjustified: "When only two waves of data are available, the difference score is a natural and useful estimate of individual change" (Rogosa, Brandt, and Zimoski 1982, p. 744). In fact, these researchers present evidence that gain scores are a more valid indicator of change than the accepted method of residualized gain, which is based on regression analysis.

The null hypotheses were rejected at the .10 level (one-tailed test). This liberal alpha level was chosen to minimize the possibility of Type II errors, namely accepting null hypotheses as true when in fact they are false.

The chapter concludes with several supplementary analyses involving correlations between variables and a comparison of present results with those reported in Good and Grouws' fourth-grade experiment.

Hypothesis 1. Teachers who participate in the staff development program (with or without principal involvement) will use active teaching techniques more frequently than teachers who do not participate in the program.

Hypothesis 2. Teachers who participate in the staff development program with principal involvement will use active teaching techniques more frequently than teachers who participate in the programs without principal involvement.

The first step in testing these hypotheses was to determine the number of behaviors in Good and Grouws' instructional model that each teacher used for the criterion time period. Because 13 behaviors were measured, a teacher's score on this variable could vary between 0 and 13. The next step in the analysis was to compute descriptive statistics on this variable for each treatment group. The results of this analysis are shown in table 4-1.

The three treatment groups were similar in their instructional behavior at the time of the observed pre-treatment lesson. Teachers used an average of approximately five behaviors at the criterion level. Following the staff development program, the two trained groups made similar changes in their behavior; each group used approximately eight of the behaviors at the criterion level. The control group remained at its pre-treatment level.

At the time of the delayed observation, the principal involvement group continued to make gains in its use of the behaviors, whereas the regular inservice group regressed slightly. The control group continued to remain near its pre-treatment level.

Table 4-2 reports the tests of the two hypotheses at two times of observation (post and delayed). Hypothesis 1 was supported: at both times of observation the difference in number of techniques used to criterion level between the two trained groups and the control group was statistically significant. Hypothesis 2 was partially supported: the two trained groups were nearly identical in the post-treatment lesson, but in the delayed lesson the superiority of the principal involvement group approached statistical significance ($p = .07$).

The homogeneity of variances assumption was not satisfied for the pre-lesson measure of number of techniques used to criterion level. However, the violation of the assumption does not appear serious since the adjustment of the post-

TABLE 4-1

Number of Techniques Used to Criterion Level by Each Treatment Group

Time of Observation		Principal Involvement Group (N=18)	Regular Inservice Group (N=17)	Control Group (N=18)
Pre-Treatment	M	4.94	5.29	5.17
	(SD)	(1.59)	(2.71)	(1.69)
Post-Treatment	M	7.72	7.82	5.28
	(SD)	(2.27)	(2.51)	(2.05)
Delayed	M	8.33	7.47	5.33
	(SD)	(1.72)	(2.37)	(1.46)

TABLE 4-2

Analysis of Covariance Summary for
Number of Techniques Used to Criterion Level

POST TREATMENT MEANS

Source	df	MS	F	p
Regression (b = .31)	1	24.65	5.14	.03
Treatment	2	37.49	7.82	.001
Within	49	4.80		
Homogeneity of regression			.94	NS
Homogeneity of variances (pre)			2.94	.05
Homogeneity of variances (post)			.32	NS

Adjusted means

Principal involvement (PI)	7.79
Regular inservice (RI)	7.77
Control (C)	5.27

PI and RI vs C t = 3.95 p = .001
PR vs RI t = .03 p = NS

DELAYED TREATMENT MEANS

Source	df	MS	F	p
Regression (b = .26)	1	11.95	3.57	.07
Treatment	2	43.84	13.08	.001
Within	49	3.35		
Homogeneity of regression			.10	NS
Homogeneity of variances (pre)			2.94	.05
Homogeneity of variances (post)			2.05	NS

Adjusted Means

Principal involvement (PI)	8.38
Regular inservice (RI)	7.43
Control (C)	5.53

PI and RI vs C t = 4.86 p = .001
PI vs RI t = 1.52 p = .07

treatment and delayed means resulting from the analysis of covariance was very slight.

The results presented above can be further understood by analyzing teachers' use of each of the 13 instructional behaviors. The percentage of teachers in each treatment group who used each of the behaviors for the criterion time period is shown in table 4-3. Substantial differences between the two trained groups and the control group were found on the behaviors most emphasized in the staff development program: assigning and checking homework; mental computation; and development and controlled practice. Fewer teachers in the trained groups, however, were able to keep their transition behavior below two minutes.

The slight superiority of the principal involvement group on the delayed lesson (see table 4-1) is described more specifically in table 4-3. Interestingly, the principal involvement group and regular inservice groups are similar except for the seatwork phase of the delayed lesson. In that phase a higher percentage of teachers in the principal involvement group gave seatwork directions, monitored seatwork, and checked seatwork for the criterion time level.

The actual amount of time spent by each treatment group on the 13 instructional behaviors is shown in table 4-4. Substantial time differences between the trained groups and control group are evident on behaviors emphasized in the staff development program: assigning and checking homework; mental computation; development and controlled practice; and avoidance of unmonitored seatwork. The two trained groups differ from each other most in their use of developmental instruction and seatwork. In the post-treatment lesson, the principal involvement group spent less time than the regular inservice group on developmental instruction and substantially more time on unmonitored seatwork. In the delayed

TABLE 4-3
Percentage of Teachers in Each Treatment Group
Using Instructional Behaviors for Criterion Time Periods

Instructional Behaviors		Percentage of Teachers Within Guidelines		
		Principal Involvement Group (N=18)	Regular Inservice Group (N=17)	Control Group (N=18)
Transitions (2 minutes or less)	Pre	61%	53%	56%
	Post	33%	2[illegible]	67%
	Delayed	33%	35%	72%
Interruptions (no interruptions)	Pre	50%	59%	44%
	Post	56%	65%	61%
	Delayed	50%	71%	50%
Review Previous Work (1 minute or more)	Pre	28%	35%	33%
	Post	61%	35%	67%
	Delayed	50%	59%	22%
Check Prior Homework (at least 2 minutes)	Pre	6%	0%	11%
	Post	67%	59%	0%
	Delayed	61%	65%	11%
Assign Homework (1 second or more)	Pre	6%	18%	22%
	Post	78%	65%	6%
	Delayed	78%	71%	6%
Quiz (3 minutes or less)	Pre	67%	76%	56%
	Post	94%	76%	83%
	Delayed	89%	82%	83%
Mental Computation (at least 3 minutes)	Pre	6%	12%	22%
	Post	50%	65%	11%
	Delayed	50%	47%	11%
Development (at least 5 minutes)	Pre	50%	47%	44%
	Post	61%	82%	39%
	Delayed	89%	76%	50%
Controlled Practice (at least 1 minute)	Pre	33%	47%	33%
	Post	33%	53%	11%
	Delayed	33%	41%	6%
Directions for Seatwork (1 minute or less)	Pre	56%	65%	67%
	Post	50%	53%	44%
	Delayed	67%	47%	72%
Monitored Seatwork (15 minutes or less)	Pre	61%	41%	61%
	Post	100%	88%	67%
	Delayed	100%	76%	72%
Unmonitored Seatwork (2 minutes or less)	Pre	56%	53%	56%
	Post	67%	71%	50%
	Delayed	94%	65%	67%
Check Seatwork at End (at least 1 minute)	Pre	17%	24%	17%
	Post	33%	41%	28%
	Delayed	56%	29%	17%

TABLE 4-4
Length of Time (in Minutes) for Each Instructional Behavior by Treatment Group

Instructional Behaviors		Principal Involvement Group (N=18) M	SD	Regular Inservice Group (N=17) M	SD	Control Group (N=18) M	SD
Transitions	Pre	1.76	1.45	2.30	1.89	1.79	1.47
	Post	3.36	2.27	3.36	1.91	2.14	1.72
	Delayed	2.91	2.17	3.13	1.95	2.44	2.91
Interruptions	Pre	.59	.77	.74	1.48	1.31	2.00
	Post	.84	1.48	.19	.30	.43	.70
	Delayed	.52	.82	.51	1.22	.75	1.28
Review Previous Work	Pre	1.73	3.36	1.82	2.92	1.63	3.13
	Post	3.54	4.72	2.66	4.47	6.81	9.75
	Delayed	2.03	2.49	1.24	1.36	1.00	2.77
Check Prior Homework	Pre	.25	1.06	.15	.39	.64	1.86
	Post	3.46	3.53	4.62	3.78	.00	.00
	Delayed	2.61	2.47	3.59	4.01	.74	3.01
Assign Homework	Pre	.03	.12	.12	.28	.16	.33
	Post	.51	.43	.84	1.19	.03	.12
	Delayed	.78	.93	.44	.39	.06	.24
Quiz	Pre	2.58	3.96	1.83	3.92	4.33	5.83
	Post	.44	1.89	1.99	4.05	2.19	5.33
	Delayed	.75	2.38	1.92	5.53	2.77	8.06
Mental Computation	Pre	.35	1.48	.83	2.21	1.75	4.36
	Post	2.15	2.31	3.16	3.86	.50	1.61
	Delayed	2.13	2.42	2.08	2.92	.54	1.67
Development	Pre	6.97	8.75	6.54	8.27	5.57	7.77
	Post	7.80	8.01	12.09	8.42	6.34	8.11
	Delayed	15.75	8.72	8.93	7.13	13.03	14.40
Controlled Practice	Pre	1.43	3.06	3.87	5.07	2.24	3.93
	Post	2.19	3.89	2.27	2.82	.97	2.76
	Delayed	1.08	1.84	2.77	3.79	.53	2.24
Directions for Seatwork	Pre	1.51	1.78	1.05	1.11	1.12	1.33
	Post	1.55	1.43	1.55	2.31	1.53	1.61
	Delayed	1.01	.66	1.56	2.00	.86	.91

(cont'd next page)

TABLE 4-4 (cont'd)

Direct Instruction Behaviors		Principal Involvement Group		Regular Inservice Group		Control Group	
		M	SD	M	SD	M	SD
Monitored Seatwork	Pre	14.06	11.76	18.33	13.22	14.44	12.23
	Post	8.80	3.90	8.53	4.87	12.72	11.16
	Delayed	7.81	3.43	12.91	10.20	12.23	10.71
Unmonitored Seatwork	Pre	5.96	7.89	5.15	6.67	7.94	10.16
	Post	6.30	11.79	1.63	1.85	8.09	9.94
	Delayed	.55	.85	2.29	3.16	4.49	9.28
Check Seatwork at End	Pre	2.01	5.88	1.08	2.17	.90	2.93
	Post	1.07	1.39	.83	1.00	.96	1.58
	Delayed	1.69	1.67	.88	1.41	1.53	3.15
Length of Lesson	Pre	39.23	6.68	43.81	8.30	43.82	8.62
	Post	42.14	7.16	43.72	10.25	42.71	11.42
	Delayed	39.61	6.93	42.25	7.53	40.97	6.91

lesson, however, the pattern was just the reverse: the principal involvement group spent much more time than the regular inservice group on developmental instruction, and much less time in unmonitored seatwork.

The data were analyzed to determine whether training effects were consistent across districts. Data from the two small districts were combined, while the large district was treated separately. It was possible to combine data for the two small districts because the instructional behavior of teachers in these districts was similar. This judgement was based on inspection of Table A-1 (see Appendix J), which shows the length of time that teachers in each district used each of the 13 instructional behaviors.

Table A-2 in Appendix J presents district-level results for average number of instructional behaviors used for criterion time periods by each treatment group. The effects observed in the total sample were generally replicated within districts. Both trained groups outperformed the control group at each time of observation, and the principal involvement group outperformed the regular inservice group in the delayed lesson.

Table A-3 presents the percentage of teachers in each treatment group using each instructional behavior for the criterion time period in District 1 and in District 2-3. Table A-4 presents the same breakdown for the time duration of each behavior. The results shown in these tables indicate that training effects were generally consistent across the district.

Summary. The primary variable used to test hypotheses 1 and 2 was the number of instructional behaviors that teachers used to criterion level. Hypothesis 1 was strongly supported; the combined trained groups used more of the techniques than the control group in the post-treatment and delayed lesson. Hypothesis 2 was partially supported, since teachers in the principal involvement treatment used more of the behaviors than teachers in the regular inservice treatment in the delayed lessons. The two trained groups performed similarly in the post-treatment lesson.

Hypothesis 3. _Students of teachers who participate in the staff development program (with or without principal involvement) will earn higher scores on mathematics achievement tests than students of teachers who do not participate in the program._

Hypothesis 4. _Students of teachers who participate in the staff development program with principal involvement will earn higher scores on mathematics achievement tests than students of teachers who participate in the program without principal involvement._

These hypotheses were tested using two different measures of student mathematics achievement -- a curriculum-referenced test and a nationally standardized test.

Curriculum-referenced achievement test. This test was administered to students of all participating teachers before and again several months after the staff development program. Descriptive statistics for total scores of each treatment group are shown in table 4-5. The pretest achievement level of students whose teachers were in the principal involvement treatment was substantially lower than students whose teachers were in the other treatment conditions.

On the posttest administration the students in the principal involvement treatment were still achieving at a lower level, but the gap between their performance and the performance of students in the other treatments had narrowed. Analysis of mean pretest-posttest gains in table 4-5 indicates that students in

TABLE 4-5

Descriptive Statistics for Performance of Each Treatment Group on the Curriculum-Referenced Test (Total Score)

Variable	Principal Involvement Group (N=16)		Regular Inservice Group (N=16)		Control Group (N=18)	
	M	(SD)	M	(SD)	M	(SD)
Pre	33.69	(9.16)	39.61	(6.77)	38.75	(4.87)
Post	41.22	(8.64)	44.93	(9.07)	44.39	(5.40)
Gain	7.53		5.32		5.64	

the principal involvement treatment made the largest gain of the three treatment groups. The gains of students in the regular inservice group and control group were similar.

The analysis of covariance summary for treatment group differences on the posttest administration of the curriculum-referenced test is shown in table 4-6. The adjustment procedure was found to be invalid since both the homogeneity of regression assumption and homogeneity of variances assumption were violated. Therefore, an alternative statistical test of hypotheses 3 and 4 was made. The _t_ test for independent means was used to determine whether the gain scores of the contrast groups differed significantly from each other.

The results of the _t_ tests are shown in table 4-7. Although the mean gain of the trained groups is higher than that of the control group, the difference is not statistically significant. The difference between the mean gains of the principal involvement and regular inservice group almost reaches statistical significance ($p = .07$). Also, it should be noted that the principal involvement group outgained the regular inservice group by approximately the same margin that it outgained the control group.

Table A-5 in Appendix J presents descriptive statistics for the curriculum-referenced test for each district and grade level combination. The principal involvement group outgained the regular inservice group at both grade levels in District 1, and at the fourth-grade level in Districts 2 and 3. Also, the principal involvement group outgained the control group at both grade levels in Districts 2 and 3, and at the fifth-grade level in District 1.

Standardized achievement test. The California Achievement Test (CAT) was administered in each district before and after the staff development program. Standard score equivalents (NCE scores) for the computation subtest, concepts

TABLE 4-6

Analysis of Covariance Summary for Curriculum-Referenced Post Test (Total Score)

Source	df	MS	F	p
Regression (b = .81)	1	1847.73	84.83	.001
Treatment	2	9.74	.45	NS
Within	46	21.78		
Homogeneity of regression			3.54	.04
Homogeneity of variances (pre)			3.03	.05
Homogeneity of variances (post)			2.31	.10

Adjusted means

Principal involvement (PI)	44.46
Regular inservice (RI)	42.92
Control (C)	43.15

PI and RI vs C   t = .39   p = NS

PI vs RI   t = .88   p = NS

TABLE 4-7

Contrasts of Treatment Groups on Curriculum-Referenced Test Gain Scores

Contrast	N	M	SD	t	p
Trained Groups	32	6.43	5.02		
vs				.59	NS
Control Group	18	5.64	3.53		
Principal Involvement Group	16	7.53	5.43		
vs				1.42	.07
Regular Inservice Group	16	5.32	3.49		

subtest, and total test were used in the statistical analyses.

The performance of the three treatment groups on these measures is shown in table 4-8. The groups were approximately at the national average for the total CAT test on the pretest administration. All groups made gains, with the regular inservice group making the largest gain and with both trained groups outgaining the control group. The same pattern of relative gains appears in the two subtests. However, the differences in gains on the concepts subtest are slight. The greatest difference in gains is found on the computation subtest, on which the regular inservice group outgained the other two groups.

Table 4-9 shows the analysis of covariance summary for the CAT total score. Neither of the planned contrasts (trained groups versus control groups; principal involvement group versus regular inservice group) is statistically significant. However, the critical assumption of homogeneity of regression was not satisfied ($p = .03$).

Similar problems occurred in the analysis of covariance for the two subtests. A significant effect for homogeneity of regression ($p = .04$) and homogeneity of pretest variances ($p = .02$) was found on the computation subtest. The homogeneity of regression effect approached significance ($p = .12$) and the homogeneity of pretest and posttest effects were significant ($p = .05$) on the concepts subtest. Therefore, hypotheses 3 and 4 were tested using the gain scores of each treatment group.

Table 4-10 presents the results of t tests for the contrasts of interest. None of the hypotheses were supported. A noteworthy result is that the regular inservice group outperformed the principal involvement group by a substantial margin on the computation subtest. This difference in achievement gain is opposite to that which was predicted.

TABLE 4-8

Descriptive Statistics for Performance of Each Treatment Group on the California Achievement Test

Variable		Principal Involvement Group (N=18) M	(SD)	Regular Inservice Group (N=16) M	(SD)	Control Group (N=18) M	(SD)
Computation	Pre	48.80	(11.12)	49.52	(5.94)	50.37	(6.18)
	Post	56.65	(11.02)	61.09	(7.98)	58.07	(9.13)
	Gain	7.80		11.57		7.70	
Concept	Pre	50.53	(8.60)	53.61	(4.30)	51.57	(5.60)
	Post	55.23	(8.27)	58.50	(4.24)	55.37	(6.22)
	Gain	4.70		4.89		3.80	
Total	Pre	48.64	(8.75)	51.93	(5.16)	50.86	(6.15)
	Post	55.42	(9.28)	60.29	(5.65)	56.63	(7.63)
	Gain	6.78		8.36		5.77	

TABLE 4-9

Analysis of Covariance Summary for the California Achievement Test (Total Score)

Source	df	MS	F	p
Regression (b = .64)	1	1119.44	31.46	.001
Treatment	2	39.67	1.11	NS
Within	46	35.59		
Homogeneity of regression	2	120.95	3.81	.03
Homogeneity of variances (pre)			2.20	NS
Homogeneity of variances (post)			1.71	NS

Adjusted means

Principal involvement (PI)	56.73
Regular inservice (RI)	59.23
Control (C)	56.35

PI and RI vs C   t = .93   p = NS

PI vs RI   t = 1.16   p = NS

91

TABLE 4-10

Contrasts of Treatment Groups on California Achievement Test Gain Scores

TOTAL TEST GAIN SCORE

Contrast	N	M	SD	t	p
Trained Groups	32	7.58	5.74		
vs				1.03	NS
Control Group	18	5.71	7.00		
Principal Involvement Group	16	6.78	3.55		
vs				.81	NS
Regular Inservice Group	16	8.38	7.35		

COMPUTATION SUBTEST GAIN SCORE

Contrast	N	M	SD	t	p
Trained Groups	34	9.61	8.41		
vs				.80	NS
Control Group	18	7.70	8.09		
Principal Involvement Group	18	7.86	6.13		
vs				-1.35	.90
Regular Inservice Group	16	11.58	10.30		

No t tests were done for contrasts involving the concepts subtest as the criterion, since the difference between treatment groups in mean gain was slight (see table 4-8).

Tables A-6 and A-7 in Appendix J present descriptive statistics for CAT total score and computation score, respectively, for each district and grade level combination. The pattern of within-district results is generally consistent with the results observed in the total sample.

Summary. Hypotheses 3 and 4 were tested using both a curriculum-referenced test and a nationally standardized test (CAT) of mathematics achievement. With respect to hypothesis 3, the trained groups made greater gains than the control group on both tests, but the differences were slight and not statistically significant. With respect to hypothesis 4, the principal involvement group made a greater gain than the regular inservice group ($p = .07$) on the curriculum-referenced test, as predicted. The regular inservice group, however, made a greater gain than the principal involvement group on the computation subtest of the CAT. This result is opposite to the hypothesized difference. The two trained groups did not differ significantly from each other on the CAT total test or the CAT concepts test.

Hypothesis 5. Students of teachers who participate in the staff development program (with or without principal involvement) will exhibit less off-task behavior during instruction than students of teachers who do not participate in the program.

Hypothesis 6. Students of teachers who participate in the staff development program with principal involvement will exhibit less off-task behavior than students of teachers who participate in the program without principal involvement.

The number of off-task students was counted at five-minute intervals for each observed mathematics lesson. The average number of off-task students across the intervals was then computed. The results for each treatment group on this measure is shown in table 4-11. The average class size for each treatment group is also shown in this table. Since class size varied across groups, it was necessary to compute for each class the average number of off-task students as a percentage of the total number of students in the class. Table 4-11 presents descriptive statistics for this variable, which is labelled "percentage of off-task students."

The three treatment groups were similar to each other in percentage of off-task students in the pretreatment lesson. Both trained groups showed a slight reduction in off-task percentage on the post-treatment lesson, while the control group remained unchanged. On the delayed lesson, more than one-fourth of the average control class was off-task. The percentage of off-task students in the two trained groups was well below this level.

Table 4-12 reports the tests of hypotheses 5 and 6 at two times of observation. The assumptions underlying the adjustment procedure were satisfied with one exception. The treatment group variances differed significantly ($p = .006$) in the delayed lesson. Violation of this assumption, however, is minimal when the sample size for each group and treatment group variances on the pretest are similar (Elashoff 1969, p. 395).

Hypothesis 5 was supported since trained teachers had a lower percentage of off-task students than did control teachers in both the post lesson ($p = .10$) and the delayed lesson ($p = .05$). Hypothesis 6 was not supported since the principal involvement group and the regular inservice teachers did not differ

TABLE 4-11

Descriptive Statistics for Student Off-Task Behavior

Variable		Principal Involvement Group (N=18) M	(SD)	Regular Inservice Group (N=17) M	(SD)	Control Group (N=18) M	(SD)
Number of Off-Task Students	Pre	3.98	(2.85)	4.23	(2.51)	4.11	(2.56)
	Post	2.67	(2.33)	3.44)	(2.80)	3.76	(1.79)
	Delayed	3.67	(2.49)	3.39	(1.60)	5.02	(3.75)
Number of Students in Class	Pre	20.31	(6.38)	22.06	(3.19)	20.94	(4.63)
	Post	19.94	(6.42)	21.29	(3.80)	20.14	(4.29)
	Delayed	19.56	(5.63)	21.47	(3.83)	19.53	(7.03)
Percentage of Off-Task Students	Pre	18.54	(13.46)	18.74	(9.60)	19.40	(11.45)
	Post	13.73	(11.92)	16.18	(13.42)	19.40	(9.82)
	Delayed	18.87	(11.99)	16.05	(7.89)	24.33	(18.19)

TABLE 4-12

Analysis of Covariance Summary for Percentage of Off-Task Students

POST TREATMENT MEANS

Source	df	MS	F	p
Regression (b = .22)	1	.03	2.41	NS
Treatment	2	.01	1.00	NS
Within	49	.01		
Homogeneity of regression			.35	NS
Homogeneity of variances (pre)			.90	NS
Homogeneity of variances (post)			.78	NS
Adjusted means				
Principal involvement (PI)	.14			
Regular inservice (RI)	.16			
Control (C)	.19			
PI and RI vs C t = 1.27 p = .10				
PI vs RI t = .61 p = NS				
DELAYED TREATMENT MEANS				
Regression (b = .19)	1	.03	1.71	NS
Treatment	2	.03	1.65	NS
Within	48	.02		
Homogeneity of regression	2	.01	.66	NS
Homogeneity of variances (pre)			.80	NS
Homogeneity of variances (post)			5.16	.006
Adjusted means				
Principal involvement (PI)	18.94			
Regular inservice (RI)	16.08			
Control (C)	24.23			
PI and RI vs C t = 1.71 p = .05				
PI vs RI t = .64 p = NS				

significantly in percentage of off-task students in either lesson.

Table A-8 in Appendix J presents the same descriptive statistics for off-task behavior as in table 4-11, but separately for District 1 and Districts 2-3. It appears that there is a district-level effect for percentage of off-task students. In District 1 the two trained groups of teachers lowered their percentage of off-task students in the post lesson and maintained this gain in the delayed lesson. In Districts 2 and 3, the trained groups made slight gains in the post lesson but lost them in the delayed lesson. Thus, the support for hypothesis 5 in the total sample is largely a function of a training effect that occurred in District 1. Hypothesis 6 was not supported in either District 1 or Districts 2-3.

Summary. The variable used to test hypotheses 5 and 6 was the percentage of off-task students in the post and delayed lessons. Hypothesis 5, involving the contrast of the trained groups and the control group, was supported. Hypothesis 6, involving the contrast of the two trained groups, was not supported.

Supplementary Analyses

Treatment Implementation

The report of Good and Grouws' fourth-grade experiment provides data on teachers' implementation of their instructional model (Good and Grouws 1979, p. 358). The present study collected data on some of the same implementation variables.

Table 4-13 presents comparisons between the two studies on similarly measured variables. Good and Grouws' report includes only post-treatment data, so table 4-13 does not contain pre-treatment data from the present study.

Comparison of the control groups in the two studies indicates that their

TABLE 4-13

Comparison of Training Effects in Good and Grouws' Fourth-Grade Experiment and in the Present Study

Variable		G and G Treatment	Principal Involvement	Regular Inservice	G and G Control	Present Control
		Percentage of Teachers				
1. Conduct review	Post	91%	61%	53%	62%	67%
	Delayed		56%	65%		22%
2. Check homework	Post	79%	72%	88%	20%	0%
	Delayed		83%	71%		11%
3. Mental computation	Post	69%	61%	76%	6%	11%
	Delayed		56%	53%		11%
4. At least 5 minutes on development	Post	45%	61%	82%	51%	39%
	Delayed		89%	76%		50%
5. Controlled practice	Post	33%	39%	59%	20%	17%
	Delayed		33%	41%		6%
6. Hold accountable for seatwork	Post	59%	72%	59%	31%	44%
	Delayed		78%	53%		50%
7. Seatwork directions longer than 1 minute	Post	18%	50%	47%	23%	56%
	Delayed		33%	53%		28%
8. Assign homework	Post	66%	78%	65%	13%	6%
	Delayed		78%	71%		6%

level of use of Good and Grouws' instructional model was similar. The only noteworthy differences were in conducting reviews (delayed lesson) and in giving seatwork directions. On the latter variable, the actual mean length of time for giving directions in the control group post lesson in the present study was 1.53 minutes. Thus, this group was not much different in performance from the control group in Good and Grouws' study.

The performance of trained teachers was also similar across the two studies. The only noteworthy differences were on the variables of review, development, and seatwork directions. Trained teachers in Good and Grouws' study achieved better implementation of review and seatwork direction procedures, while teachers in the present study achieved better implementation of the time criterion for developmental instruction.

Correlation Between Variables

Table 4-14 shows the correlations between implementation and off-task variables. The correlations between the three measurements of implementation (i.e., the number of techniques used for the criterion time level) are just slightly positive. It appears that individual differences in implementation are relatively unstable from one time of observation to the next. The correlations between the three measurements of student off-task behavior are even lower, indicating that this classroom behavior was quite unstable across the three observed lessons.

An interesting result in table 4-14 is the consistent negative correlation between teachers' implementation of Good and Grouws' instructional model and student off-task behavior. It appears that implementation of the model was associated with a decrease in student off-task behavior, especially in the post lesson (r = -.53). This effect of Good and Grouws' staff development program

TABLE 4-14

Correlation Matrix of Implementation and Off-Task Variables[1]

	Implement Pre	Implement Post	Implement Delay	Off Task Pre	Off Task Post	Off Task Delay
Implement Pre	--	.27	.20	-.27	-.36	-.21
Implement Post		--	.41	-.15	-.53	-.12
Implement Delay			--	-.04	-.35	-.34
Off Task Pre				--	.22	.19
Off Task Post					--	.20
Off Task Delay						--

Note. -- N for all correlations is 53, except for Off-Task, Delay (N=52)

[1] The implementation variable is the number of techniques that a teacher used for the criterion time level. The off-task variable is the average percentage of off-task students in a lesson.

has not been previously investigated.

Table 4-15 reports the intercorrelations of the achievement test variables. As expected, there was a high correlation between the pre and post administrations of the curriculum-referenced test ($\underline{r}$ = .75) and between the pre and post administrations of the California Achievement Test ($\underline{r}$ = .65). The lower pre-post correlation for the latter test may be the result of the longer time lag between the two administrations and the fact that a different version of the test was administered on each testing occasion.

Although the two tests are correlated, the degree of association is not high ($\underline{r}$ = .54 for total post score). It appears that, to a large extent, the curriculum-referenced test and the nationally standardized test measure different aspects of mathematics achievement.

Analyses of the relationship between teachers' implementation of Good and Grouws' instructional model and student achievement gains are in progress. These analyses will be presented in a subsequent report.

TABLE 4-15

Correlation Matrix of Curriculum-Referenced Test Variables and California Achievement Test Variables

	Curric Pre	Curric Post	Comput Pre	Comput Post	Concepts Pre	Concepts Post	Total Pre	Total Post
Referenced Test Pre	--	.75	.27	.38	.40	.40	.41	.45
Referenced Test Post		--	.27	.49	.39	.50	.38	.54
Computation Subtest Pre			--	.57	.80	.60	.94	.58
Computation Subtest Post				--	.49	.79	.53	.95
Concepts Subtest Pre					--	.72	.95	.64
Concepts Subtest Post						--	.69	.93
CAT Total Pre							--	.65
CAT Total Post								--

N's for correlation coefficients vary between 49 and 51.

CHAPTER FIVE

Discussion

This chapter is organized into four sections. In the first section we discuss the meaning of observed differences in the effects of the principal involvement and regular inservice treatments. Do the results demonstrate that the involvement of a building principal can enhance the impact of a staff development program on his or her teachers? The second section of the chapter considers the significance of the findings concerning the relative effectiveness of the staff development treatments (ignoring the factor of principal involvement) and the control treatment.

In the third section we draw upon the study's results to reconsider the leadership functions that a principal, or other designated person, can perform in promoting the implementation of a staff development program of the type used in this study. Next, we discuss the implications of the study's results for school organization and school improvement. The chapter concludes with a set of recommendations for research on instructional leadership.

Principal Involvement in Teachers' Staff Development

The major purpose of this study was to determine the effects of directly involving elementary school principals in their teachers' staff development activities. Previous research suggests that this type of involvement characterizes principals of effective schools. The research, however, consists entirely of descriptive and correlational studies, which are weak designs for determining whether observed relationships between principal behavior and teacher and student performance are causal. An experimental design, in which principal behavior is manipulated by intervention, is necessary for determining whether principal behavior directly affects teacher instruction and student achievement. In the present study the manipulation was effected by having one group of principals participate in the teachers' staff development program (the principal involvement treatment) while withholding participation from another group of principals (the regular inservice treatment).

Effects on Teacher Instruction. Teachers in both treatment groups implemented Good and Grouws' instructional model to a nearly identical level in the post lesson. In the delayed lesson, however, teachers in the principal involvement treatment increased their implementation of the instructional model, while teachers in the regular inservice treatment regressed slightly. The difference between treatment groups on the implementation measure approached statistical significance ($p = .07$).

Why were differences in implementation not observed until the delayed lesson? Two plausible explanations are suggested by the way in which the experimental events were timed. The first explanation involves the timing of the principal involvement treatment. The principals in this treatment attended the second staff development session in mid-February and their first instructional leadership session at the end of February. They were given until April 1 to complete their first observation-and-conference. The post lesson observations, however, occurred between March 8 and March 18. Thus, the principals may have had too little time to exert influence on the teachers' post lesson instruction.

The situation at the time of the delayed lesson was different. All of the principals in the principal involvement treatment had completed their first observation-and-conference by April 1, and several had completed a second. Also, all of the principals attended a second instructional leadership meeting at the end of April. Since the delayed lesson observations occurred between May 3 and May 13, there was ample time for principals to exert their influence on teachers' implementation of the instructional model. Their influence had the desired impact, as reflected in the difference in implementation scores of teachers in the two trained groups.

The second explanation for the absence of a principal involvement effect until the delayed lesson is that the experiment commenced several months after the school year began. Both principals and teachers commented that homework and whole group instruction are difficult practices to implement once the pattern of instruction is set in September. For example, in Chapter 3 we referred to the principal who stated that teachers who had practiced some form of individualized instruction in the first semester were hard pressed to alter their practices in February simply on the basis of two brief training sessions in Good and Grouws' model. (In this respect it is interesting to note that the training in Good and Grouws' fourth-grade study occurred during the first month of the school year.)

Although principals generally had a positive reaction to the instructional model, they and their teachers needed time to think through procedures for shifting instructional practices. The homework policy in the instructional model challenged the participants, but most of them confronted it in the first staff development session, and they began implementation by the second session. In both trained groups the implementation level for homework assignment increased dramatically from the pre lesson to the post lesson, and stayed at the same level in the delayed lesson (see table 4-3).

The recommendation for whole group instruction in the model was much more difficult to implement. Individualized instruction was entrenched in many of the classrooms. Principals and teachers found it difficult to think of ways to handle individual differences among students -- especially low ability students -- in the context of whole group instruction. The teachers manual provided little information about ways to handle this problem. When the research staff visited principals and teachers in the principal involvement treatment in March and April, these groups reported some success in implementing whole group instruction. This gradual success is reflected in the observational data. Teachers in the principal involvement treatment spent about eight minutes on development in the post lesson and almost twice that amount of time in the delayed lesson (see table 4-4). According to the scoring system, teachers are not given credit for development unless the teacher is engaged in this instructional mode with the whole or nearly the whole class. Also, these teachers greatly reduced the amount of time they spent in seatwork (see table 4-4), which is another indicator of a shift away from individualized instruction.

The finding of a treatment group difference in the delayed lesson, but not in the post lesson, was unanticipated. Also, implementation was only measured at two points in time following training. Despite these problems, the post hoc interpretation offered above makes good sense. Principals need time to work with teachers to influence their instruction. Principals in the involvement treatment had adequate time by the delayed lesson, but not by the post lesson.

Effects on student achievement (curriculum-referenced test). Students in the principal involvement treatment made greater gains on the curriculum-referenced test of mathematics achievement than did students in the regular inservice treatment. The difference was statistically significant at the .07 level (one-tailed test). Several explanations for this result seem plausible.

The superiority of students in the principal involvement group on the curriculum-referenced test is explained most directly by their teachers' implementation of Good and Grouws' instructional model. In the delayed observation, teachers in the principal involvement group spent approximately 16 minutes of the lesson on developmental instruction, whereas teachers in the regular inservice group spent only 9 minutes on this activity (see Table 4-4). In other words, teachers who had involved principals spent more time engaged in instruction of new material. Since the curriculum-referenced test was designed specifically to measure amount of curricuum covered and mastered, it seems reasonable that students whose teachers spent more time on developmental instruction would perform better on the test than students whose teachers spend less time on developmental instruction.

This explanation of treatment group differences on the curriculum-referenced test suggests that the involved principals affected student achievement by promoting teacher implementation of critical aspects of Good and Grouws' instructional model. Teachers whose principals were not involved in the staff development program did not have a person in authority encouraging them to implement -- and to continue implementing -- these critical aspects.

Another explanation of the curriculum-referenced test results is that the involved principals exerted influence by encouraging teachers to improve student performance on the post-administration of this test. The involved principals in the two small districts had worked with this test for several years and so were quite familiar with its format and meaning. The involved principals in the large district had not previously experienced this test. As mentioned in Chapter 3, these principals were given information about the test in their initial instructional leadership session in February. Peer pressure might have operated to influence the involved principals to work with teachers so that their school would produce favorable gains on this test.

The involved principals were further sensitized to the curriculum-referenced test when the results of the pretest were given in late February to each participating teacher for his or her class and to all principals for each grade level in their schools. Also, the involved principals knew that the curriculum-referenced test would be re-administered in May to determine student achievement gains. All of these sensitization procedures could have influenced the involved principals to help teachers enhance student performance on math concepts and procedures measured by the curriculum-referenced test.

It can be argued that principals in the other two treatment groups were similarly sensitized because they too received grade-by-grade results for the pretest, and each teacher under their supervision received pretest results for his or her classroom. No doubt these procedures resulted in some sensitization, but two important differences between these groups of principals and the involved principals should be noted. First, only the involved principals attended the training sessions in Good and Grouws' program, at which time they learned that student mathematics achievement could be affected by having teachers make changes in their instructional behavior. Although this point seems obvious, participants in the sessions had not known previously about the type of correlational process-product research on which Good and Grouws' program and similar programs are based.

The second difference between the involved principals and other principals is that only the former group had the opportunity to meet together (in the first leadership session) to discuss the VEC test results and their meaning in relation to Good and Grouws' instructional model. Thus, type and frequency of sensitization were substantially different for involved principals than for principals in the regular inservice and control treatments.

In summary, the above discussion suggests that the involved principals could have influenced the results on the curriculum-referenced test in two ways. One type of influence involves helping teachers to implement critical aspects of Good and Grouws' instructional model. The other type of influence involves encouraging teachers to help students improve their achievement test performance. These two types are not mutually exclusive. There is evidence that the involved principals could have used one or both types of influence.

The treatment group differences on the curriculum-referenced test can be attributed to other factors than principal influence. It must be noted that the pretest achievement of students in the principal involvement treatment was substantially lower than that of students in the regular inservice treatment (see table 4-5). The involved principals undoubtedly were aware that their schools had lower-achieving students than other schools in the district. This disparity may have stimulated the involved principals to devote extra effort working with teachers to reduce the disparity.

Still another explanation for the treatment effect involving the curriculum-referenced test is that it is an artifact of the statistical procedure used to determine the presence of the effect. As shown in table 4-5, students in the principal involvement treatment made a 7.5 point mean gain on this test, starting from a pretest mean score of 33.7. The gain of the other two treatment groups was not quite so large: they gained approximately 5.5 points, starting from a mean of approximately 39 points. With students starting from a lower initial achievement level, teachers and principals in the principal involvement treatment may have found it easier to stimulate improved achievement than was the case for teachers and principals working with students who started at a higher initial achievement level. This argument, however, can be reversed. It may actually be more difficult to help lower-achieving students make appreciable learning gains because they lack motivation and a belief that they can do so. Without evidence to the contrary, this argument seems as plausible as its converse.

The different levels of pretest achievement would further complicate interpretation of treatment effects if the posttest measure of achievement had a low ceiling. The results shown in table 4-5, however, indicate that the two treatment groups with higher initial achievement scored approximately 45 points on the posttest. Since the posttests for grades 4 and 5 contained 65 and 70 items, respectively, it seems unlikely that achievement gains for these two groups were depressed by a ceiling effect.

An interaction between pretest achievement level and achievement gain cannot be ruled out as an explanation of the observed difference in gain between students in the principal involvement treatment and students in the regular inservice treatment. Thus, this explanation stands as a plausible alternative to the explanation that the observed differences occurred because principal involvement augments the effects of a staff development program.

Effects on student achievement (standardized test). Students whose teachers were in the regular inservice treatment made greater gains on the computation scale of the nationally standardized test than did students whose teachers were in the principal involvement treatment. How can this result be explained, especially in view of the fact that the same group of students had lower performance on the curriculum-referenced test?

The explanation we favor is that participating teachers emphasized aspects of the mathematics curriculum to which they were sensitized in their particular treatment condition. Support for this explanation comes from two sources.

First, the results shown in table 4-15 indicate that the curriculum-referenced test and the computation test measured substantially different aspects of the mathematics curriculum. The correlation between the pretests was only .27, and the correlation between the posttests was .49. The percentage of shared variance was much less than the percentage of independent variance. If teachers in the two treatments were emphasizing different aspects of the mathematics curriculum, this would likely result in differential performance of their students on the two tests.

The instructional model presented in the staff development program was not curriculum-specific. An examination of the teachers manual reveals that it refers to instruction in math "concepts," "skills," and "problem-solving." Almost without exception, however, the examples refer to computational skills. In the main part of the manual, 18 of the 19 examples involved the basic computational skills of addition, subtraction, multiplication, and division. In the appendix, 42 of the 46 examples involved the same computational skills.

This analysis of the manual suggests that it would be easy for teachers to get the impression that the purpose of the instructional model was to increase students' proficiency in basic computational skills. Also, the substantive content of the two staff development sessions was focussed almost entirely on the instructional model rather than on curriculum issues. In the absence of other input, teachers in the regular inservice treatment might well have concentrated on improving students' computational skills rather than on improving their performance across the varied topics covered in the fourth- and fifth-grade mathematics curriculum.

Evidence in support of this argument comes from differential treatment group implementation of the instructional model. We have already discussed the fact that the regular inservice teachers spent less time on developmental instruction in the delayed observation lesson. Also, these teachers spent substantially more time on seatwork than did teachers in the principal involvement treatment (see table 4-4). The former group spent a total of 17 minutes on seatwork directions and seatwork, whereas the latter group spent a total of 9 minutes on the same activities.

Seatwork is typically used for practice of computational skills. (The observers noted virtually no instruction or practice involving mathematical problem-solving.) Since the regular inservice group spent more time on this activity, it is reasonable to expect that their students would score higher on a measure of computational skill.

This explanation suggests that the regular inservice teachers were trying to implement the instructional model, but primarily those aspects that are oriented to curriculum objectives involving computational skills. The teachers with involved principals were also trying to implement the model, but with more emphasis on the total range of curriculum objectives in mathematics.

Effects on students' off-task behavior. The classes of teachers in the two treatment groups did not differ significantly from each other in percentage of off-task students. A non-significant difference was found in both the post and delayed lessons (see table 4-12).

Even though no treatment effect occurred for this variable, the finding is interesting because it supports the interpretation that we have advanced for the effects mentioned above. Reduction of student off-task behavior is not a major goal of Good and Grouws' instructional model. It is only discussed in the section on seatwork in the teachers manual (6 pages of the 47 page manual). There was some discussion of off-task behavior in the staff development sessions, but it was at best a minor focus. Teachers expressed more concern about students' motivation for learning mathematics and for the various activities of the instructional model than they did about whether students were on task from moment to moment. Also, there was no explicit attention to off-task reduction as a goal in the instructional leadership sessions for principals in the principal involvement sessions.

The fact that off-task reduction was not a highlighted goal of the instructional model could explain why principals and teachers in the two treatments did not differ in reduction of off-task behavior. The goals that were differentially highlighted for the involved principals -- instructional model implementation and performance on the curriculum-referenced tests -- were those for which treatment effects were observed.

Another explanation of the results for off-task behavior is that the observed percentages are at or near the lower limits for this behavior without special intervention being focussed on it. McIntyre, Copenhaver, Byrd, and Norris (1983) recently determined off-task rates for samples of third-, fifth-, and seventh-grade mathematics lessons. The off-task rate was 23 percent, 25 percent, and 27 percent for the three grade levels, respectively. Fisher and his colleagues (1980) observed similar percentages of off-task elementary students in the Beginning Teacher Evaluation Study.

In the present study, however, the treatment groups were below this level in the prelesson (less than 20 percent) and one treatment group achieved a low off-task rate of 13.73 percent after training (see table 14-11). Only one measurement point (the control group's delayed lesson) approximated the off-task rate found in other studies. These comparisons suggest that instructional management in the present sample of classes was better than average. If this was the case, teachers in the trained groups may not have felt a particular need to focus on improving students' on-task rate.

Magnitude of observed effects. The statisticaly significant differences between the principal involvement group and regular inservice group were of small magnitude. Several features of the experimental situation may explain this result. First, the pretest performance of the participating students on the nationally standardized test (see table 4-8) indicates that they were approximately at the fiftieth percentile. In contrast, the mean pretest performance of students in Good and Grouws' fourth-grade study was approximately at the twenty-eighth percentile on a nationally normed test. It may be that principals in the principal involvement treatment did not feel the same intensity of need for improvement of student achievement that participants in Good and Grouws' study felt. If this was the case, they would not have attempted to work with teachers to bring about the substantial gains in student mathematics achievement found in Good and Grouws' study.

Another explanatory factor is the time and duration of the experiment. The last measurements of instructional model implementation and student achievement occurred approximately three months after the staff development session and first instructional leadership session. Differences between the two treatments might have been accentuated if subsequent measurements were made after the involved principals had more time to work with their teachers. Also, as we indicated above, the fact that the training sessions occurred in February, well after the start of the school year, limited what the involved principals could accomplish.

The intensity of the instructional leadership training may have been another factor limiting treatment group differences. The involved principals met together for just two sessions. Participation in additional sessions might have served to increase their leadership skills and commitment to teachers' implementation of the model. Also, the involved principals observed most of their participating teachers a total of two times. Additional observations might have helped these teachers fine-tune their use of the instructional model, resulting in a higher implementation level relative to the teachers in the regular inservice treatment.

Finally, another factor that may have limited treatment group differences is the design of the staff development sessions. It was not feasible to conduct separate sessions for teachers in the two treatment groups. Therefore, the sessions included teachers in the regular inservice treatment, teachers in the principal involvement treatment, and principals in the principal involvement treatment. The presence of principals from other schools at the sessions may have led the regular inservice teachers to perceive more need to implement the instructional model than they would have perceived had they attended the sessions only with other teachers.

We do not know the extent to which initial student ability level and the timing and intensity of training affects the impact of principal involvement on teachers' staff development. We also do not know whether certain features of the experimental design, such as including both treatment groups in the same training sessions, obscured treatment effects. Further research is needed to determine whether principals can be involved in teachers' staff development so that the effects achieve practical significance in addition to the statistical significance found in this study.

Nonrandom assignment of principals to treatments. The sampling unit used in the experimental design was the individual school. Thus, the principal and all of the participating teachers in each school were assigned to the same treatment group. Because of restrictions imposed by the school districts, it was not possible to randomly assign the schools to treatment groups.

Because of nonrandom assignment, the possibility exists that some or all of the observed group differences on the dependent variables reflect pre-existing group differences rather than treatment effects. As we indicated in Chapter 2, principals were assigned to the principal involvement treatment because of their interest in the study, their reputation as effective leaders, and/or their current workload. If interest in instruction and effective leadership were indeed more prevalent among the involved principals, these characteristics could account for the observed effects favoring the teachers and students of these principals.

The first step in testing the validity of this explanation is to realize that these were characteristics ascribed to the involved principals by central office administrators. Some of the data collected in the experiment can be used to check the validity of these ascribed characteristics. For example, table 4-1 shows the number of instructional techniques used by teachers in each treatment group before the experiment began. The results indicate very little difference between treatment groups. (In fact, the teachers supervised by the involved principals had the lowest initial level of criterion use of the techniques.) Similarly, table 4-11 indicates very little difference between treatment groups in percentage of off-task students at the outset of the experiment. These data suggest that the involved principals were not initially more effective instructional leaders than principals in the other two treatment groups.

The issue of treatment group similarity is complicated somewhat by the pretest data on student mathematics achievement. The pretest data for the curriculum-referenced test, shown in table 4-5, clearly indicate that the involved principals worked in buildings with lower-achieving students. (The data for the nationally standardized test, shown in table 4-8, also show the same pattern of difference, but the magnitude of the differences is slight.) It can be argued that these principals were instructionally effective because their teachers' classroom performance was similar to that of principals of students who were achieving initially at a higher level on the curriculum-referenced test.

This argument would be given weight if it could be shown that teachers' classroom performance is affected by student achievement level. The data in this study provide only weak evidence that such a relationship exists. The correlation between student achievement on the curriculum-referenced pretest and pretreatment teacher implementation of the instructional model is only .16. The correlation between the former variable and pretreatment percentage of student off-task behavior is only -.22. It seems that teacher and student behavior in the classroom is largely independent of initial student achievement level. Therefore, if the involved principals were more effective instructional leaders, they would not have been hampered in affecting their teachers' classroom performance by a lower-achieving student population.

This analysis suggests, then, that the principals in the three treatment groups did not differ initially in instructional leadership effectiveness, even though they were assigned nonrandomly to treatment groups. Furthermore, to the extent that initial differences existed, they were controlled for their effect on the dependent variables by the use of analysis of covariance or gain score analysis. Nonetheless, pre-existing differences between principals in the two trained groups cannot be ruled out definitely as an alternative explanation of observed group differences on the dependent variables. Further research using a better controlled design will be necessary to determine more definitely the true impact of instructional leadership manipulation on teacher and student performance.

Concluding comment. Even if other factors operated to affect the observed treatment group differences, the weight of the evidence suggests that manipulation of principal behavior did produce positive effects on teachers and students in this experiment. Although it is just one experiment, its results are consistent with the many descriptive and correlational studies that preceded it. Many of these studies (see table 2-6) identified principals' involvement in teachers' staff development and classroom observation as characteristic of effective principals.

Another characteristic of effective principals identified in these studies is the effort they direct toward the goal of high academic achievement for students. Similarly, the involved principals in this study were given clear academic goals and asked to achieve them. They were asked to help teachers implement an academically-oriented instructional model and to improve student achievement as measured by a curriculum-referenced test. These are goals that they accomplished to a degree. Conversely, the involved principals did not perform differently than the non-involved principals on goals that were not emphasized in the principal involvement treatment. These goals involved reduction in off-task behavior and computational skill.

Present and previous research suggest then, that the critical determinants of principal effects on teacher instructional performance and student academic achievement are (1) the clear articulation of these outcomes by the principal as goals to be achieved and (2) personal effort directed toward these goals.

Effectiveness of the Staff Development Program

The major objective of this study was to determine the effects produced by involving school principals in a staff development program for teachers. A secondary objective was to determine the effectiveness of the staff development program itself, irrespective of the principal involvement factor. The inherent effectiveness of the program was of interest to us because it might affect principal involvement, which was the variable of primary interest. If the program was difficult to implement and did not produce benefits for students, principals in the principal involvement treatment might choose to become disengaged from it irrespective of their agreement to work on improving their use of instructional leadership skills. The effectiveness of the program was determined by comparing the combined performance of both trained groups (i.e., principal involvement and regular inservice) with that of the control group. This comparison constitutes a replication of the design of Good and Grouws' fourth-grade experiment.

The results of these comparisons generally favored the trained groups. Their implementation of the instructional model was significantly higher ($p = .001$) in both the post lesson and delayed lesson. The results are very similar to the results obtained in Good and Grouws' fourth-grade study, as shown by the implementation levels for specific behaviors in table 4-13.

The comparisons involving student achievement measures were not statistically significant. As we discussed above, however, each trained group may have emphasized a different aspect of student achievement. If true, comparisons of the combined trained groups may be misleading. For each achievement test, the control group should be compared with the trained group that emphasized the aspect of achievement measured by that test. When this is done, the results favor the trained groups. The principal involvement group outgained the control group on the curriculum-referenced test by approximately the same level (1.89 points) that it outgained the regular inservice group (2.21 points). Similarly, the regular inservice group outgained the control group on the CAT computation subtest by approximately the same level (3.87 points) that it outgained the principal involvement group (3.77 points). While these student achievement effects are not as impressive as those found in Good and Grouws' study, they are in the same direction.

The combined treatment groups achieved a lower percentage of off-task students in both the post lesson ($p = .10$) and delayed lesson ($p = .05$). Although this training outcome was not reported in Good and Grouws' study, it is a reasonable outcome to expect from this type of staff development program. Correlational results shown in table 4-14 indicate that there is a moderate negative relationship between implementation of the instructional model and student off-task behavior.

The results summarized above raise several questions. Why did the staff development program work? Why did it work for teachers whose principals were not involved in the program (that is, the regular inservice group)? And why did it not work for all teachers? We shall discuss each of these questions in turn.

Why did the program work? Good and Grouws' *Active Mathematics Teaching* worked, we believe, for two reasons. First, the design of the program includes a number of features that research has found lead to implementation of the program's content. (Reviews of research on effective inservice practices are available in Gall and colleagues 1982, and Sparks 1983). For example, the objectives are clearly articulated in the form of a daily and weekly lesson plan; the sessions provide opportunity for expression of teachers' personal and technical concerns; there is opportunity for practice between sessions; the instructional model is generally consistent with teachers' belief systems about mathematics instruction; and the program generally requires "fine-tuning" of existing skills rather than acquisition of a new skill repertoire. These features of the program were probably instrumental in promoting the observed changes in teachers' instructional behavior.

The effects on student achievement and off-task behavior can be accounted for by another aspect of the program's design. The instructional model taught in the program originated in correlational research on effective mathematics instruction. (This research was discussed in Chapter 2). Thus, each element of the instructional model was included because of evidence that it promotes student achievement and on-task behavior. Indeed, as teachers gained experience with the model, they saw that it was having positive effects on students, and this perception encouraged further implementation of the model. Teachers reported in the training and end-of-experiment sessions that students responded especially well to the predictable daily lesson plan, to mental computation, to homework, and to the fact that their work was checked.

Why did the program work for the regular inservice group? The principals of teachers in the regular inservice group did not have the same instructional leadership training and support provided to the involved principals. Yet the results indicated that Good and Grouws' program worked almost as well with the regular inservice group as with the principal involvement group. As we discussed earlier in the chapter, statistically significant differences between the two treatment groups were generally of small magnitude.

One explanation for this result is that principals in the regular inservice group also engaged in instructional leadership behaviors. All of them were experienced principals in their districts. Principals in the large district were using Madeline Hunter's Instructional Theory Into Practice model for supervising teachers' instructional behavior. Principals in the two small districts were using the Valley Education Consortium's model for assessing student achievement and for linking assessment to curriculum.

The district-sponsored models, then, focussed principals' attention on academic achievement goals, classroom instruction, and teacher supervision. Thus, the difference in instructional leadership between the two trained treatments probably was a matter of degree rather than of present-absent.

It is important to note in the preceding analysis that the initiative for principals' instructional leadership came from the school district central office. Participation of principals in the Valley Education Consortium or in Madeline Hunter's program represented district initiatives. This fact, together with other features of the experiment, suggests to us that _the leadership behavior of district administrators was an important determinant of the effects observed in this study_. Indeed, the experiment could not have been conducted without the involvement of district administrators. The assistant superintendent in each district was contacted initially to determine the possibility of having district schools participate in the experiment. These contacts were made by a member of the research staff who had a reputation for state-level leadership in school improvement, and who was a central figure in the development of the Valley Education Consortium. His authority as an expert and person of influence was doubtless instrumental in persuading the assistant superintendents to take a close look at this experimental project.

The next step in the process was to meet with the principals in each district to secure their support. These meetings involved all of the principals who eventually participated in all three treatment groups. Thus, principals in the regular inservice group were subjected to the same elements of persuasion as principals in the principal involvement group. The principals then had the opportunity to meet with their teachers before making a final commitment to the project.

This pattern of recruitment is not unusual in field research in education. For example, Good and Grouws (1981) recruited participants for their junior high school experiment using this procedure:

> The investigators met with school administrators during the summer in order to explain the project and to obtain permission to do the study. School administrators explained the project to principals, who in turn described the project to classroom teachers (p. 40).

The fact that this is a typical recruitment procedure makes it easy to overlook the role of school district administrators in research on classroom instruction. This oversight is also augmented by the fact that district administrators usually fade into the background once a project of this type gets underway.

It appears, then, that principals in the regular inservice group were subject to persuasion by school district administrators and by the research staff to participate in the experiment. Also, after the experiment was approved, these principals knew that it had district support and sanction. Furthermore, teachers of these principals were given cues that the experiment had the support of district administrators behind it. The primary trainer in two of the districts was a central office supervisor; in the third district this person was a longtime teacher with a reputation as a district leader. Both principals and teachers in the regular inservice group knew that data on instruction and behavior were being collected and would be reported in the form of summary statistics to district administrators.

This analysis suggests that leadership behavior by district administrators had an influence on the effectiveness of the regular inservice group. It is difficult to imagine that a building principal and his or her teachers could initiate or sustain a staff development program of this type without the involvement of district administrators.

The investigation of instructional leadership by district administrators was not an objective of this experiment. The results, however, suggest that it is a distal, but nonetheless important determinant of teachers' instructional effectiveness. We suggest as a hypothesis for further investigation that district administrators and building principals do not produce leadership effects independently, but rather they work interactively to produce them. Surprisingly, the research on school and teacher effectiveness has focused almost exclusively on the principal's instructional leadership. The analysis of present results suggests that a broader framework that includes both district and principal leadership behavior is needed. In the next section we propose some initial elements of this framework by identifying leadership functions at different levels of administration that may affect school and teacher effectiveness.

Why was the program not more effective? Although the trained groups were more effective than the control group on most outcome variables, the training effects were not exceptionally strong. For example, the trained groups used an average of eight elements of the instructional model to the criterion level (see table 4-1) but the complete model includes thirteen elements. Approximately 50 percent of the trained teachers did not do mental computation exercises for at least three minutes in the posttraining lessons; about 60 percent did not do controlled practice for at least one minute; about 25 percent did not assign homework; and about 60 percent did not spend at least one minute checking seatwork (see table 4-3). Almost 20 percent of students were off-task during mathematics instruction, even after their teachers had participated in the staff development program (see table 4-11).

Some of these implementation failures can be explained by individual differences among teachers. In a review of research on this problem, Mohlman, Coladarci, and Gage (1982) found that teachers who perceive an instructional practice as easily used and congruent with their instructional approach are more likely to implement the practice than teachers who perceive it as difficult and incongruent. Similarly, Ebmeier and Good (1979) found that implementation of instructional practices in the fourth grade was partly a function of teacher characteristics.

The finding that teacher characteristics affect program implementation seems very reasonable. Further analysis, however, suggests that the relationship between these factors may be mediated by administrative policy. If teachers perceive that they have a choice about implementing the practice, their personal characteristics will determine their implementation behavior. If administrative constraints are placed on choice, however, personal characteristics are much less relevant.

Our experience with the training sessions in the present experiment suggests that there was a delicate balance between administrative constraints and teacher choice. The involved principals occasionally made recommendations, but they were careful to avoid directly requiring implementation of the model. The principals did make offers to help in implementation, the most notable example being the sending of letters over their signature explaining that teachers would be instituting homework. Several principals also assured teachers that they would help teachers handle parents who might object to having their children do homework. (This possibility did not materialize, to our knowledge).

For the most part, consensus decision-making seemed to operate. Teachers generally agreed to try the model, but if a teacher vigorously objected to implementation, no coercive behavior was used. In fact, only one teacher objected to the point that he wished not even to have his classroom observed at the time of the post and delayed lesson. He had the principal's support in this decision.

This analysis is consistent with other analyses that have been made of school organization and administration. For example, the concepts of school cellular organization (Lortie 1975) and loose coupling (Weick 1976) acknowledge the relatively weak control of administrators over classroom instruction. Their relevance to the present experiment is that they place a limit on the extent to which principals can exert control over teachers' implementation of practices contained in a staff development program. The leadership interventions used in the principal involvement treatment (training session attendance; classroom observation and conference) may have a low upper limit of effectiveness. To increase implementation beyond a certain limit, it is probably necessary to institute mechanisms that place constraints over teacher choice. Otherwise, teacher characteristics will operate and determine whether the instructional practice is implemented. Although these characteristics may be modifiable, the process is likely to be time-consuming and expensive.

Nonrandom assignment of principals to treatments. In discussing the differential effectiveness of the principal involvement and regular inservice treatments, we noted the problems of interpretation created by the nonrandom assignment of principals to treatment. The same problem affects comparisons involving the trained groups and the control groups. Because of nonrandom assignment, there was a systematic bias in the type of principal selected for the control group. As stated in Chapter 2, control group principals, even before this experiment, participated less often in teachers' inservice sessions than did principals in the two treatments involving training. Also, control group principals in District 1 less often used Madeline Hunter's Instructional Theory Into Practice as their basis for teacher supervision.

Despite the pre-existing treatment group differences on these dimensions, the control group was very similar to the trained groups on pre-treatment instructional variables relating to teacher behavior, student behavior, and student achievement. Pre-existing differences within and between treatment groups were controlled statistically for their effect on the dependent variables. Also, treatment group differences on implementation variables were very similar to differences found in Good and Grouws' fourth-grade experiment (see table 4-13), where random assignment was present.

Because of these factors it seems very unlikely that the superiority of the trained groups relative to the control group on the dependent variables was due to initial superiority of trained group principals in instructional leadership. The observed treatment group differences on the dependent variables can be attributed with a high level of certainty to the staff development program that was used in the principal involvement and regular inservice treatments.

The pre-existing differences between treatment group principals does pose problems for generalizability of findings, however. There is little assurance that selection of any group of principals for instructional leadership training of the type operationalized in the principal involvement treatment would produce similar effects to those observed here. Rather, it seems likely that instructional leadership training would only be selectd and implemented by principals who were predisposed to be instructional leaders.

Individual differences between principals in attitude toward instructional leadership should come as no surprise to educators. Research on staff development for teachers has documented repeatedly that some teachers implement training more than other teachers as a function of personal characteristics (Mohlman, Coladarci, and Gage 1982). Principals should be no different in this respect. The present experiment was not designed to explore the effects of individual differences in principal characteristics on treatment implementation, but this could be done in future research. Hall, Rutherford, Hord, and Huling (1984), for example, have identified characteristics of principals that look like promising predictors of implementation of instructional innovations. These characteristics could be included as variables in research on principal involvement in teachers' staff development.

Instructional Leadership Functions

The present experiment was designed to focus on one leadership role - the elementary school principalship - and the leadership functions of this role that might facilitate staff development programs for teachers. We decided on this focus because of recent research interest in building principals and their influence on school effectiveness. The results of the experiment, discussed above, suggest that other educator roles and other leadership functions may be equally important in understanding school effectiveness.

Gersten and his colleagues (1982) made a similar observation. They argued that it is more productive to understand instructional leadership by analyzing it into functions, or "critical behaviors," than by focusing on a single leadership role like the principalship. Following their argument, we did a retrospective analysis, based on the present experiment, of the leadership functions that facilitate the implementation and effectiveness of a staff development program of the type developed by Good and Grouws.

Leadership functions. The analysis yielded this set of leadership functions for facilitating staff development (and, more generally, school improvement):

1. Priority-setting function. It requires an act of leadership to propose priorities for school improvement. Educators in each of the three participating districts had identified improvement of mathematics instruction and staff development as priorities for their districts. The salience of these priorities was used by district administrators as the rationale for involving their staff in the experiment.

2. Resource acquisition function. Leaders in the two small districts chose to involve their schools in the Valley Education Consortium over a period of several years because they viewed it as a continuing resource for school improvement. The leaders in the large district made a major commitment to Instructional Theory Into Practice over a period of 10 years for the same reason. Leaders in each district viewed the present experiment as an opportunity to strengthen existing resources at virtually no cost. (In fact, honoraria and no-cost training were provided to participating schools.)

3. Monitoring functions. These leadership functions include seeing that all elements of the school improvement process occur as intended. Timelines need to be established; people need to be contacted; meetings must be called; resources need to be secured; trouble-shooting must occur as problems arise. Unless someone oversees this process, it can break down.

4. Compliance function. This is one of the most critical leadership functions, but it is not well articulated in the recent literature on instructional leadership. There appears to be an assumption that teachers will implement an instructional practice if they are trained well, if their concerns are

addressed, and if they are involved in decision-making. Our observations suggest that leaders can and do supplement these strategies by various behaviors that are intended to bring about compliance with the recommended instructional practice. Examples of these behaviors are use of rewards (including encouragement and emotional support) for compliance and sanctions for noncompliance; persuasive argument, a quality that one researcher has labelled "communicative potency" (Robert Mattson, personal communication); appeal to a higher authority on the need for compliance; expressions of enthusiasm for the merits of the instructional practice; and direct orders to comply (not observed in the present experiment).

5. Training function. A leader may do all or some of the training called for in a particular staff development program, including training to develop teachers' instructional skills. Training can involve such techniques as holding meetings, modelling the practice, and conducting classroom observations and conferences.

6. Instructional policy-making function. This function surprisingly has been overlooked by researchers of instructional leadership. Certain instructional practices involve norms, resources, and staff beyond the individual teacher and his or her classroom. For example, we discovered that a teacher feels uncomfortable about initiating homework unless his or her colleagues also do so. Two teachers at the same grade level may wish to teach their math lesson at the same time so they can group students by ability level. This practice involves resources (for example, a certain type of school schedule) and use of staff that require shifts in instructional policy. Leadership is required to recognize instructional policy issues and to make decisions about them.

7. Assessment function. An important goal of school improvement efforts,

including the present experiment, is student achievement gain. Leadership is needed to assure that appropriate measures of achievement are administered, reported, and used in decision-making. In many but not all respects, the assessment function overlaps the instructional policy-making function. For example, assessment results can be reported as group averages (the usual procedure) or by individual students or subgroups to focus attention on students with special instructional needs.

8. External relations function. Leadership is required to maintain healthy relationships with individuals and agencies that wish to affect or are affected by the school improvement process. For example, assignment of homework was a critical element of the improvement process in our experiment. This instructional practice necessitated a change in school-home relationships, and leadership was required to insure that the transition proceeded smoothly for all involved. Also, the relationship between the Valley Education Consortium and the research staff for this experiment needed occasional attention to insure that the work of each group was mutually beneficial.

9. Maintenance function. One of the most critical leadership functions is to establish mechanisms for insuring that a school improvement process is maintained over time. Substantial momentum was generated in the three districts to deliver Good and Grouws' staff development program to all fourth- and fifth-grade teachers in the schools assigned to the two training treatments. Once the experiment ended, the research staff largely withdrew from the districts. One staff member in each district was trained to deliver the program, but continuing leadership will be required for these trainers to deliver the program to control group teachers and to new teachers entering the district. Also, policy decisions need to be made about delivering the program to teachers at other

grade levels.

These nine leadership functions are very briefly sketched in the above analysis. The significance of the analysis for present purposes is that it reveals the magnitude and complexity of instructional leadership required to bring about school improvement. Previous lists of leadership functions (for example, Sweeney 1982) are much narrower in scope, we believe, because they focus on a single leadership role -- the building principal role -- rather than on the range of external agency, district, and school roles required to launch, implement, and sustain a school improvement process.

The design of the principal involvement treatment reflected this narrow scope because it was based on research on principal leadership. The treatment emphasized the principal leadership functions of attending teachers' staff development sessions, reviewing student achievement scores, and conducting observation-conference cycles with teachers. It was the relative success of the regular inservice group that led us to view instructional leadership as a broader phenomenon and to analyze it into the nine functions described above.

Leadership roles. Some research on principals supports the importance of their role as instructional leaders, whereas other research questions the importance of that role (Gersten et al. 1982). From our current perspective, these disparate findings are not necessarily contradictory. Some of the nine leadership functions identified above can be performed effectively by principals, or by other personnel. For example, district supervisors and curriculum specialists probably have the most expertise and most time to perform the training, assessment, and maintenance functions. The building principal may be the best qualified, by virtue of interest and designated role, to provide leadership for the compliance, coordination, instructional policy-making, and school-community relations functions. The superintendent and assistant superintendents, again by virtue of interest and designated role, may be the most qualified to perform the priority-setting and resource acquisition functions. Teachers may be best qualified to provide leadership, in conjunction with administrators, for the training, instructional policy-making, and compliance functions.

We do not mean to imply that unique leadership functions can be ascribed to each educator role in a district. The functions clearly overlap roles. The point of this brief analysis is that the present experiment suggests the need to include a variety of leadership roles and functions in planning a staff development process intended for school improvement. These roles and functions extend well beyond the school building level. Current opinion, however, is that a building level focus is necessary and sufficient. For example, Neale, Bailey, and Ross (1981) stated that staff development should be "located in the local school building, directed by the principal and staff of that building to meet educational needs identified by the staff and clientele of that school." Our experience in the present study suggests that the principal and staff can provide much leadership for staff development, but that additional leadership resources outside the school are needed too.

The Relationship Between Staff Development, School Organization, and School Improvement

It may seem presumptuous at first glance to generalize from a single study about the relationship between staff development programs and school organization. The attempt seems less presumptuous, though, when we pause to consider that the experiment extended over three different school districts, 53 teachers

(35 of whom were specially trained), and 15 principals (five of whom were specially trained). Also, the staff development program used in the study is representative of a new generation of such programs in which instructional behaviors have been empirically linked to student achievement gains. (These programs are reviewed in Gage and Giaconia 1981).

Our experience in this experiment suggests that the manner in which schools are presently organized creates obstacles to successfully initiating, implementing, and sustaining staff development programs of the type represented by Good and Grouws' Active Mathematics Teaching. Similarly, present school organization acts as an inhibitor to the performance of leadership functions required in using this type of program.

Weick (1976) and other organizational theorists have characterized schools as loosely coupled systems. This term is used to characterize an organization in which elements of the system are attached, but the attachment is "circumscribed, infrequent, weak in its mutual affects, unimportant, and/or slow to respond" (Weick 1976, p. 3). With this meaning in mind, Gall and his colleagues (1982) interpreted current patterns of staff development as a loosely coupled system in school districts: "[Current staff development programs] consist predominantly of isolated inservice activities focusing on teacher development rather than school improvement. The activities are loosely connected to priority goals and to assessment of student outcomes" (p. 117).

An analysis of Good and Grouws' staff development program indicates that it implies a tightly coupled school system, or at least an effort to tighten the coupling of school elements. For example, the effects of the program can be directly related to district and school goals for improvement. Teacher instructional behavior is linked to student behavior in the classroom and subsequently to student achievement. Each of these elements (teacher behavior, student behavior, and student achievement) can be measured, and so there is the potential for administrative control over them that does not exist in staff development programs designed for a loosely coupled system.

If traditional school organization and staff development form a loosely coupled system, the imposition of a tightly coupled program of staff development seems bound to create substantial organizational tension. Our experience suggests that the tension results from the interdependence and complexity of resources required for implementing a tightly coupled program. For example, many human resources must be redirected from other activities in order to get the program underway. A massive effort is required to assess student achievement on curriculum-referenced tests over several points in time. (If this work had not been previously done by the Valley Education Consortium, the assessment could not have been done within project resources.) The need for supervision of teachers' behavior as they learn Good and Grouws' instructional model is still uncertain, but if it is necessary, substantial resources are needed. Also, substantial collaborative decision-making is required to develop the instructional policies implied by a tightly coupled model of instruction. This collaboration is difficult because teachers usually work in isolation from each other and from principals. Indeed, Tye and Tye (1984) recently argued that most school reforms will fail because "teachers are normally isolated from one another...and...do not often come together in their schools to discuss curricular and instructional changes" (p. 319).

The marshalling of resources for a tightly coupled system, then, is substantial. In the face of such requirements, it is easy to see that a loosely coupled organization would resist the system. The resistance would come, in part, from district leaders who see that the requirements exceed the authority they traditionally have assumed within their districts and even the capabilities of their organization.

This analysis of school organization, staff development, and school improvement is based primarily on this one project. Descriptions of other projects suggest that similar conditions exist elsewhere. Levine and Stark (1982), for example, conducted case studies of three districts that recently attempted to improve student achievement in inner-city schools. Their observations of the change process reveal that massive organizational and instructional re-arrangements were necessary for this purpose. These re-arrangements involved such elements as "years of staff development," "systematically coordinating federal, state, and local resour s for compensatory education," "reorganized teachers' resource centers so that all the materials available in the school were coded for essential skills," and taking "great pains to minimize teachers' recordkeeping" (pp. 42-43).

If our analysis has merit, it suggests that there are severe limits on the potential of a building principal to implement a tightly controlled instructional program. The resource requirements of the program exceed the designated leadership authority of most principals. The only alternative, then, is to invoke leadership at higher levels of authority--leadership found at the district, state, and federal levels. Building principals cannot be ignored, however, because as we found, they do make a difference in implementation of a tightly coupled program once the necessary resources have been assembled and organized.

Recommendations for Future Research

This research project provided a substantial experimental test of the proposition that school principals can be enlisted as a resource to enhance the effectiveness of a staff development program for teachers. This proposition, which is grounded in previous descriptive and correlational research, was supported by the results of the experiment.

The experimental test was by no means conclusive, however. Additional research is necessary to establish that the observed effects were not chance events or artifacts created by systematic bias in the formation of treatment groups. Tighter experimental control and a longer timeframe for tracing effects are obvious recommendations. In addition, we recommend that future research include studies of the process by which the effects are produced. It is not at all clear how principals work with teachers on a day-to-day basis to influence classroom instruction. How did principals in the principal involvement treatment talk to teachers about the need for homework or for more whole-class developmental instruction? What did they tell teachers after observing their classrooms for implementation of Good and Grouws' instructional model? What compliance techniques, if any, did they use? Our research methdology did not enable us to answer these important questions, but it is certainly feasible to design studies that would do this.

We recommend, too, that instructional leadership be reconceptualized in future research. Our notions about this concept were largely grounded in the research literature on elementary school principals. Our analysis of events and outcomes of the experiment suggests the importance of a broader conceptual framework. An expanded set of leadership functions and district-level roles needs to be considered. A broader framework will complicate the problem of designing experiments on instructional leadership, but we believe that it will also lead to the design of more effective interventions than the one tested in this project.

Lastly, we issue a note of caution about research testing the generalizability of the findings of this project. The experiment was focused on elementary school principals, intermediate grade teachers, mathematics instruction, and a "tightly coupled" staff development program. The meaning of instructional leadership and effective instruction is likely to shift substantially as one focuses on other levels of schooling, on other curriculum domains, and on other types of staff development. If this is true, the generalizability of the observed treatment is likely to be low. Rather than testing the limits of generalizability, we recommend replicative studies of the same situation for the purpose of developing deeper understanding of instructional leadership phenomena. When the operation of instructional leadership in one situation is well understood, it should be possible to develop more powerful studies of related phenomena in other school contexts.

REFERENCES

Anderson, L.M.; Evertson, C.M.; and Brophy, J.E. "An Experimental Study of Effective Teaching in First-Grade Teaching Groups." The Elementary School Journal 79 (1979): 193-221.

Andros, K., and Freeman, B. "The Effects of Three Kinds of Feedback on Math Teaching Performance." Paper presented at the annual meeting of the American Educational Research Association, April 1981, Los Angeles.

Austin, G.R. Process Evaluation: A Comprehensive Study of Outliers. Baltimore: Maryland State Department of Education, 1978.

Berman, P., and McLaughlin, M.W. Federal Programs Supporting Educational Change. Vol. VIII: Implementing and Sustaining Innovations. Santa Monica, California: Rand Corporation, May 1978.

Borg, W.R., Kelley, M., Langer, P., and Gall, M.D. The Minicourse: A Microteaching Approach to Teacher Education. Beverly Hills, California: Macmillan Educational Services, 1970.

Bossert, S.T.; Dwyer, D.C.; Rowan, B.; and Lee, G.V. The Instructional Management Role of the Principal: A Preliminary Review and Conceptualization. San Francisco: Far West Laboratory for Educational Research and Development, October 1981.

Brookover, W.B., and Lezotte, L.W. Changes in School Characteristics Coincident with Changes in Student Achievement. East Lansing, Michigan: Michigan State University, Institute for Research on Teaching, 1979. (ED 181 005).

Campbell, R.J. "Knowledge, Control and In-Service Education and Training for Teachers." British Journal of Teacher Education 7 (1981):149-163.

Charters, W.W., Jr., and Jones, J.E. "On the Risk of Appraising Non-Events in Program Evaluation." Educational Researcher 2 (1973):5-7.

Colardarci, T. A classroom-based experiment assessing the impact of teacher behavior and student achievement of a direct-instruction approach to teaching. Unpublished doctoral dissertation, Stanford University, 1980.

Collins, R., et al. Inservice Education Bibliography. Syracuse, New York: National Council of States on Inservice Education, 1979.

Crawford, J., et al. An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade, Vols 1-3. Stanford, California: Center for Educational Research at Stanford, Stanford University, 1978. ERIC document ED 160 148.

DeBevoise, W. "Synthesis of Research on the Principal as Instructional Leader." Educational Leadership 41 (1984):14-20.

Duckworth, K. Specifying Determinants of Teacher and Principal Work. Eugene, Oregon: OR: University of Oregon, Center for Educational Policy and Management, 1983.

Ebmeier, H., and Good, T. "The Effects of Instructing Teachers about Good Teaching on the Mathematics Achievement of Fourth-Grade Students." American Educational Research Journal 16 (1979):1-16.

Edmonds, R. "Effective Schools for the Urban Poor." Educational Leadership 37 (1979):15-27.

Elashoff, J.D. "Analysis of Covariance: A Delicate Instrument." American Educational Research Journal 6 (1969):383-401.

Fisher, C.; Berliner, D.; Filby, N.; Marliave, R.; Cahen, L.; and Dishaw, M. "Teaching Behaviors, Academic Learning Time, and Student Achievement: An Overview." In Time to Learn, edited by C. Denham and A. Lieberman. Washington, D.C.: National Institute of Education, 1980.

Gage, N.L., and Giaconia, R. "Teaching Practices and Student Achievement: Causal Connections." New York University Education Quarterly 12 (1981): 2-9.

Gall, M.D.; Haisley, F.B.; Baker, R.G.; and Perez, M. "The Relationship Between Inservice Education Practices and Effectiveness of Basic Skills Instruction." Eugene, Oregon: University of Oregon, Center for Educational Policy and Management, December 1982.

Gersten, R.; Carnine, D.; and Green, S. "The Principal as Instructional Leader: A Second Look." Educational Leadership 40 (1982):47-53.

Good, T.L., and Grouws, D.A. "Teaching Effects: A Process-Product Study in Fourth Grade Mathematics Classrooms." Journal of Teacher Education 28 (1977):49-54.

Good, T.L., and Grouws, D.A. "The Missouri Mathematics Effectiveness Project: An Experimental Study in Fourth Grade Classrooms." Journal of Educational Psychology, 71 (1979):355-362.

Good, T.L., and Grouws, D.A. Experimental Research in Secondary Mathematics Classrooms: Working with Teachers. (Final Report of National Institute of Education Grant NIE-G-79-0103.) Columbia, Missouri: University of Missouri-Columbia, May 1981.

Good, T.L.; Grouws, D.A.; and Ebmeier, H. Active Mathematics Teaching. New York: Longman, 1983.

Hall, G.E., and Loucks, S.F. Program Definition and Adaptation: Implications for Inservice. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1980.

Hall, G.; Rutherford, W.L.; Hord, S.M.; and Huling, L.L. "Effects of Three Principal Styles on School Improvement." Educational Leadership 41 (1984): 22-29.

Hunter, M. Improved Instruction. E. Segundo, California: TIP Publications, 1976.

Joyce, B.R.; Howey, K.; and Yarger, S. I.S.T.E. Report I. Palo Alto: Stanford Center for Research and Development in Teaching, 1976.

Kean, M.H., et al. What Works in Reading? Philadelphia: Philadelphia School District, Office of Research and Evaluation, May 1979.

Keziah, R. "Implementing Instruction Behaviors That Make a Difference." Centroid (North Carolina Council of Teachers of Mathematics) 6 (1980):2-4.

Leithwood, K.A., and Montgomery, D.J. "The Role of the Elementary School Principal in Program Improvement." Review of Educational Research 52 (1982): 309-339.

Levine, D.U., and Stark, J. "Instructional and Organizational Arrangements That Improve Achievement in Inner-City Schools." Educational Leadership 40 (1982):41-46.

Linn, R.L., and Slinde, J.A. "The Determination of the Significance of Change Between Pre- and Posttesting Periods." Review of Educational Research 47 (1977):121-150.

Lortie, D. School Teacher. Chicago: University of Chicago Press, 1975.

Madden, J.V., et al. School Effectiveness Study: State of California. Sacramento: California Department of Education, 1976.

McIntyre, D.J.; Copenhaver, R.W.; Byrd, D.M.; and Norris, W.R. "A Study of Engaged Student Behavior Within Classroom Activities During Mathematics Class." Journal of Educational Research 77 (1983):55-59.

Miller, L., and Wolf, T.E. "Staff Development for School Change." In Staff Development: New Demands, New Realities, New Perspectives, edited by A. Lieberman and L. Miller. New York: Teachers College, Columbia University, 1979.

Mohlman, G.G.; Coladarci, T.; and Gage, N.L. "Comprehension and Attitude as Predictors of Implementation of Teacher Training." Journal of Teacher Education 33 (1982):31-36.

Neale, D.C.; Bailey, W.S.; and Ross, B.E. Strategies for School Improvement. Boston: Allyn and Bacon, 1981.

New York State Office of Education. School Factors Influencing Reading Achievement: A Case Study of Two Inner City Schools. Albany, New York: Office of Education, March 1974.

Pitner, N.J. "Training of the School Administrator: State of the Art." Eugene, Oregon: Center for Educational Policy and Management, University of Oregon, 1981.

Rogosa, D.; Brandt, D.; and Zimoski, M. "A Growth Curve Approach to the Measurement of Change." Psychological Bulletin 92 (1982):726-748.

Rosenshine, B. "Classroom Instruction." In The Psychology of Teaching Methods: The 75th Yearbook of the National Society for the Study of Education, Part I, edited by N.L. Gage. Chicago: University of Chicago Press, 1976.

Schalock, D. "A Summary of Findings From Research on the Continued Professional Development of School Personnel in Oregon." Unpublished report prepared by a Profession Wide Study Group on the Continued Professional Development of School Personnel in Oregon, November 1977.

Shavelson, R.J. Statistical Reasoning for the Behavioral Sciences. Boston: Allyn and Bacon, 1981.

Shoemaker, J., and Fraser, H.W. "What Principals Can Do: Some Implications from Studies of Effective Schooling." Phi Delta Kappan 63 (1981):178-182.

Showers, B. Transfer of Training: The Contribution of Coaching. Eugene, Oregon: University of Oregon, Center for Educational Policy and Mangement, December 1982.

Slavin, R.E., and Karweit, N.L. Mathematics Achievement Effects of Three Levels of Individualization: Whole Class, Ability Grouped, and Individualized Instruction. Technical Report 349. Baltimore: The Johns Hopkins University, Center for Social Organization of Schools, 1984.

Sparks, G.M. "Synthesis of Research on Staff Development for Effective Teaching." Educational Leadership 41 (1983):65-72.

Stallings, J. "Allocated Academic Learning Time Revisited, or Beyond Time on Task." Educational Researcher 9 (1980):11-16.

Sweeney, J. "Research Synthesis on Effective School Leadership." Educational Leadership 39 (1982):346-352.

Tye, K.A., and Tye, B.B. "Teacher Isolation and School Reform." Phi Delta Kappan 65 (1984):319-322.

Venezky, R.L., and Winfield, L.F. Schools That Succeed Beyond Expectations in Reading. (Studies on Education Technical Report No. 1.) Newark, Delaware: University of Delaware, 1979.

Weber, G. Inner-City Children Can Be Taught to Read: Four Successful Schools. Washington, D.C.: Council for Basic Education, 1971.

Weick, K.E. "Educational Organizations as Loosely Coupled Systems." Administrative Science Quarterly 21 (1976):1-19.

Wellisch, J.B., et al. "School Management and Organization in Successful Schools." Sociology of Education 51 (1978):211-226.

APPENDIX A

Characteristics of Treatment Groups

APPENDIX A

Characteristics of Treatment Groups *

Teacher ID	District ID	School ID	Grade	Sex
Principal Involvement Group				
01	A	2	4	F
02	A	2	4	M
03	A	2	5	M
04	A	2	5	F
05	A	7	5	M
06	A	7	4	F
07	A	8	5	F
08	A	8	5	F
09	A	8	4	M
10	B	4	5	F
11	B	4	5	M
12	B	4	4	F
13	B	4	5	F
14	B	4	4	M
15	C	14	4	F
16	C	14	4	F
17	C	14	5	M
18	C	14	5	F
Regular Inservice Group				
19	A	4	4	F
20	A	4	5	M
21	A	4	4	F
22	A	4	5	M
23	A	5	4	F
24	A	5	5	M
25	A	9	4	F
26	A	9	4	F
27	A	9	5	M
28	A	9	5	M
29	B	10	4	F
30	B	10	5	F
31	B	10	4	M
32	B	10	5	M
33	C	13	4	F
34	C	13	5	M
35	C	13	5	F

* The letters A, B, and C are used to designate the three participating districts. The school ID's are those used in the original coding of the data, and so are not in numerical order in this table.

Teacher ID	District ID	School ID	Grade	Sex
Control Group				
36	A	1	4	M
37	A	1	4	F
38	A	1	5	F
39	A	1	5	M
40	A	3	4	M
41	A	3	4	M
42	A	3	5	F
43	A	3	5	F
44	A	6	4	F
45	A	6	5	F
46	B	12	4	M
47	B	12	4	F
48	B	12	5	M
49	B	12	4	F
50	C	15	5	M
51	C	15	5	F
52	C	15	4	F
53	C	15	4	F

SH468

APPENDIX B

Interview Schedule for Principals

Date ________ School ________ District ________ Principal ________

Gathering Information about the Form and Level
of a Principal's Involvement in Instruction-Related Activity

Principals in all participating districts
Interview Schedule

Supervision

1. Do you follow a general model or approach to supervision, e.g., ITIP? If so, please describe the assumptions underlying the approach and the procedures you use to implement it. If not, describe the approach you take to supervision generally.

____

____

____

____

____

2. During the last 10 days how many different classrooms have you observed in?

____

About how long was the typical observational period?

____

What kinds of processes did the observations focus on?

____

____

What did you do with information collected through the observations?

____

____

3. During the year generally about how much time do you spend weekly observing teachers teach?

4. Do you (or does the district) require teachers to submit instructional plans?
_____Yes _____No

If yes check all that apply.

_____ long-range; i.e., yearly or semester plans
_____ unit plans
_____ weekly plans
_____ daily plans

5. If teachers are required to submit plans, who reviews them?
_____You _____Other (specify)

- How do you review them? e.g., just check to make sure they're done? make comments on them?

____

____

- Does a teacher's instructional plan influence what you observe during classroom visitations? If yes, indicate how.

____

____

6. Do teachers submit to you on a regular basis information on their students' progress toward attaining desired learning goals? (Over and above the quarterly reports they prepare for parents) _______

If yes, what form does this information take, e.g., test scores? teacher judgments? observational records?

____

____

<u>Placement</u>

7. What role, if any, do you play in placing students in your building?

Role in Grade or Classroom Placement Decisions	Role in Promotion/Retention Decisions
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________

What factors do you consider when making grade or classroom placement decisions?

___

___

___

What factors do you consider in making promotion/retention decisions?

___

___

___

Program Monitoring and Evaluating

8. What kind of evidence do you look to during the year as a sign that the math program in your building is producing desired effects, or is falling short of expectations, e.g., test scores, teacher perceptions, parent comments, personal observations, etc.?

___

___

___

9. What evidence, if any, do you look to at the end of the year, or over a series of years, to evaluate the effectiveness of the program? Please rank these factors according to their importance in determining overall effectiveness?

___

___

___

10. In what ways, if any, do you share information about the effectiveness of the program with teachers?

___

With central office staff?

___

With parents?

___

With local Board members?

___

Involvement with Teacher Inservice

11. To what extent, if any, do you plan, schedule, or coordinate inservice activities for teachers in your building?

____

____

____

12. Do teachers in your building regularly exchange ideas and information with each other about instruction-related matters?

____

____

____

If they do, please describe the form through which these exchanges take place and your perceptions of the kinds of outcomes produced by these exchanges.

____

____

____

What role, if any, do you play in facilitating these exchanges?

____

____

____

13. What incentives, if any, do you offer to teachers who stimulate and sustain their involvement in professional development?

____

____

____

14. To what extent do you participate in teacher inservice activities?____________

____

If you do participate, list those activities in which you have participated during 1982-83.

____

____

What activities did you participate in last year?

15. What is your overall impression of the quality and effects of the inservice opportunities provided teachers in your district?

<u>Managing the School as a Learning Environment</u>

16. Recent research has demonstrated a rather obvious fact: student absences contribute significantly to the problem of student low achievement. Do you or your staff use any particular procedures to control absence and tardy rates, e.g., hiring aides to call parents at night about absent or tardy students?

17. Studies also have shown that students' learning suffers in classrooms that frequently are intruded upon, e.g., by requests for students to leave the room, by announcements on the intercom, by tardy students entering the room, etc. Does your school have any particular policies designed to minimize classroom intrusions? Please describe.

18. As you well know, time allotted to instruction is associated with student learning. Does your school have any guidelines for teachers about the amount of time they should spend per week, or per day, on instruction in a particular subject area? If yes, please explain.

Questions about this Study

19. Do you have any questions or concerns about this study?

DF103

May 1983

MATH INSERVICE STUDY
Exit Interviews with Principals

Schools with Involved Principals

Training Sessions

1. Were there any particular positive effects or negative effects associated with your participation in the training sessions last February?

Positive Effects	Negative Effects
for the principal	for the principal
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
for the participating teachers	for the participating teachers
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________

2. What was your perception of the quality and appropriateness of the training sessions? of the teachers' handbook that supported the training effort? (ask if principal had opportunity to review the handbook outside the context of the training session)

____

____

____

____

3. Are there any ways you can think of to improve the training program?

____

____

____

____

Principal's Observation of Teachers' Use of Model

4. Did you have an opportunity to make two formal observations of teachers' use of the Good & Grouws' model? Yes______ No______

   If no, how many observations were made? ______

5. What procedure did you use to conduct formal observations of teachers' use of the model? (e.g., the checklist prepared by project staff, a self-made observational code)

   ____

   ____

   ____

   ____

   How did this procedure relate to the supervisory procedures you ordinarily use?

   ____

   ____

   ____

   ____

6. What did you do with the information obtained through the observation?

   ____

   ____

   ____

   ____

7. How do you feel teachers responded to the observations you conducted and the information coming from them?

   ____

   ____

   ____

   ____

8. How often did you make informal observations of teachers' math lessons?

_____ weekly
_____ about every other week
_____ about once a month
_____ once every couple of months
_____ once or twice between February and June

9. If you did make informal observations, did you learn different things from them from the formal observations?

____

____

____

____

10. What would you say was the most difficult part of the Good & Grouws' model for teachers to use? Why?

____

____

____

____

What part was the easiest?

____

____

____

____

What part, or combination of parts seemed to produce the largest positive effects for teachers and students?

____

____

____

____

Principal's Role in Fostering Colleague Exchange

11. Did you have an opportunity to arrange a meeting with participating teachers in your building to discuss and resolve any problems that developed about using the model? If so, what specifically was discussed at the meetings and what decisions were reached?

If not, was there any other way in which you promoted teacher interaction around issues related to the Good & Grouws' model?

Review of VEC Test Data with Teachers

12. Did you have an opportunity to review with 4th & 5th grade teachers results from the VEC mid-year test? If so, what procedures did you use to review the test information?

13. How did teachers respond to the VEC test data?

14. What is your impression of the quality and utility of the VEC test data?

Support in Carrying Out Role of Involved Principal

15. Do you feel you received enough guidance and support to carry out your role in the inservice program?

16, What additional or different forms of guidance and support would have been of help?

Perceptions of Role as an Involved Principal

17. Were the demands of time and energy made by the project manageable or unmanageable for you?

18. Was there anything you did in your role as an involved principal that seemed particularly productive, either for you or your staff?

____

____

____

____

Was there anything that seemed particularly unproductive or disappointing?

____

____

____

____

Future Possibilities

19. Would you recommend that other teachers in your building receive training next year in the Good & Grouws' model?

____

____

____

____

20. Do you think the kind and level of involvement that you had in this project would be appropriate for other staff development efforts in your building or district?

____

____

____

____

SP645

May 1983

MATH INSERVICE STUDY
Exit Interviews With Principals

Schools Receiving Teacher Inservice Without Principal Involvement

1. What reactions do you have, if any, to the Good and Grouws' approach to math teaching that 4th and 5th grade teachers in your building receiv. ' training in this year?

   Possible Probes - In your view, is there anything particularly distinctive about this approach to math? If so, what leads you to think this? What's your sense of the teachers' response to the model? Did it help or hinder them in any particular way? What's your sense of the impact or potential impact, of the model on student learning?

2. What impressions do you have of the quality of the training sessions for teachers that took place last February? What's the basis of your impressions?

3. How did teachers react to the classroom observations that were made by project staff?

4. During the time the project was being carried out, February 1 to the end of the year, how many formal observations of each 4th & 5th grade teachers' math lessons did you make? (formal observation = observation has preestablished focus; observation extends full lesson; information from observation discussed with teacher)

______

About how often did you make informal observations of these teachers' math lessons?

_____ weekly
_____ about every other week
_____ about once a month
_____ once every couple of months
_____ once or twice between February and June

5. Based on the information available to you, what would you say was the most difficult part of the Good & Grouws' math model for teachers to use? Why?

______

______

______

What was the easiest?

______

______

______

Which part, or combination of parts, if any, seemed to produce the largest positive effects for teachers and students?

______

______

______

6. Have you done anything different in relation to the math program in your school, or in relation to observations of teachers' math lessons, during the last semester? If so, are these differences linked to the math inservice project in any way?

______

______

______

______

7. What, if anything, have you heard about the work done by the principal(s) in your district who have taken an active part in the math inservice program?

8. Would you recommend that other teachers in your building receive training next year in the Good & Grouws' model?

9. If similar inservice training was to be provided next year, would you consider it worthwhile to attend the training sessions?

Supplementary Section on VEC Test Data

10. Did you have an opportunity to review with 4th & 5th grade teachers' results from the VEC mid-year tests? If so, what procedures did you use to review the test information?

11. How did teachers respond to the VEC test data?

12. What is your impression of the quality and utility of the VEC test data?

May 1983

MATH INSERVICE STUDY
Exit Interviews With Principals

Control Schools

1. What, if anything, have you heard about the math inservice project that some of the other elementary schools in the district have participated in this year? From whom did you get this information?

   Possible Probes - What have you heard about: a) the Good & Grouws' model? b) the training sessions for teachers that took place last February? c) the observations that were made of 4th & 5th grade teachers' math lessons? d) what the principals who were involved in the staff development did, and how they felt abut what they did?

Heard from teachers in your building	Heard from teachers who received training
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________

   Heard from principals or other administrators in the district

   ____

   ____

   ____

2. If principals have heard something about the project, ask this:

   Based on what you've heard, would you like your teachers to receive training in the Good & Grouws' model next year? If so, what do you think the main benefits would be?

   ____

   ____

   ____

   ____

   ____

3. During the time the project was being carried out, February 1 to the end of the year, how many formal observations of each 4th & 5th grade teachers' math lessons did you make? (formal observation = observation has preestablished focus; observation extends full lesson; information from observation discussed with teacher)

____

About how often did you make informal observations of these teachers' math lessons?

_____ weekly
_____ about every other week
_____ about once a month
_____ once every couple of months
_____ once or twice between February and June

4. Do you feel teachers in your building have done anything different in math this last semester from what they have done in the past? If so, are these differences linked to the math inservice project in any way?

5. Have you done anything different in relation to the math program in your school, or in relation to observations of teachers' math lessons, during the last semester? If so, are these differences linked to the math inservice project in any way?

6. If similar inservice training was to be provided next year, would you consider it worthwhile to attend the training sessions?

SP645

APPENDIX C

An Introduction to the Valley Education Consortium

APPENDIX C

An Introduction to the

VALLEY EDUCATION CONSORTIUM

WHAT IS THE VALLEY EDUCATION CONSORTIUM (VEC)?

It is people from many realms of education, including teaching, management, service, teacher preparation, and research, working together to solve educational problems of common concern.

HOW DOES IT WORK?

In addition to a formal organizational structure of a Board (the chief administrator from each member agency) and a Council (a key middle-management representative from each agency), the Consortium accomplishes its work through cross-district work groups, seminars, and work sessions, school-based support teams, and special district projects.

WHO BELONGS TO THE CONSORTIUM?

Currently active members of the Consortium include: the Carlton Elementary, Cascade, Central, Dallas, Dayton, McMinnville, Salem, Silverton Union High, Willamina, and Woodburn School Districts; the Marion, Polk, and Yamhill Education Service Districts; Western Oregon State College, and the Teaching Research Division of the Oregon State System of Higher Education. By virtue of their ESD's involvement, all school districts in Marion, Polk, and Yamhill Counties are associate members.

WHY DID IT BEGIN AND WHERE IS IT GOING?

The Consortium began operating on an informal basis in the early 1970's to help Western Oregon State College (formerly Oregon College of Education) develop its teacher preparation program. During 1974-75 the Consortium grew in membership as districts and agencies sought common ways of responding to the State Department's new Minimum Standards for Elementary and Secondary Schools in Oregon. In 1976 a formal charter for the Consortium was adopted. Since that time VEC has focused its efforts on designing a comprehensive approach to the improvement of school programs (using Oregon's Standards for Elementary and Secondary Schools as a base) and on providing support to schools involved in program improvement. Work thus far has centered largely on the basic skills, but it is now expanding to include more sophisticated subject areas, including the development of critical thinking and problem solving (reasoning) skills. The Consortium also is working actively with faculty and administrators from WOSC to develop programs that will prepare teachers to function effectively within Standards-based instructional programs.

APPENDIX D

Trainer Guidelines for Good and Grouws' Staff Development Program

1/20/83

First Training Session
Good and Grouws' Inservice Program

Objectives:

1. To develop teacher enthusiasm for the program.
2. To provide a rationale for why the program works.
3. To help teachers implement the program in their classroom.
4. To help teachers learn how to write math lesson plans based on the program.

Why the District Supports the Program:

1. Relationship of the program to district goals and philosophy.
2. The program is research-validated. Research demonstrates that it helps teachers and students.
3. The program is based on what effective teachers do. ("Effective" teachers are those whose kids learn math well.) The teaching model builds on what teachers already do. Some teachers follow the model closely even without special training.
4. The teaching model does not mean more work for the teacher. Also, it can be used with any curriculum.

Research Project Associated With the Program:

1. Some district teachers are not getting the training now. Some of you are getting it, but not your principal. Some of you are getting it along with the principal.
2. It's important not to talk to teachers or principals in other schools so we can find out which inservice approach is best. We will share the results with you at the end of the test.
3. Any questions about the classroom observations that have occurred over the past two weeks?
4. Reassure everyone that all of the research data are confidential.

ASCD Videotape:

1. Videotape is approximately one-half hour.

2. Answer questions about the tape afterwards.

Overview of the Teaching Model:

1. Major components of the model

   a. daily review
   b. lesson development
   c. seatwork
   d. homework
   e. special reviews

2. The essence of the model is an integrated sequence of: development-practice-review.

3. Overhead transparencies can be used to present the overview.

4. Answer questions about the overview.

Distribute Teacher Handbook:

1. As time permits, have teachers and principals read several sections of the handbook.

2. Answer questions.

3. Assignment: read rest of handbook by next training session.

Construct a Lesson Plan Using the Teaching Model

1. Distribute packet of lesson plan forms.

2. Have teachers and principals construct a math lesson plan using the form provided.

Small Group Discussion:

1. Give teachers and principals a chance to form small groups and discuss their reactions to the teaching model. Teachers whose principals are involved should form separate groups school by school.

2. Allow approximately 20-30 minutes for this activity.

Concluding Remarks:

1. Repeat assignment: read rest of teacher handbook.

2. Start implementing the model to the extent possible.

3. Date of next session.

4. Purpose of next session is to discuss your experiences in implementing the model, and to share ideas about how to make the model work.

Whole Group versus Individualized Instruction

Problems with Totally or Largely Individualized Math:

1. Students tend to get off task easily and stay off task for long periods of time.

2. Children tend to get frustrated because it is not easy to get teacher assistance.

3. Many students at fourth and fifth grade level aren't able to work independently under individualized instruction.

4. Not enough opportunity for teachers to explain concepts and procedures in detail.

Key Points:

1. Good and Grouws' teaching model is not opposed to individualized instruction. Seatwork and homework can be individualized. The model just suggests doing certain activities (review, mental computation, development) with the whole group.

2. Research shows that children benefit by spending as much time as possible with the teacher. In individualized instruction each child spends very little time with the teacher.

3. The teacher is critical in elementary school because many children don't have the reading skills necessary to learn from math textbooks and learning packets. Effective whole group instruction maximizes children's exposure to the teacher.

Second Training Session
Good and Grouws' Inservice Program

Introduction

1. Answer questions about points made in the teachers handbook.

2. Have teachers share experiences in implementing the teaching model.

3. Show second part of videotape (complete math lesson illustrating the teaching model).

4. Answer questions about the videotape.

Review the Teaching Model

1. Review each part of the model, and bring out the following points.

2. DAILY REVIEW. The review should cover the previous day's work. It needn't take but a few minutes. The review helps children stay on track and reinforces what they have just learned.

3. COLLECT HOMEWORK. Homework can be checked quickly by putting answers on the blackboard or on transparencies. Children can check their own answers. Spot-check a few problems to see if they've been done correctly.

4. MENTAL COMPUTATION. It's not important when in the lesson this strategy occurs. Children should do the computations in their head. That is the point of mental computation.

5. DEVELOPMENT. Teach for meaning. Teach not just how to do the computations, but why. Stress real-life applications of the computations. Use manipulatives to develop concepts; manipulatives are quite appropriate for fourth and fifth graders. Try spending more time on development: children will retain what they've learned better and their seatwork and homework will be more successful. Let children reason the process through.

6. CONTROLLED PRACTICE. Make sure you check for understanding before assigning seatwork. This is done by assigning a few problems similar to those included in the seatwork. Have the whole class do each problem, and check on the accuracy of their response. If students show lack of understanding, engage in further development (step 5).

7. SEATWORK. Try to assign enough seatwork so that students are busy the whole time. Monitor seatwork continuously rather than working on other activities.

8. HOMEWORK. Ten to fifteen minutes is sufficient. The purpose of homework is to provide successful, distributed practice. Homework should reinforce the day's development and seatwork.

9. SPECIAL REVIEWS.

10. OFF TASK BEHAVIOR.

    a. minimize transitions from one activity to another.
    b. do not allow unlimited use of the bathroom.
    c. do not interrupt the math lesson with other activities, e.g., distributing lunch tickets.
    d. avoid excessive student movement in the room or leaving the room while the math lesson is in progress.
    e. do not interrupt children during seatwork.
    f. if a child is at the blackboard, keep other children involved.

Construct a Lesson Plan Using the Teaching Model

1. Ask teachers whether they've used the lesson plan forms provided in the first session.

2. Have teachers and principals construct a math lesson plan using the form provided.

Whole Group and Individualized Instruction

1. Discuss problems in adapting the teaching model to individual differences among children in learning math.

Small Group Discussion

1. Provide opportunities for discussion among participants. Teachers whose principals are involved should form separate groups school by school.

2. Allow approximately 20-30 minutes for this activity.

Conclusion

1. Remind participants not to talk with teachers and principals in other schools about this inservice program.

2. Ask participants not to mention to observers which experimental group they're in.

3. Give approximate dates of two classroom observations and student testing.

4. Get teachers' math lesson schedules.

5. Have teachers and principals complete end-of-session form.

1/20/83

Questions and Answers About the Teaching Model

Q: How closely must we stick to the time allocations?

R: We are asking that you stay reasonably close to the suggested time lines. Previous research suggests that these requests will result in higher student achievement and we are anxious to give this idea a rigorous test. Having said this, we realize that from time to time some variance will be necessary; we ask only that you give your best effort.

Q: What about a student who often does not do the homework assignments?

R: It is important that students do the homework assignments. Try to determine why the assignments are not being done...too difficult? laziness? and act accordingly. Sometimes a talk with the student will help. Other times, a chat with the parents may be appropriate. We leave it to your professional judgment as to what is best in your particular situation.

Q: How can I increase the amount of time I spend on development?

R: First you might consider whether you are spending appropriate amounts of time on initial development of meaning, assessment, controlled practice, and reteaching. Beyond this, initial development can be profitably expanded by using new or additional concrete materials, or providing supplemental activities (that fit with the objectives of the day) generated from game books, other textbooks, professional journals, or from colleagues. Remember that presenting ideas in several different contexts is part of development and often helps students to transfer the ideas to new situations.

Q: Have you got some additional ideas for mental computation?

R: You might work on multiplication and division by ten. Maybe we could share some ideas on this with one another. Mrs. Wilson what have you...

Q: What about grading practices with this program. Does the homework count as part of the grade?

R: We ask that you evaluate student performance using the standards you have always used. If you feel seatwork and homework performance should be a part of this assessment that certainly would be okay. In fact, that seems to be a very reasonable sort of thing to do. What part of the grade is again left to your discretion.

Q: I have always used periodic quizes when I teach math. Can I use them with this program?

R: There would be no problem with continuing to use these quizes. Of course using them on a daily basis would tend to interfere with the instructional pattern that we are asking you to follow.

Q: Is it required that the reviews be done on Monday?

R: We would like you to do them as often as possible on Monday so that we can get in the reviews at approximately equal intervals. However, if things work out so that you can finish a unit or chapter on a Monday then by all means put off the large review until Tuesday and if you wish to give a unit test give it on Wednesday.

Q: What about the frequency with which I test my students?

R: This is up to you and your judgement.

WEEK OF: UNIT:

	WEEKLY REVIEW 20 MINUTES	DEVELOPMENT (10 MINUTES)	SEATWORK (10 MIN)	ASSIGNED HOMEWORK	COMMENTS
MONDAY					

	8 minutes: CHECK HOMEWORK	REVIEW PREVIOUS DAY	MENTAL COMPUTATION	DEVELOPMENT (20 MINUTES)	SEATWORK (15 MIN)	ASSIGNED HOMEWORK	COMMENTS
TUESDAY							
WEDNESDAY							
THURSDAY							
FRIDAY						NONE! 	

APPENDIX E

Notes on Training Sessions

APPENDIX E
NOTES ON TRAINING SESSIONS

First Training Session in District 1 - January 31, 1983

1. 24 teachers and principals were in attendance.
2. Trainers put up two posters. Poster 1:

   Schedule

8:00 - 8:15	Coffee
8:15 - 8:45	Introduction
8:45 - 9:30	Overview
9:30 - 9:45	Break
9:45 - 10:30	Good & Grouws model
10:30 - 11:00	Lesson Planning
11:00 - 11:30	Summary & Next Steps

   Poster 2:

   Objectives

   Participants will:
   - have an awar... of the Good & Grouws model
   - know and explai.. the components and the lesson plan
   - design a lesson using the lesson plan
3. Teacher asked, should mental computation relate to the current lesson?
4. Teacher said, the model says nothing about testing and quizzing.
5. Teacher said, the model does not talk about what to do with kids who finish seatwork early.
6. Teacher asked, should you assign homework if students do seatwork wrong?
7. Teacher asked, how should you check homework? Should you spotcheck?
8. Teacher asked, how do you deal with different ability groups in the classroom?
9. Trainer related Good and Grouws' concept of development to ITIP concepts of: "teaching to an objective," "monitor and adjust," "applied practice," "modelling," and "anticipatory set."
10. Trainer had participants read selected pages from teacher handbook, especially section on homework.
11. Teacher said, students shouldn't have homework. Students have enough work to do in seatwork. Parents can foul students up by teaching things wrong. Research staff members responded, research shows homework is effective.
12. Teacher said, his daughter, who is in a second grade private school, has homework every night. He thinks homework is good and assigns it. He checks homework when students start the math lesson. If a child hasn't done his homework, he has to do it during recess. This is very effective.
13. Several teachers said, they teach small groups during math, and this model of whole-group teaching will be a problem for them.

First Training Session in District 2 - February 1, 1983

1. Nine teachers and one principal were in attendance.
2. Teachers said, if children get homework in every subject, that's 2 or 3 hours of homework a night. This is a concern of the teachers.
3. Teacher asked, what do you do if you have a "b[illegible] [illegible]her (a preservice teacher doing a practicum) or student teacher [illegible]room?

4. Teacher said, Japanese students do very well, and they have total group instruction and lots of homework.
5. Principal said, don't spend so much time worrying about the bottom ten percent of the class. Work with the children who are ready to work. (This was said after some of the teachers said that slow students in the class wouldn't do homework or get on-task quickly as lesson shifted from one activity to another.)

First Training Session in District 3 - February 3, 1983

1. Session started with assistant superintendent giving a pep talk on the district's support of the project. He left after about ten minutes.
2. Teacher asked, how does one form classroom groups to handle individual differences? Another teacher suggested, if you have two fourth-grade classrooms, you can form two groups. One teacher can take the high group, and the other teacher can take the low group.
3. Teacher complained, "We don't have a choice. District wants us to implement the program." Teacher referred to the assistant superintendent's pep talk at the beginning of the session.
4. Teachers recommended, district should send the letter to parents concerning homework. (Good and Grouws' teachers manual contains a sample letter than can be sent to parents explaining the shift to use of homework assignments.) Much conversation about this procedure.
5. Teachers talked a lot about slow learners in their classroom.

Second Training Session in District 1 - February 17, 1983

1. A schedule was presented on a poster:

8:15 - 8:30	Set
8:30 - 9:00	Review
9:00 - 10:00	Problem-Solving
10:00 - 10:15	Break
10:15 - 10:30	Observation Form
10:30 - 10:45	Test Results
10:45 - 11:15	School Groups
11:15 - 11:30	Summary

2. Another poster stated objectives:
   Participants will:
   - refine their understanding of the Good and Grouws lesson design
   - review solutions for classroom implementation and problems
   - explain how to read the observation form
   - understand the VEC test printout.
3. Trainer made constant reference to the relationship between Good and Grouws' model and Madeline Hunter's model.
4. Teacher asked, suppose children don't get the concept during seatwork. Should you assign homework? Trainer responded, Yes, to maintain the expectation that there will be homework.
5. Teacher asked, how do you get sufficient time to reteach the small group of children who aren't getting it.

6. Teachers expressed tension about doing a review and finding out children didn't get it. Do you march on to new concepts or reteach it?
7. A few teachers said they have highly individualized programs and would find it very difficult to switch to the whole group approach recommended in Good and Grouws' model.
8. Teachers said, there is a great range of SES differences in their district.

Second Training Session in District 2 - February 16, 1983

1. An account of this session by a CEPM staff person not affiliated with this research project is presented at the end of this appendix.
2. Eight teachers, the principal, and a block student were present.
3. A teacher complained about two very slow learners in her group. They're just not learning math in his classroom.
4. Teacher said, it's good to collect homework at start of the day so children don't do homework in morning before math period.
5. Teacher said, Good and Grouws model is good because low ability students find security in its structure and routine.
6. Teacher observed, students are making good progress after she started giving homework consistently.
7. Teacher said, she really likes mental computation part of the lesson.
8. Principal asked, do teachers use manipulatives in their math lessons. Teachers said there was a lack of manipulatives for them to use.
9. Teachers said, they feel bound by having to march children through the curriculum.
10. Teacher said, one parent wrote her a letter about the new homework policy, "It's about time."
11. Teacher told the principal, tell office not to call me over the intercom during math period.

Second Training Session in District 3 - February 18, 1983

1. Teachers said, children like a predictable structure. They like knowing there will be mental computation each day.
2. Teachers spent a lot of time talking about the few children who are way behind in math.
3. Teacher said, there is a benefit to multiple demonstrations. More and more children tune in with each demonstration. The light bulb goes on for them.
4. Some teachers report, they are doing switching. Slow children go to one teacher; bright children go to the other teacher.
5. Twelve participants attended today: one principal, 9 teachers, and 2 student teachers.
6. Teacher said, math is very abstract. Students need concrete manipulatives. (This remark and similar remarks in the other districts were prompted by the extensive use of manipulatives in the videotape shown in second session.)
7. Teacher said, children feel it's important to finish their seatwork, even at the expense of accuracy. If seatwork is too long, children tend to get off task. It is helpful to use some children as tutors.
8. Teacher said, she assigned this homework problem - make up a story problem for the other children to solve.
9. Several teachers said, they really get bogged down with hard-to-teach kids.
10. Teacher asked, what is the best time of day to teach math?
11. Teacher said, spending more time on math takes away from other subjects.

Wednesday Morning Live: Observations on a Staff Development Meeting

By Wynn De Bevoise

The boardroom in this elementary-school-turned-district-office looks out on a large warehouse. Almost lost against the structure's gray weathered shingles, pigeons gather and disperse, their meetings brief and fitful.

Inside the boardroom, elementary teachers take their seats slowly and deliberately. It is eight-fifteen in the morning and they are waiting for a training session to begin. As the university professor and the district trainer prepare the audiovisuals, the teachers joke and grumble about being there. They regard the questionnaires they have been asked to fill out with evident distaste. Some push them away. Their faces resemble those of students who feel they've been given an unfair homework assignment. The professor, a consultant for this staff development program, mildly explains the importance of their answers. He does not react to the restrained hostility. As the meeting is ready to begin, most of the teachers jump up to get a cup of coffee from the kitchen down the hall. One laughs and observes that their behavior is "off-task."

By the time everyone returns, the principal has arrived and the atmosphere is more subdued. Some are now filling out the questionnaires that ask for their reactions to the first two weeks of using a new instructional method in their math lessons.

For the next 50 minutes, all participants view a videotape of a teacher using Thomas L. Good and Douglas A. Grouws' Active Teaching of Mathematics system in a fourth-grade classroom. It is Good and Grouws' model that the teachers gathered in this room have been following for the past two weeks. As they watch the videotape, several teachers continue to fill out the questionnaire. They are intent and serious as they write, but they also manage to follow the taped presentation. The principal's attention does not waver from the videotape. He seems to be digesting every activity and interaction in the filmed lesson.

In the exchange of views that follows, the teachers bring up problems they encounter in class that are not addressed in the videotape—discipline, length of time it takes for students to correct each other's papers, difficulties in planning lessons, and grouping problems. They complain that "canned" presentations never quite reflect their own unique classrooms. As each teacher mentions a problem, several others offer suggestions and techniques that work for them.

Initially, the interaction takes place among the teachers without interruption by the trainer, principal, or professor. Increasingly, however, questions are posed directly to the trainer or the professor, and the principal asks questions with greater freedom. He sits between the trainer and the professor, opposite most of the teachers. "The amount of time on task in the videotape was impressive," he remarks. The comment hangs motionless for a moment, but no one suggests that the videotape represents an unachievable ideal.

Discussion returns several times to issues that are clearly of great concern to the teachers. One such issue is homework. Before the district agreed to participate in the present experiment, teachers adhered to a policy of not assigning homework. Under the Good

De Bevoise, Wynn. "Wednesday Morning Live: Observations on a Staff Development Meeting." R&D Perspectives (Winter 1983): 6-8. Eugene, Oregon: Center for Educational Policy and Management, College of Education, University of Oregon.

and Grouws system, they are expected to assign homework daily. In the preceding meeting, there was resistance to this new policy. Teachers predicted that half the students would not do the work and that parents would complain. The principal intervened at this point and reiterated Good and Grouws' specifications for homework assignments: they should be given daily, require no more than 15 minutes to complete, and reinforce skills the students have already acquired. In addition, he signified that the teachers should feel positive about the 50 percent who do complete the assigned work. "View the cup as half full," he remarked, "rather than as half empty."

Apparently the teachers have taken the principal's point to heart and have gained a new perspective on the issue. Now they are not discussing whether they should give assignments to be completed outside of class, but how they can most efficiently handle the whole process. One young woman wonders when to collect homework. "If homework is to be reviewed the first thing during math class," she says, "and math is a fifth-period class, do we collect homework at the beginning of the school day or after the work has been reviewed?" An older male teacher, obviously a veteran of many years, points out that if students are expected to complete their homework outside of class, it has to be collected first thing in the morning. Another man, younger and not afraid to show enthusiasm, suggests making a transition from the preceding class to math by passing back the homework that had been collected at the beginning of the day. "When possible," he adds, "I have already corrected the work by the time I give it back for review. Of course, that's not always possible."

The teachers' comments indicate that left to themselves they might not have devised a consistent method for handling practice work and grading. The necessity imposed by the new instructional method, the guidance provided by the training program, and the principal's reassurance give them a sense of security in attempting new activities. With this support all appear to have established organized ways of coping with the assignment and grading of homework.

"I'm getting a positive response," remarks one teacher. "There's cooperation from home; everyone's expecting homework." She seems slightly surprised by her success. "In fact, I think the kids like the structure of regular homework. They like to know what's coming. And they're eager to see how they've done on last night's assignment."

"Yes," adds another. "And now when a student or the parents come in and demand to know why I gave a low grade, I don't have to worry. I just open my gradebook and it's all right there."

Another issue important to the teachers is that of student grouping. Before the experiment, many had used some form of individualized instruction. Good and Grouws' instructional system emphasizes whole group instruction. All teachers express concern about how to handle classes in which most of the students catch on to new concepts quickly but a few lag behind. Do you aim for mastery by all students, slowing down the pace and risking loss of interest in the brightest group? Or do you quicken the pace and accept the fact that the slowest students will not fully grasp most concepts?

Several teachers use students who understand a lesson to help those who do not understand. The professor assures them that the model is flexible and easily allows for this type of adaptation. He adds other suggestions to help the teachers cope with students who learn at different speeds. "Explain the material in a new way for the students who didn't understand the initial explanation. Sharing a different way to approach a problem deepens everyone's understanding. Or give students different seatwork and homework assignments, depending on their abilities."

Several teachers nod. The question comes up again. "Does the system allow for individualization of activities during the time devoted to seatwork?"

"Absolutely," responds the trainer. "Look, this system is designed to help you. We're not trying to change what you were already doing, we're trying to enhance it." Relief is expressed in smiles around the table.

The principal, who has been reinforcing teachers' suggestions, becomes more actively involved in the discussion at this point. He wants to know how often they use a discovery activity similar to the one demonstrated in the videotape. Their responses seem honest, not bound by what they think he would like to hear. Would it help, he wonders, if teachers gave the "greyhounds" creative work in class rather than homework. In unison, the teachers say no. That type of differentiation would reward the bright students and it is the slow ones that need to be rewarded. He then questions them about manipulatives. "What manipulatives do you have," he asks, "and what ones would you like to have?"

One woman answers that she has some geo boards, which she uses for teaching about fractions.

"Where did you get them?" asks the principal.

"I made them," she responds and describes the materials needed.

Clearly, the principal is becoming familiar with the small details of classroom management under the Good and Grouws system.

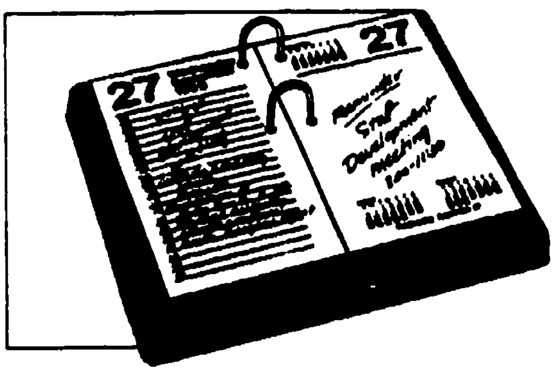
At the conclusion of this meeting, which is the last training session, teachers surprise the trainer by asking about getting together again at the end of the school year to review progress. Their desire to devote further time to discussion indicates that they are finding the training beneficial and belies the reluctant atmosphere with which the session began.

These teachers do not feel they are voluntary participants in the implementation of the Good and Grouws system. The project was approved and mandated by district administrators. It is not surprising, then, that the participants might feel misgivings and some resentment about the training. What has become apparent, however, as the teachers discuss problems and share solutions related to using the model is their increasing desire to see the effects of its use on the work habits and skills of their students. The encouragement and flexibility of the principal, trainer, and professor have helped the teachers move from a position of resistance to one of involvement.

Chattering in relaxed tones, the teachers move into the hallway. The training session has offered them a rare opportunity to talk about problems and common experiences in teaching. They share last-minute information before climbing into their cars, again as isolated as in their separate classrooms.

Meanwhile, in the nearly deserted boardroom, the trainer and the professor review the morning's discussion. They are pleased. The teachers' request to meet again has offered unexpected confirmation that the experiment seems to be working. The assistant superintendent stops by briefly to say that the principal considers the session a clear success.

Despite barred windows, empty chairs, and a series of tables littered with empty coffee cups, the boardroom radiates warmth. This is collegiality on a level unknown to pigeons.



APPENDIX F

A Checklist for Observing Teachers' Use
of the Missouri "Active Teaching" Model in Mathematics

APPENDIX F

A Checklist for Observing Teachers' Use of the Missouri "Active Teaching" Model in Mathematics

Date ____________ Teacher ____________ Observation # ____________

Place a checkmark ( ) in the blank next to all statements that hold true.

Daily Review

_____ reviewed the concepts and skills associated with the homework

_____ collected and dealt with homework assignments

_____ asked several mental computation exercises

_____ spent a total of about 8 minutes in doing all of the above*

Development

_____ briefly focused on prerequisite skills and concepts

focused on meaning and promoting student understanding by:

- ____ explaining the processes used to reach a particular solution or answer
- ____ using concrete illustrations and examples
- ____ using manipulatives

assessed student comprehension

- ____ used process/product questions
- ____ used controlled practice

_____ repeated and elaborated on the meaning portion as necessary

_____ spent a total of about 20 minutes on the development phase of the lesson

Seatwork

_____ provided uninterrupted practice

_____ got everyone involved

_____ continuously monitored student's work and provided feedback to individuals as needed

_____ let students know that their work would be checked at the end of the period

_____ checked students' work at the end of the period

_____ spent a total of about 15 minutes on seatwork

Homework Assignment

_____ assigned homework

* It is intended that on Mondays about 20 minutes will be allocated for reviewing concepts and skills covered during the previous week.

SG417

APPENDIX G

Time Line Observation Form for Recording Good and Grouws' Variables

1-14-83

Observation Form

Teacher's Name________________________________

Your Name________________________________

Date________________________________

District________________________________

No. of Students in Class Today________________________________

Time Math Class Started________________________________

Before class starts, ask teacher:

1. What will your math lesson be about today?

2. Will you be reviewing any of yesterday's work?

3. Last time did you assign any homework to be done out of class? If "yes," will you do anything with the homework today?

4. Will your students be doing any seatwork today? If "yes," what are acceptable things for the children to do if they finish early?

5. Will any of the children be leaving during the class today?

6. Will today be a typical math lesson for you?

ransition nterruption o. of students off-task	T I	T I	T I	T I	T I	T I	T I	T I	T I	___	T I	T I	T I	T I	T I
eview previous work heck prior homework ssign homework uiz	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q		R C A Q	R C A Q	R C A Q	R C A Q	R C A Q
ental computation evelopment ontrolled practice in dev	M D P	M D P	M D P	M D P	M D P	M D P	M D P	M D P	M D P		M D P	M D P	M D P	M D P	M D P
irections for Seatwork eatwork (monitored +) heck seatwork at end	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E		S + - E	S + - E	S + - E	S + - E	S + - E

	8:00	8:30	9:00	9:30	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30	14:00	14:30	15:00
ransition nterruption o. of students off-task	T I	T I	T I	T I	___	T I	T I	T I	T I	T I	T I	T I	T I	T I	___
eview previous work heck prior homework ssign homework uiz	R C A Q	R C A Q	R C A Q	R C A Q		R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	
ental computation evelopment ontrolled practice in dev	M D P	M D P	M D P	M D P		M D P	M D P	M D P	M D P	M D P	M D P	M D P	M D P	M D P	
irections for seatwork eatwork (monitored +) heck seatwork at end	S + - E	S + - E	S + - E	S + - E		S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	

	15:30	16:00	16:30	17:00	17:30	18:00	18:30	19:00	19:30	20:00	20:30	21:00	21:30	22:00	22:30
Transition Interruption No. of students off-task	T I	T I	T I	T I	T I	T I	T I	T I	T I	———	T I	T I	T I	T I	T I
Review previous work Check prior homework Assign homework Quiz	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q		R C A Q	R C A Q	R C A Q	R C A Q	R C A Q
Mental computation Development Controlled practice in dev	M D P	M D P	M D P	M D P	M D P	M D P	M D P	M D P	M D P		M D P	M D P	M D P	M D P	M D P
Directions for Seatwork Seatwork (monitored +) Check seatwork at end	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E		S + - E	S + - E	S + - E	S + - E	S + - E

184

	23:00	23:30	24:00	24:30	25:00	25:30	26:00	26:30	27:00	27:30	28:00	28:30	29:00	29:30	30:00
Transition Interruption No. of students off-task	T I	T I	T I	T I	———	T I	T I	T I	T I	T I	T I	T I	T I	T I	———
Review previous work Check prior homework Assign homework Quiz	R C A Q	R C A Q	R C A Q	R C A Q		R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	R C A Q	
Mental computation Development Controlled practice in dev	M D P	M D P	M D P	M D P		M D P	M D P	M D P	M D P	M D P	M D P	M D P	M D P	M D P	
Directions for seatwork Seatwork (monitored +) Check seatwork at end	S + - E	S + - E	S + - E	S + - E		S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	S + - E	

	30:30	31:00	31:30	32:00	32:30	33:00	33:30	34:00	34:30	35:00	35:30	36:00	36:30	37:00	37:30
ransition	T	T	T	T	T	T	T	T	T		T	T	T	T	T
nterruption	I	I	I	I	I	I	I	I	I		I	I	I	I	I
o. of students off-task										___					
eview previous work	R	R	R	R	R	R	R	R	R		R	R	R	R	R
heck prior homework	C	C	C	C	C	C	C	C	C		C	C	C	C	C
ssign homework	A	A	A	A	A	A	A	A	A		A	A	A	A	A
uiz	Q	Q	Q	Q	Q	Q	Q	Q	Q		Q	Q	Q	Q	Q
ental computation	M	M	M	M	M	M	M	M	M		M	M	M	M	M
evelopment	D	D	D	D	D	D	D	D	D		D	D	D	D	D
ontrolled practice in dev	P	P	P	P	P	P	P	P	P		P	P	P	P	P
irections for Seatwork	S	S	S	S	S	S	S	S	S		S	S	S	S	S
eatwork (monitored +)	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -		+ -	+ -	+ -	+ -	+ -
heck seatwork at end	E	E	E	E	E	E	E	E	E		E	E	E	E	E

	38:00	38:30	39:00	39:30	40:00	40:30	41:00	41:30	42:00	42:30	43:00	43:30	44:00	44:30	45:00
ransition	T	T	T	T		T	T	T	T	T	T	T	T	T	
nterruption	I	I	I	I		I	I	I	I	I	I	I	I	I	
o. of students off-task					___										___
eview previous work	R	R	R	R		R	R	R	R	R	R	R	R	R	
heck prior homework	C	C	C	C		C	C	C	C	C	C	C	C	C	
ssign homework	A	A	A	A		A	A	A	A	A	A	A	A	A	
uiz	Q	Q	Q	Q		Q	Q	Q	Q	Q	Q	Q	Q	Q	
ental computation	M	M	M	M		M	M	M	M	M	M	M	M	M	
evelopment	D	D	D	D		D	D	D	D	D	D	D	D	D	
ontrolled practice in dev	P	P	P	P		P	P	P	P	P	P	P	P	P	
irections for seatwork	S	S	S	S		S	S	S	S	S	S	S	S	S	
eatwork (monitored +)	+ -	+ -	+ -	+ -		+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	
heck seatwork at end	E	E	E	E		E	E	E	E	E	E	E	E	E	

	45:30	46:00	46:30	47:00	47:30	48:00	48:30	49:00	49:30	50:00	50:30	51:00	51:30	52:00	52:30
ransition	T	T	T	T	T	T	T	T	T		T	T	T	T	T
nterruption	I	I	I	I	I	I	I	I	I		I	I	I	I	I
lo. of students off-task										___					
eview previous work	R	R	R	R	R	R	R	R	R		R	R	R	R	R
heck prior homework	C	C	C	C	C	C	C	C	C		C	C	C	C	C
ssign homework	A	A	A	A	A	A	A	A	A		A	A	A	A	A
uiz	Q	Q	Q	Q	Q	Q	Q	Q	Q		Q	Q	Q	Q	Q
lental computation	M	M	M	M	M	M	M	M	M		M	M	M	M	M
evelopment	D	D	D	D	D	D	D	D	D		D	D	D	D	D
ontrolled practice in dev	P	P	P	P	P	P	P	P	P		P	P	P	P	P
irections for Seatwork	S	S	S	S	S	S	S	S	S		S	S	S	S	S
eatwork (monitored +)	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -		+ -	+ -	+ -	+ -	+ -
heck seatwork at end	E	E	E	E	E	E	E	E	E		E	E	E	E	E

BEST COPY AVAILABLE

	53:00	53:30	54:00	54:30	55:00	55:30	56:00	56:30	57:00	57:30	58:00	58:30	59:00	59:30	60:00
ransition	T	T	T	T		T	T	T	T	T	T	T	T	T	
nterruption	I	I	I	I		I	I	I	I	I	I	I	I	I	
lo. of students off-task					___										___
eview previous work	R	R	R	R		R	R	R	R	R	R	R	R	R	
heck prior homework	C	C	C	C		C	C	C	C	C	C	C	C	C	
ssign homework	A	A	A	A		A	A	A	A	A	A	A	A	A	
uiz	Q	Q	Q	Q		Q	Q	Q	Q	Q	Q	Q	Q	Q	
lental computation	M	M	M	M		M	M	M	M	M	M	M	M	M	
evelopment	D	D	D	D		D	D	D	D	D	D	D	D	D	
ontrolled practice in dev	P	P	P	P		P	P	P	P	P	P	P	P	P	
irections for seatwork	S	S	S	S		S	S	S	S	S	S	S	S	S	
eatwork (monitored +)	+ -	+ -	+ -	+ -		+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	
heck seatwork at end	E	E	E	E		E	E	E	E	E	E	E	E	E	

1-14-83

Post Observation Form

1. Student comprehension when teacher explained concepts and procedures.

   ____ Students followed teacher's explanation with no confusion.

   ____ Good comprehension, with occasional confusion.

   ____ Lots of confusion.

   ____ Not applicable

2. Student comprehension during seatwork.

   ____ Students worked independently all or most of the time.

   ____ Students occasionally asked for teacher assistance.

   ____ Heavy demand for teacher assistance.

   ____ Not applicable.

3. Were word problems included in the lesson?

   ____ No.

   ____ Yes, for 2 minutes or less.

   ____ Yes, for more than 2 minutes.

4. Notes on classroom organization and unusual occurrences.

APPENDIX H

Manual for Scoring Good and Grouws' Variables

MANUAL FOR SCORING GOOD AND GROUWS VARIABLES

Definition of Variables

Transition (T)

Transition refers to the process of going from one phase of the class period to another. Transition also involves (a) the period of time from when the class is scheduled to begin and when it actually gets productively underway; (b) when the teacher signals end of class is approaching and students start putting things away before the lesson has actually ended; and (c) the distribution or collection of seatwork, homework, and other materials.

Interruptions (I)

Interruptions include such events as fire alarms, intercom messages, and classroom discipline problems.

Review of Previous Work and Quizzes (R)

Review refers to work on old material. It deals with concepts and skills that the students have learned prior to the lesson being observed. The concepts and skills can be from the previous day or from much earlier. Use of games, puzzles, worksheets, etc., which are used to review concepts and skills fall into this category. Review of previously-taken quizzes and tests and giving of quizzes are coded as Q.

Checking of Prior Homework (C)

This variable is similar to R, but is limited to review of homework previously assigned. The "checking" process can include in-class grading of homework, the showing of solutions to homework problems, and answering of questions about the homework.

Assignment of Homework (A)

This variable is an index of the time spent giving an homework assignment. It includes such activities as distributing homework problems and giving homework directions.

Quiz (Q)

This variable is the time spent giving a quiz in class.

Mental Computation (M)

This variable refers to time spent on computation without the aid of paper-and-pencil, blackboard, or calculators. The computations are done mentally. If the computations are not done mentally, code as "Seatwork" or "Controlled Practice Within Development." The problems that students are asked to solve mentally may refer to new or old computational skills.

Development (D)

The development part of a math lesson involves helping students comprehend new concepts and skills. Development activities often include teacher explanations and demonstrating use of manipulative materials, making of comparisons, search for patterns, class discusion, group work, and use of audio-visual materials. Brief review of prerequisite concepts and skills may occur, but this activity is counted as time spent on development if its purpose is to facilitate the development part of a lesson rather than to review previous work (R).

Controlled Practice Within Development (P)

This variable includes controlled practice in which one or two problems are stated and then immediate feedback is given on the correctness of the responses. Controlled practice is generally used to help the teacher identify and correct student misunderstandings just before seatwork to make sure the students can do the seatwork successfully.

Directions for Seatwork (S)

This variable refers to time spent on giving directions to students on how to do their seatwork. Time spent by students getting ready to do seatwork (e.g., getting materials ready) should be coded as a transition (T).

Seatwork (+ or -)

Seatwork refers to practice on concepts and skills covered in the day's lesson. It may include some review of previous work. If the problems are entirely a review of previous work, though, the activity should be coded as "Review of Previous Work" (R). If the teacher is monitoring the seatwork or assisting students during the time period being observed, it is coded as "+". If the teacher is not doing either of these activities, it is coded as "-".

Checking of Seatwork at End (E)

This variable includes such activities as the teacher alerting students that their seatwork will be checked, reading correct answers to the seatwork problems, or asking students to check their answers in some way.

Number of Students Off Task

A student is on task if he is engaged in an activity assigned by the teacher or that is academically appropriate to the situation. Examples of on-task activities are: listening to the teacher, doing seatwork, answering the teacher's questions, working at the blackboard. Examples of off task activities are: chatting with another student about a nonacademic matter, waiting for the teacher's help during seatwork, daydreaming, standing around, waiting for the lesson to start or end.

Decision Rules for Observing Good and Grouws' Variables

1. A math lesson begins when the teacher signals that it has begun. Some students may not be at their seats or even in the room. Some students may be at their desks, but are not ready to begin their math work. Code the interval between the teacher's signal and students' setting down to work as TRANSITION.

2. A math lesson ends when the teacher indicates that it has ended. For example, "put your math books away now." Do not code any TRANSITION time after the end-of-lesson signal has occurred.

3. All of the codes except SEATWORK are reserved for whole group instruction. Individualized math instruction is to be coded as SEATWORK, even when the teacher works with a small group of students or with several small groups.

4. If a student teacher is in the classroom, do not code anything that he or she or his or her students are doing. Do not count the students whom the student teacher is instructing as part of the total number of students in the class. Also, do not code these students as on or off task. However, do describe this situation in your post-observation notes.

5. Code REVIEW (R) only when the teacher is reviewing the prior day's work. The review should be in the form of a summary of what was studied the previous day. If the review refers to an unspecified time frame or a time frame other than yesterday, it is likely that DEVELOPMENT (D) is occurring. If the teacher starts with a review of yesterday's work, but proceeds to reteach the content, start by coding REVIEW and then switch to DEVELOPMENT.

6. CHECK PRIOR HOMEWORK refers specifically to checking the answers to homework. This code applies to the teacher giving or asking for the correct answers to the homework. It also applies to stating or asking how the answers were obtained. If you are unsure whether the teacher is checking homework or seatwork, ask the teacher to clarify when the lesson is over. Change your coding if necessary.

7. CHECK SEATWORK AT END (E) will usually occur at the end of a lesson. However, it can also occur at the beginning of the lesson if the teacher goes over the previous day's seatwork. If the teacher collects seatwork without saying anything, ask teacher after lesson if she intended to check the seatwork and give feedback to the students; if yes, circle two minutes of Es.

8. Use the ASSIGN HOMEWORK (A) only if homework is assigned to be done out of class. A teacher may assign homework, but allow students to complete it in class. The ASSIGN HOMEWORK code should only be used if there is an expectation that the students will be doing math work outside of the class or some time later after the math lesson is ended. Ask the teacher what his or her intentions are if you are not certain.

9. MENTAL COMPUTATION refers specifically to quick computations that the student does in his or her head. The student can write the answers on paper, but he or she should not do the computations on paper. If the computations are done on paper, it is likely that the QUIZ code should be used.

10. CONTROLLED PRACTICE should be used only when: the teacher assigns one or a few problems to be done; the teacher checks the accuracy of students' answers right away; the problems occur near the end of the development period (unless students can't do the problems and the teacher needs to do further development); and the problems should be of the same type as will appear in seatwork. All of these conditions should be present for the CONTROLLED PRACTICE (P) code to be used.

11. TRANSITION (T) occurs whenever students are waiting for something to occur (e.g., waiting for the teacher's help), or are getting ready to do something.

12. Students are OFF TASK whenever they are not engaged in instructionally relevant behavior that meets the teacher's expectations.

13. Whole class instruction occurs only when teacher is working with at least three-fourths of the class in a group activity. Otherwise it is seatwork.

14. If teacher is instructing a small group of the students (less than one-fourth) while other students are doing seatwork, count as S-. The exception is if teacher stops working with subgroup and assists a seatwork student (code S+).

15. If you observe a teacher on Friday, ask the teacher after class whether she normally assigns homework. If yes, code one homework interval.

16. Ask the teacher after class whether she collected and checked homework earlier in the day. (Do this only if no homework collection or checking occurred.) If yes, have teacher estimate time and fill in appropriate number of "check prior homework" intervals.

17. If the teacher is engaged in "T" or "I" behavior while you are doing an "off task" tally, count all students as off task because they are not engaged in instructionally relevant behavior.

APPENDIX I

Rationale for Time Criteria for Behavior Categories in Good and Grouws' Instructional Model

APPENDIX I

Rationale for Time Criteria for Behavior Categories in Good and Grouws' Instructional Model

The rationale for the time criteria used to analyze the data from the experiment was derived primarily from the Teachers Manual for Good and Grouws' training program. The Teachers Manual is included in a report by Good and Grouws (1981). The essential behaviors described in the manual are summarized in Table 2-2.

Quoted material in the following sections is from the Teachers Manual, using the pagination of the Good and Grouws (1981) report.

1. TRANSITIONS - 2 minutes or less

The need to minimize transitions is only mentioned specifically with respect to seatwork. "To help optimize the effectiveness of seatwork, three general principles should be observed... The first principle, momentum, ...mean[s] keeping the ball rolling without any sharp break in teaching activity and in student involvement" (p. 103). The need to minimize transitions is also implied in the section on instructional pacing (p. 120).

The Teachers Manual states no time criteria for transitions. Since some transitional time seems inevitable, the decision was made to use the pretreatment mean as a minimal criterion. This mean is 1.94 minutes (see table 4.2), and was rounded off to 2 minutes.

2. INTERRUPTIONS - 0 minutes

Although interruptions are not discussed explicitly in the Teachers Manual, the desirability of avoiding them is implied in the discussion of instructional momentum (p. 103) and pacing (p. 120). Since it seems possible to avoid interruptions completely with careful instructional planning, 0 minutes was set as the criterion.

3. REVIEW PREVIOUS WORK - 1 minute or more

The Teachers Manual recommends "1-2 minutes on review" (p. 122).

4. CHECK PRIOR HOMEWORK - 2 minutes or more

"... it should take only a couple of minutes to check homework" (p. 121). "In general, we think the following situation will be most applicable... 3-4 minutes checking homework" (p. 122). The more modest figure of 2 minutes implied by "couple of minutes" was used as the criterion for this behavior category.

5. ASSIGN HOMEWORK - 1 second or more

The Teachers Manual does not contain a recommendation on the amount of time that assigning homework should take. It is only important that the behavior occur, and this is reflected by the "1 second or more" criterion.

6. QUIZ - 3 minutes or less

This behavior category is not mentioned in the Teachers Manual, except in the section on cumulative monthly reviews (p. 115). It is mentioned there as just one of several methods of review. This category was included in the classroom observation instrument, however, because it occurred frequently in pilot observations of teachers' lessons. The mean length of this category in the pretreatment lesson was 2.94 minutes, and this score (rounded to 3 minutes) was used as the criterion. It seemed implicit in Good and Grouws' model that behaviors like quizzes should not dominate a math lesson. A quiz longer than 3 minutes might leave less time for other, more important behaviors advocated in their instructional model.

7. MENTAL COMPUTATION - 3 minutes or more

The Teachers Manual states, "We would like for you to include 3-5 minutes on mental computation activities each day..." (p. 117).

8. DEVELOPMENT - 5 minutes or more

The summary table on page 94 in the Teachers Manual (reproduced in the present report as table 2-) recommends "about 20 minutes" for development.

However, the data from the present experiment suggests that this time criterion is unrealistically high. In fact, the treatment implementation analysis in Good and Grouws' fourth-grade study includes the percentage of teachers who spent "at least five minutes on development" (see table of the present report). This lesser time criterion was selected for use in the present study.

9. CONTROLLED PRACTICE - 1 minute or more

The Teachers Manual does not make a specific time recommendation for this behavior category. There is mention that controlled practice (defined in the present study as the assignment of practice problems) and oral questioning "can be completed in 3-5 minutes" (p. 97). Controlled practice can be very brief if students have understood the teacher's prior demonstration. Therefore, a minimal time criterion of "1 minute or more" was used to indicate satisfactory use of this behavior.

10. DIRECTIONS FOR SEATWORK - 1 minute or less

The Teachers Manual states that the teacher should initiate seatwork "with a simple and direct statement" (p. 103). The treatment implementation analysis in Good and Grouws' fourth-grade study included a variable described as, "Did seatwork directions take longer than one minute?"

11. MONITORED SEATWORK - 15 minutes or less

The Teachers Manual states, "...we recommend that about 10-15 minutes each day be allotted for seatwork. Ten to fifteen minutes allows sufficient time for students to work enough problems to achieve increased proficiency but not so long as to bring about boredom, lack of task involvement...If practice time does not exceed 15 minutes, few students are likely to be bored" (p. 102).

12. UNMONITORED SEATWORK - 2 minutes or less

The Teachers Manual does not state explicitly that teachers should avoid unmonitored seatwork. It is strongly implied, however, in the discussion of seatwork (pp. 101-106). It seems unrealistic to expect teachers to monitor seatwork one hundred percent of the time, so a less strict criterion was used. The decision was made to use the lowest mean amount of unmonitored seatwork achieved by a treatment group in the pre-observation. This mean was 2.39 minutes, achieved by the regular inservice group in Districts 2-3. This figure was rounded to 2 minutes in determining the time criterion.

12. CHECK SEATWORK AT END - 1 minute or more

The Teachers Manual states, "At the very end of the seatwork period, hold students accountable for their work by asking individual students to give the answer to a few of the assigned problems. This checking of answers should be very rapid and you need only check 3 or 4 of the problems..." (pp. 105-106). A more specific time criterion is not stated. Therefore, the decision was made to use the pretreatment mean as a minimal criterion. The actual mean was 1.27 minutes, which was rounded to 1 minute to provide the criterion.

APPENDIX J

Supplementary Tables

211

TABLE A-1
Adjusted Post and Delayed Treatment Means for Instructional Behaviors by District

Direct Instruction Behavior	Time of Observation	Treatment Means Adjusted by Pre Scores on Corresponding Variable								
		District 1			District 2			District 3		
		Principal Involvement (N=9)	Regular Inservice (N=10)	Control (N=10)	Principal Involvement (N=6)	Regular Inservice (N=4)	Control (N=4)	Principal Involvement (N=4)	Regular Inservice (N=3)	Contr (N-4
Transitions	Post	3.41	4.08	2.06	3.44	1.75	2.52	3.24	2.77	2.11
	Delayed	3.20	2.70	3.31	3.25	4.08	1.28	2.11	2.70	1.66
Interruptions	Post	1.41	.15	.29	.30	.27	.25	.28	.21	.84
	Delayed	.72	.17	.86	.30	1.71	.32	.39	.03	.82
Review Previous Work	Post	5.31	1.83	6.66	1.78	4.47	6.48	2.25	3.35	8.24
	Delayed	2.12	1.46	.00	1.44	.98	1.22	2.56	.87	3.24
Check Prior Homework	Post	2.89	4.85	.00	2.83	6.95	.08	5.58	.91	.08
	Delayed	1.92	2.71	1.17	1.91	7.32	.05	5.07	1.69	.05
Assign Homework	Post	.31	.39	.05	.71	1.87	.00	.70	.97	.00
	Delayed	.44	.42	.00	1.17	.58	.22	1.08	.33	.01
Quiz	Post	.00	3.37	1.61	.09	.00	.00	1.99	.00	6.15
	Delayed	1.02	1.07	1.76	.64	.69	1.85	.21	9.23	.80
Mental Computation	Post	1.90	2.06	.24	1.46	5.45	.00	3.69	3.86	1.62
	Delayed	1.62	2.20	.27	1.78	2.37	.00	4.08	1.45	1.59
Development	Post	4.36	9.55	4.21	11.20	16.16	10.22	11.31	15.11	7.81
	Delayed	16.43	7.79	8.74	19.53	14.70	16.73	9.25	6.54	22.13
Controlled Practice	Post	.93	1.73	1.76	4.99	2.72	.12	1.87	3.04	.00
	Delayed	1.51	1.69	.90	.55	5.20	.00	.46	3.54	.01
Directions for Seatwork	Post	1.60	2.47	1.62	1.83	.37	2.35	1.00	.80	.92
	Delayed	1.26	1.94	.93	.76	1.04	1.19	.74	1.17	.46
Monitored Seatwork	Post	7.46	9.23	13.61	9.00	4.04	17.62	12.31	10.13	5.78
	Delayed	6.27	13.17	12.71	8.59	4.20	12.50	11.55	16.38	11.38
Unmonitored Seatwork	Post	10.82	1.72	9.50	2.19	2.12	3.79	.83	1.22	6.31
	Delayed	.00	3.16	5.48	1.18	1.33	3.70	1.08	1.02	1.18
Check Seatwork at End	Post	1.37	.37	1.45	.76	1.48	.00	1.04	1.35	.60
	Delayed	1.72	.74	2.60	2.17	.53	.44	.84	1.92	.06

TABLE A-2
Number of Techniques Used to Criterion Level
by Each Treatment Group in Each District

District 1

	Principal Involvement (N=9)		Regular Inservice (N=10)		Control (N=10)	
	M	(SD)	M	(SD)	M	(SD)
Pre	4.33	(1.32)	4.30	(2.36)	5.00	(1.56)
Post	7.00	(2.69)	6.40	(2.17)	5.50	(1.84)
Delayed	8.11	(2.09)	6.90	(1.91)	4.90	(.74)

Districts 2-3

	Principal Involvement (N=9)		Regular Inservice (N=7)		Control (N=8)	
	M	(SD)	M	(SD)	M	(SD)
Pre	5.56	(1.67)	6.17	(2.69)	5.38	(1.92)
Post	8.44	(1.59)	9.86	(1.21)	5.00	(2.39)
Delayed	8.56	(1.33)	8.29)	(2.87)	5.88	(1.96)

TABLE A-3
Percentage of Teachers Using Instructional
Behaviors for Criterion Time Periods

		District 1			Districts 2-3		
		Principal Involvement (N=9)	Regular Inservice (N=10)	Control (N=10)	Principal Involvement (N=9)	Regular Inservice (N=7)	Control (N=8)
Transitions (2 minutes or less)	Pre	78%	40%	50%	44%	71%	63%
	Post	33%	0%	70%	33%	71%	63%
	Delayed	11%	50%	60%	56%	14%	88%
Interruptions (no interruptions)	Pre	44%	50%	40%	56%	71%	50%
	Post	44%	70%	70%	67%	57%	50%
	Delayed	33%	70%	50%	67%	71%	50%
Review Previous Work (1 minute or more)	Pre	11%	10%	30%	44%	71%	75%
	Post	56%	10%	60%	67%	71%	75%
	Delayed	56%	70%	0%	44%	43%	50%
Check Prior Homework (at least 2 minutes)	Pre	0%	0%	20%	11%	0%	0%
	Post	44%	60%	0%	89%	57%	0%
	Delayed	67%	60%	10%	56%	71%	0%
Assign Homework (1 second or more)	Pre	11%	20%	20%	0%	14%	0%
	Post	67%	50%	10%	89%	86%	0%
	Delayed	78%	70%	0%	78%	71%	13%
Quiz (3 minutes or less)	Pre	89%	70%	40%	44%	86%	75%
	Post	100%	60%	90%	89%	100%	75%
	Delayed	100%	80%	80%	78%	86%	88%
Mental Computation (at least 2 minutes)	Pre	0%	0%	10%	11%	29%	38%
	Post	44%	50%	10%	56%	86%	13%
	Delayed	44%	40%	10%	56%	57%	13%
Development (at least 5 minutes)	Pre	33%	30%	50%	67%	71%	38%
	Post	44%	70%	30%	78%	100%	50%
	Delayed	89%	70%	40%	89%	86%	63%
Controlled Practice (at least 1 minute)	Pre	0%	40%	20%	67%	57%	50%
	Post	22%	50%	20%	44%	57%	0%
	Delayed	44%	30%	10%	22%	57%	0%
Directions for Seatwork (1 minute or less)	Pre	56%	90%	90%	56%	29%	38%
	Post	67%	40%	50%	33%	71%	38%
	Delayed	56%	40%	70%	78%	57%	75%
Monitored Seatwork (15 minutes or less)	Pre	56%	40%	60%	67%	43%	63%
	Post	100%	80%	60%	100%	100%	75%
	Delayed	100%	70%	70%	100%	86%	75%
Unmonitored Seatwork (2 minutes or less)	Pre	44%	40%	50%	67%	71%	63%
	Post	44%	60%	50%	89%	86%	50%
	Delayed	100%	50%	60%	89%	86%	75%
Check Seatwork at End (at least 1 minute)	Pre	11%	0%	20%	22%	57%	13%
	Post	44%	30%	40%	22%	57%	13%
	Delayed	56%	10%	30%	56%	57%	0%

TABLE A-4

Length in Time (in Minutes) for Instructional Behaviors by Treatment Group Within Districts

Instructional Behaviors		District 1: Principal Involvement (N=9) M (SD)	District 1: Regular Inservice (N=10) M (SD)	District 1: Control (N=10) M (SD)	Districts 2-3: Principal Involvement (N=9) M (SD)	Districts 2-3: Regular Inservice (N=7) M (SD)	Districts 2-3: Control (N=8) M (SD)
Transitions	Pre	1.24 (1.38)	2.49 (1.89)	1.79 (1.46)	2.28 (1.41)	2.04 (2.00)	1.79 (1.59)
	Post	3.32 (1.69)	4.17 (1.31)	2.03 (1.89)	3.39 (2.85)	2.20 (2.13)	2.29 (1.61)
	Delayed	2.98 (1.35)	2.87 (2.21)	3.26 (3.60)	2.84 (2.86)	3.51 (1.60)	1.41 (1.31)
Interruptions	Pre	.61 (.70)	1.06 (1.86)	1.09 (2.04)	.58 (.89)	.29 (.50)	1.58 (2.06)
	Post	1.39 (1.93)	.17 (.33)	.31 (.72)	.28 (.51)	.21 (.27)	.59 (.69)
	Delayed	.71 (1.02)	.18 (.30)	.87 (1.39)	.33 (.56)	.97 (1.85)	.60 (1.19)
Review Previous Work	Pre	.75 (2.16)	.70 (2.04)	2.18 (4.05)	2.71 (4.14)	3.43 (3.36)	.94 (1.32)
	Post	4.70 (6.28)	1.20 (2.34)	6.78 (11.36)	2.37 (2.21)	4.74 (6.04)	6.85 (8.08)
	Delayed	2.14 (2.17)	1.48 (1.51)	.00 (.00)	1.92 (2.90)	.90 (1.14)	2.25 (3.92)
Check Prior Homework	Pre	.00 (.00)	.20 (.48)	1.15 (2.43)	.50 (1.50)	.07 (.19)	.00 (.00)
	Post	2.81 (3.94)	4.83 (2.97)	.00 (.00)	4.12 (3.16)	4.31 (4.97)	.00 (.00)
	Delayed	1.87 (1.44)	2.70 (2.83)	1.33 (4.03)	3.36 (3.11)	4.87 (5.26)	.00 (.00)
Assign Homework	Pre	.06 (.17)	.15 (.34)	.08 (.18)	.00 (.00)	.07 (.19)	.25 (.46)
	Post	.31 (.24)	.39 (.50)	.05 (.16)	.71 (.49)	1.49 (1.60)	.00 (.00)
	Delayed	.43 (.30)	.43 (.37)	.00 (.00)	1.12 (1.21)	.47 (.44)	.13 (.35)
Quiz	Pre	1.06 (3.17)	2.50 (4.75)	6.29 (6.88)	4.11 (4.24)	.88 (2.32)	1.89 (3.09)
	Post	.00 (.00)	3.38 (4.90)	1.50 (4.74)	.89 (2.67)	.00 (.00)	3.06 (6.21)
	Delayed	.00 (.00)	1.03 (1.98)	4.29 (10.68)	1.50 (3.28)	3.21 (8.50)	.88 (1.94)

Table A-4 Continued

Mental Computation	Pre	.00 (.00)	.16 (.41)	1.68 (5.30)	.70 (2.10)	1.79 (3.32)	1.85 (3.19)
	Post	1.87 (2.35)	2.03 (2.36)	.25 (.79)	2.44 (2.38)	4.79 (5.11)	.81 (2.30)
	Delayed	1.50 (1.77)	2.20 (3.51)	.33 (1.03)	2.76 (2.90)	2.04 (2.06)	.81 (2.30)
Development	Pre	6.53 (10.56)	5.18 (9.04)	4.25 (4.77)	7.41 (7.14)	8.49 (7.24)	7.23 (10.57)
	Post	4.19 (4.35)	8.11 (5.53)	5.80 (8.78)	11.42 (9.39)	17.77 (8.89)	7.03 (7.72)
	Delayed	16.32 (8.73)	7.19 (5.54)	7.80 (10.20)	15.18 (9.21)	11.41 (8.80)	19.58 (16.80)
Controlled Practice	Pre	.00 (.00)	4.16 (6.10)	1.78 (4.11)	2.87 (3.91)	3.46 (3.53)	2.81 (3.87)
	Post	.76 (1.53)	1.83 (2.66)	1.70 (3.62)	3.62 (5.02)	2.91 (3.14)	.06 (.18)
	Delayed	1.67 (2.33)	1.60 (2.81)	.95 (3.00)	.50 (1.00)	4.44 (4.58)	.00 (.00)
Directions for Seatwork	Pre	1.36 (1.52)	.48 (.53)	.59 (.77)	1.66 (2.09)	1.86 (1.26)	1.79 (1.61)
	Post	1.57 (1.77)	2.17 (2.87)	1.36 (1.52)	1.53 (1.11)	.68 (.59)	1.74 (1.79)
	Delayed	1.26 (.79)	1.86 (2.49)	.86 (1.08)	.77 (.38)	1.13 (.99)	.85 (.70)
Monitored Seatwork	Pre	17.66 (14.36)	18.50 (14.16)	15.83 (14.88)	10.45 (7.67)	18.08 (12.86)	12.69 (8.50)
	Post	7.76 (3.95)	9.62 (4.69)	13.72 (13.12)	10.02 (3.73)	6.99 (5.06)	11.48 (8.83)
	Delayed	7.31 (3.67)	14.52 (10.14)	13.08 (12.40)	8.31 (3.31)	10.61 (10.62)	11.16 (8.85)
Unmonitored Seatwork	Pre	8.97 (10.06)	7.08 (7.75)	8.84 (10.97)	2.94 (3.28)	2.39 (3.68)	6.81 (9.67)
	Post	11.64 (15.17)	2.09 (2.12)	10.29 (11.78)	.95 (.93)	.96 (1.23)	5.35 (6.76)
	Delayed	.37 (.58)	3.39 (3.70)	5.98 (11.72)	.73 (1.06)	.71 (1.08)	2.63 (5.03)
Check Seatwork at End	Pre	2.81 (8.14)	.05 (.16)	1.38 (3.93)	1.22 (2.40)	2.54 (2.88)	.31 (.53)
	Post	1.24 (1.69)	.48 (.56)	1.45 (1.98)	.89 (1.09)	1.32 (1.31)	.34 (.52)
	Delayed	1.81 (1.95)	.66 (1.64)	2.60 (3.97)	1.57 (1.45)	1.20 (1.02)	.19 (.26)

216

TABLE A-5
Descriptive Statistics for Curriculum-Referenced Test
by Treatment Group Within District and Grade Level

District and Grade Level	Principal Involvement			Regular Inservice			Control		
		Pre Test	Post Test		Pre Test	Post Test		Pre Test	Post Test
	N	M(SD)	M(SD)	N	M(SD)	M(SD)	N	M(SD)	M(SD)
District 1 Grade 4	2	33.76 (7.52)	42.18 (10.77)	5	36.98 (7.78)	43.61 (9.58)	5	39.41 (4.71)	48.50 (2.50
		Gain = 8.42			Gain = 6.63			Gain = 9.09	
District 1 Grade 5	5	25.76 (9.37)	34.08 (7.19)	4	38.33 (5.53)	40.53 (7.19)	5	37.82 (5.58)	42.42 (7.27
		Gain = 8.32			Gain = 2.20			Gain = 4.60	
Districts 2-3 Grade 4	4	38.48 (5.16)	48.41 (1.15)	3	42.44 (8.49)	49.64 (13.28)	5	40.54 (5.84)	44.81 (4.80
		Gain = 9.93			Gain = 7.20			Gain = 4.27	
Districts 2-3 Grade 5	5	37.78 (8.36)	42.22 (8.99)	4	42.03 (6.38)	47.45 (7.74)	3	36.24 (2.09)	40.18 (2.08
		Gain = 4.44			Gain = 5.42			Gain = 3.94	

217

TABLE A-6

Descriptive Statistics for Standardized Achievement Test (Total Score) by Treatment Group Within District and Grade Level

District and Grade Level	Principal Involvement			Regular Inservice			Control		
		Pre Test	Post Test		Pre Test	Post Test		Pre Test	Post Test
	N	M(SD)	M(SD)	N	M(SD)	M(SD)	N	M(SD)	M(SD)
District 1 Grade 4	2	48.20 (1.89)	54.31 (1.51)	5	51.56 (5.29)	65.16 (5.90)	5	53.09 (4.60)	64.73 (4.26)
	Gain = 6.11			Gain = 13.60			Gain = 11.64		
District 1 Grade 5	5	44.82 (8.12)	52.68 (9.63)	4	50.53 (6.27)	59.24 (3.98)	5	51.01 (6.69)	58.23 (5.19)
	Gain = 7.86			Gain = 8.71			Gain = 7.22		
Districts 2-3 Grade 4	4	50.13 (4.47)	57.88 (1.31)	3	49.48 (5.04)	59.62 (2.5)	5	52.42 (2.78)	51.94 (5.77)
	Gain = 7.75			Gain = 10.14			Gain = -.48		
Districts 2-3 Grade 5	5	51.46 (13.22)	56.65 (14.51)	4	55.66 (3.46)	55.75 (4.91)	3	44.27 (9.53)	48.27 (2.78)
	Gain = 5.19			Gain = .09			Gain = 4.00		

TABLE A-7
Descriptive Statistics for Standardized Achievement Test
(Computation Score) by Treatment Group Within District and Grade Level

District and Grade Level	Principal Involvement			Regular Inservice			Control		
		Pre Test	Post Test		Pre Test	Post Test		Pre Test	Post Test
	N	M(SD)	M(SD)	N	M(SD)	M(SD)	N	M(SD)	M(SD)
District 1 Grade 4	2	45.45	51.85	5	50.54	68.46	5	53.12	67.93
		(4.07)	(1.82)		(6.45)	(7.31)		(5.21)	(6.96)
		Gain = 6.40			Gain = 17.92			Gain = 14.81	
District 1 Grade 5	5	40.96	51.27	4	46.58	60.94	5	49.68	59.35
		(/.61)	(12.21)		(6.72)	(3.96)		(7.13)	(5.90)
		Gain = 10.31			Gain = 14.36			Gain = 9.67	
Districts 2-3 Grade 4	4	47.35	58.03	3	47.19	60.74	5	51.66	51.12
		(5.06)	(2.44)		(6.08)	(3.99)		(2.85)	(6.86)
		Gain = 10.68			Gain = 13.55			Gain = -.54	
Districts 2-3 Grade 5	5	51.67	56.21	4	52.92	52.29	3	44.78	51.06
		(12.30)	(13.00)		(4.16)	(5.13)		(8.99)	(2.49)
		Gain = 4.54			Gain = -.63			Gain = 6.28	

TABLE A-8
Descriptive Statistics for Student Off-Task Behavior

		District 1			Districts 2-3		
		Principal Involvement (N=9)	Regular Inservice (N=10)	Control (N=10)	Principal Involvement (N=9)	Regular Inservice (N=7)	Control (N=8)
		AVERAGE NUMBER OF OFF-TASK STUDENTS					
Pre	M	5.44	4.84	3.73	2.51	3.35	4.59
	(SD)	(2.77)	(2.68)	(2.18)	(2.79)	(2.12)	(3.05)
Post	M	3.51	4.15	3.31	1.83	2.44	4.32
	(SD)	(2.57)	(3.20)	(1.76)	(1.83)	(1.90)	(1.78)
Delayed	M	3.62	3.50	6.58	3.71	3.23	3.08
	(SD)	(2.82)	(1.71)	(4.50)	(2.28)	(1.55)	(.68)
		NUMBER OF STUDENTS IN CLASS					
Pre	M	22.06	22.60	20.50	18.56	21.29	21.50
	(SD)	(2.30)	(3.17)	(2.22)	(8.62)	(3.30)	(6.72)
Post	M	21.72	22.40	19.50	18.17	19.71	20.94
	(SD)	(4.54)	(3.27)	(2.46)	(7.75)	(4.19)	(5.97)
Delayed	M	20.44	23.00	19.00	18.67	19.29	21.63
	(SD)	(3.54)	(3.30)	(2.92)	(7.28)	(3.64)	(7.07)
		AVERAGE PERCENTAGE OF OFF-TASK STUDENTS					
Pre	M	24.89	21.20	18.40	12.00	15.29	20.63
	(SD)	(13.20)	(9.96)	(10.88)	(10.67)	(8.94)	(12.76)
Post	M	17.22	18.70	17.50	10.22	12.86	22.00
	(SD)	(13.79)	(15.30)	(9.98)	(8.93)	(10.53)	(9.80)
Delayed	M	18.00	15.00	33.50	19.78	17.57	15.63
	(SD)	(13.74)	(6.70)	(21.28)	(10.57)	(9.68)	(5.55)