Research Monograph No. 13

Form and Substance in Inservice Teacher Education

Mary Kennedy

National Institute for Science Education
University of Wisconsin-Madison

December 1998

Abstract

Much of the reform rhetoric about professional development is geared toward the form that such development should take. This literature advocates collaboration among teachers, schoolwide participation in professional development, programs that extend over time and are interspersed with classroom practice, programs that include classroom visitations, and so forth. Much less has been said about what the content of such programs should be. This paper reviews studies of in-service programs that aim to enhance mathematics and science teaching. It focuses exclusively on studies that examine effects of programs on student learning. The review suggests that the differences among programs that mattered most were differences in the content that was actually provided to teachers, not difference in program forms or structures.

Introduction

The one-shot workshop is a much maligned event in education. This event has been criticized by virtually every teacher who has ever participated in it and by virtually everyone else even vaguely interested in improving teaching. In a survey of teachers' ratings of different sources of learning, Smylie (1989) found that district-sponsored inservice programs ranked dead last among 14 possible sources of learning. The top-rated sources of learning were teachers' own classroom experiences, consultation with other teachers, independent study, and observations of other teachers. Researchers and policy analysts, also critical of the one-shot workshop, have generated a number of proposals for how inservice education programs ***should*** be organized. Frequently recommended features of "good" inservice programs include that they be lengthy rather than brief, that teachers have a role in defining the content rather than having the topics imposed on them, that the scheduled meetings be interspersed with classroom practice rather than concentrated, and that they allow teachers to work in groups, rather than in isolation.

If it is the case that many of the recommended features of inservice programs have been proposed as correctives to the one-shot workshop, it is also possible that these proposals are correcting for the wrong flaw. One-shot workshops may be guilty of being overly brief, but they may also be guilty of being irrelevant. They may be guilty of treating teachers as passive receptacles, but they may also be guilty of addressing the wrong topics. Which flaws need to be corrected? Surprisingly, the reform proposals rarely mention the ***content*** that inservice teacher education programs provide to teachers. When content is mentioned, it is mentioned as something that should be coordinated over time rather than randomly changing from one event to the next. But what the content should be-teaching techniques versus research findings on how students learn, for instance-is rarely discussed.

When I say content, I do not necessarily mean the school subject-matter content. I mean the topics are that are dealt with in a program. Inservice teacher education content might include, for instance, classroom management and discipline techniques, techniques for working with parents, legal definitions of sexual harassment, knowledge about specific school subject matter, knowledge about how students learn specific school subject matter, knowledge of how to teach specific school subject matter, or other issues. When I say that reformers fail to discuss the content of inservice teacher education, I mean that they do not discuss which of these topics are most important for teachers. Instead, they discuss the length of time that should be devoted to inservice programs, the schedule of the programs, the way in which teachers are engaged in the programs, or other features that are unrelated to the content actually taught to teachers.

My goal in this paper is to examine the relevance of the content of inservice teacher education. I address this question by reviewing literature on the effects of various approaches to inservice teacher education. My review differs from others in two ways. First, I am more interested in the content of the programs being studied than I am in their structures, formats, or schedules. Second, I am more interested in whether these programs have any eventual influence on student learning than I am in what teachers think about the programs. I focus on studies that examine the relationship between inservice programs and eventual improvements in student learning.

I am particularly interested in programs that aim to improve student learning in either mathematics or science, two subjects that have received a great deal of attention by would-be reformers. These two subjects are frequently mentioned in a single breath, as if they were siblings. But as ***school*** subjects, especially in elementary schools, these subjects are remarkably different, largely because mathematics is considered a “basic skill”–one of the 3 Rs–whereas science is not. Many other important differences follow from this single distinction. For instance, science content is not routinely included on standardized achievement tests, whereas mathematics content is. Therefore researchers interested in studying student learning in science must devise their own outcome measures, whereas researchers interested in mathematics have many standardized instruments available to them.

Similarly, schools rarely purchase science textbook series that are integrated across the entire elementary grade span, even though virtually every elementary school in the country owns such a textbook series in mathematics. And even though teachers exercise a great deal of discretion in their teaching of mathematics, skipping portions of the text occasionally and adding supplemental material of their own here and there (Schwille et al., 1983; Porter, 1989), teachers exercise even greater discretion in their teaching of science. Teachers who are not interested in science may not teach much of it, or even any of it at all. When teachers do choose to teach science, they may use any number of ancillary materials or may devise special units on particular topics. There is no coordination of these efforts across grade levels, as there is in mathematics.

This status difference between mathematics and science also has implications for research on subject-specific inservice teacher education. One implication, already mentioned above, is that researchers working in mathematics have readily available standardized tests to measure student outcomes, while those working in science do not. Another implication is that those working in mathematics must acknowledge public interest in basic skills whereas science researchers are not so constrained. Most researchers engaging in inservice programs are interested in moving teaching practice away from basic skills and toward content that requires more analytic reasoning and problem solving. Because mathematics is considered a basic skill, however, researchers in mathematics must attend to the basic-skills aspect of their subject. Science, on the other hand, is not considered a basic skill, and there are no generally recognized basic facts or skills within science that we expect all elementary students to master. So whereas researchers in mathematics need to show that students have learned certain basic things, such as number facts or the multiplication tables, those working in science are not under such strong public oversight.

Yet another implication for researchers in these two areas is that, because elementary curricula in the sciences are more discretionary and more variable than those in mathematics, inservice programs in science are far more likely than those in mathematics to provide teachers not only with a set of teaching behaviors or teacher knowledge, but also with curriculum materials and teachers’ guides to accompany them.

The Literature

Although the literature on inservice programs is voluminous, that volume subsides quickly when you limit yourself, as I did, to studies that include evidence of student learning. The studies I found, and that I will discuss in this paper, are shown in Table 1. I have organized them into four groups according to the content they provide teachers:

- Those that prescribe a set of teaching behaviors that are expected to apply generically to all school subjects: These behaviors might result from process-product research or might include things like cooperative grouping. In either case, the methods are expected to be equally effective across school subjects.
- Those that prescribe a set of teaching behaviors that seem generic, but are proffered as applying to one particular school subject, such as mathematics or science: Though presented in the context of a particular subject, the behaviors themselves have a generic quality to them, in that they are expected to be generally applicable in that subject.
- Those that provide a general guidance on both curriculum and pedagogy for teaching a particular subject and that justify their recommended practices with references to knowledge about how students learn this subject.
- Those that provide knowledge about how students learn particular subject matter but do not provide specific guidance on the practices that should be used to teach that subject.

Within these groupings, I also distinguish the school subject matter on which they concentrate: mathematics or science.

The first point to notice in Table 1 is the distribution of studies in mathematics versus science. I found only four studies of science inservice programs that provided evidence of student learning, and all four of these studies are located in Group 2. This situation is quite different from that in mathematics, which has at least one study in each of the four groups.

Why does this strong difference in study characteristics exist? Part of the reason lies in the differences I outlined above between the status of mathematics and sciences in the elementary school curriculum. For instance, group 1 consists of studies in which the researchers claim no particular interest in either mathematics or science. However, because mathematics is included in all standardized achievement tests, these researchers frequently provide evidence of student learning in mathematics along with their evidence of student learning in reading. No such evidence is available for science, hence we have no examples of group 1 studies in science.

Second, the lack of uniformity in elementary science classrooms may motivate science researchers to be more prescriptive as well, which could account for their tendency to fall into group 2 rather than in groups 3 or 4. But another important reason is that the mathematics education community has a rather strong body of research on how children learn early arithmetic. The approaches to teaching science that appear in the literature are not based on evidence of how children learn, but instead are based on idealized models of how scientists themselves learn. These models of scientific reasoning define a set of routines that are presumed to apply to virtually all science content.

Table 1
Studies Included in This Review, by Content Focus

Citation	*Subject matter context*	*Grade span of participating students*	*Source of participants*	*Form and distribution of inservice time*	*Total inservice contact hours*[a]	*Study duration in months*[a]
Group 1: Focus on Teaching Behaviors Applying Generically to All School Subjects						
Stallings & Krasavage (1986)	Math	2-4	Schoolwide projects	Distributed workshops		16
Stevens & Slavin (1995)	Math	K-6	Schoolwide projects	Distributed workshops		8
Group 2: Focus on Teaching Behaviors Applying to a Particular Subject						
Good, Grouws, & Ebmeier (1983)	Math	4-12	Individual volunteers	2 @ 1.5	3	4
Good & Grouws (1979)	Math	4	Individual volunteers	2 @ 1.5	3	4
Mason & Good (1993)	Math	4-6	Individual volunteers	3 @ 1.5	4.5	5
Otto & Schuck (1983)	Science	8	Individual volunteers	5 @ variable	16	2.5
Rubin & Norman (1992)	Science	6-9	Individual volunteers	University course (10 @ 3)	30	3
Lawrenz & McCreath (1988)	Science	1-8	Individual volunteers	University course (15 @ 3)	45	8
Marek & Methven (1991)	Science	1-5	Individual volunteers	4-week Summer Institute	100	8
Group 3: Focus on Curriculum or Pedagogy Justified by how Students Learn						
Cobb et al. (1991)	Math	2	Individual volunteers	1-week Summer Institute + Distributed	150	8
Wood & Sellers (1996)	Math	2-3	Individual volunteers	1-week Summer Institute + Distributed	150	16
Group 4: Focus on how Students Learn and how to Assess Student Learning						
Carpenter et al. (1989)	Math	1	Individual volunteers	4-week Summer Institute	80	

[a]Some of these estimates of contact time had to be estimated from general descriptions of programs. For estimates of program durations, I assumed a "school year" was roughly 8 months, a semester was 4 ½ months, and two school years was 16 months.

The inservice programs being examined in these studies differ in more ways than just the content they provide teachers, however. They also differ, for instance, in the researchers' apparent optimism; some of these researchers expected to influence teaching practice by spending only several hours with teachers, while others devoted dozens of hours to their programs. And the studies differ in the researchers' apparent confidence in the power of their programs; the last column of Table 1 shows us the length of time researchers allowed between pre- and posttests of student achievement. I view these time intervals as indications of the confidence these researchers had in the strength of their programs, for the longer the interval, the more likely that some other event would counteract the treatment influence. Factors that could intervene and mitigate the program's influence include teacher illness; change in classroom composition; a new principal, superintendent, or board member with different values; a new activist parent in the community; a fire, flood, or other local natural disaster.

Longer time periods also ***decrease the*** likelihood that the program will retain its fidelity, for teachers may become bored with the new program or may simply drift in their practices over time. Researchers who tested their programs over whole-school-year intervals (indicated in Table 1 as month intervals) apparently had enough confidence in their programs to expect their programs to withstand these myriad other influences.

One other difference between these groups that does not show in Table 1, but is nonetheless important, is their tacit model for how they expect their inservice programs to eventually influence student achievement. Underlying these different approaches to inservice teacher education are different assumptions about the intervening variables that lie in the path between the program and eventual improvements in student learning. I summarize these differing sets of assumptions in Figure 1. Researchers in groups 1 and 2 expect their inservice programs to change teacher behaviors and expect that these behavioral changes will, in turn, lead to student learning. With this idea in mind, these programs focus their inservice on the specific teaching behaviors that they believe will make a difference. Researchers in groups 3 and 4, on the other hand, expect their programs to change teacher knowledge, and they tend to be relatively less prescriptive about teaching practices. The group 3 programs provide teachers with knowledge about how students learn mathematics, with some curriculum materials, and with some ideas about new practices that will better promote student learning. The program in group 4, focuses more on teacher knowledge and less on teaching practice. Researchers in group 4 do, of course, expect teaching practice to change, and they have ideas about the kinds of changes they want to see. However, instead of prescribing all the details of the new practice, they are more inclined to assume that changes in teacher knowledge or beliefs, coupled with examples of practice, will stimulate teachers to devise their own new teaching practices that will, in turn, lead to student learning. Under this model, then, the teaching behaviors that eventually emerge are more discretionary than are those expected by researchers in groups 1 and 2. These four groups, then, seem to move along a continuum from more prescriptive to less, from more focused on behavior to more focused on ideas. The studies in group 3 represent a balance between the two ends of these continua, giving teachers both some practices and some knowledge that justifies those practices.

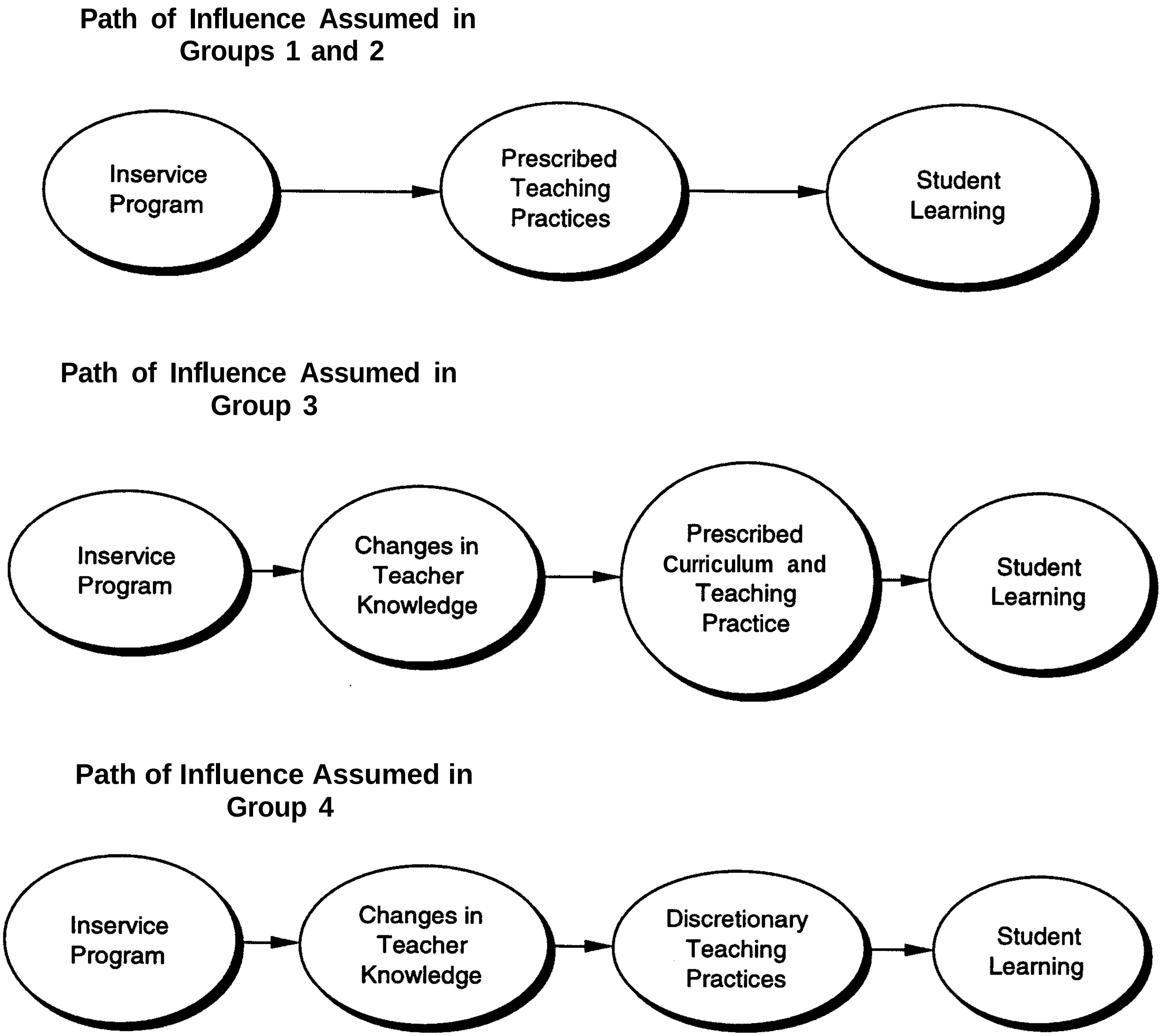

Figure 1: Three Paths to Student Learning

This difference in presumed paths of influence leads researchers to gather evidence about different types of intermediate events. For instance, if researchers in the first two groups include outcome measures other than student achievement, these outcomes measures tend to document the degree to which teachers implemented the prescribed practices. Behavioral changes among teachers provide important evidence of the intermediate impact of the program. Conversely, researchers in groups 3 and 4 tend to be more interested in measuring teachers' knowledge, attitudes, or beliefs, for these changes constitute the intermediate impact of these programs. Evidence about these differing intermediate events is important in interpreting patterns of outcomes in student achievement later on.

Having laid out the central variations among these programs, I now turn to the available evidence of effectiveness of these different approaches to inservice teacher education. The remainder of this paper is divided into three main sections. The first reviews programs designed to improve student learning in mathematics and the second reviews programs designed to improve student learning in the sciences. In each of these two sections I compare program effectiveness across the groupings indicated in Table 1. Finally, in the third section I examine a variety of other popular hypotheses regarding features of inservice teacher education that might make a difference to student learning.

Programs Aimed at Improving Student Learning in Mathematics

Because mathematics is one of the 3 Rs, we are able to examine some studies that focus on generic teaching skills and include standardized mathematics achievement scores in the set of outcomes. I have not done a thorough search for studies in group 1, but am including here two such studies as illustrative of this line of work. These two studies were both done by prominent and highly respected researchers, and both examined generic approaches to teaching that are also highly regarded and widely advocated. Stallings and Krasavage (1986) examined a Madeline Hunter program while Stevens and Slavin (1995) examined a Cooperative Learning program. In both cases, inservice was extensive and distributed throughout the school year. And in both cases, the study duration spanned at least one full school year.

The group 2 studies focusing on mathematics consist entirely of programs sponsored by Tom Good and his colleagues (Good & Grouws, 1979; Good, Grouws, & Ebmeier, 1983; Mason & Good, 1993) and all are variations of the Missouri Mathematics Model. The Missouri Mathematics Model is summarized by the set of admonitions listed in Table 2. These inservice programs typically consist of just two 1 1/2 hour sessions during which the specific behaviors and their rationales are explained. Teachers also receive a manual with more detailed discussion of the model. The program provides a way of organizing both time and students during mathematics lessons, but offers little guidance on the mathematical content itself, on which mathematical ideas might be especially difficult for students to understand, or on how to help students understand any particular mathematical idea.

The studies in groups 3 and 4 are similar in their theoretical orientations. Both are interested in student cognition, both assume some form of constructivist theory of learning, and both are interested in increasing teachers' attention toward mathematical problem solving and reasoning

Table 2

Summary of Key Instructional Behaviors

Daily Review **(first 8 minutes except Mondays)**
1. Review the concepts and skills associated with the homework
2. Collect and deal with homework assignments
3. Ask several mental computation exercises

Development **(about 20 minutes)**
1. Briefly focus on prerequisite skills and concepts
2. Focus on meaning and promoting student understanding by using lively explanations, demonstrations, process explanations, illustrations, etc.
3. Assess student comprehension
   a. Using process/product questions (active interaction)
   b. Using controlled practice
4. Repeat and elaborate on the meaning portion as necessary

Seatwork **(about 15 minutes)**
1. Provide uninterrupted successful practice
2. Momentum-keep the ball rolling-get everyone involved, then sustain involvement
3. Alerting-let students know their work will be checked at end of period
4. Accountability-check the students' work

Homework Assignment
1. Assigning on a regular basis at the end of each math class except Fridays
2. Should involve about 15 minutes of work to be done at home
3. Should include one or two review problems

Special Reviews
1. Weekly review/maintenance
   a. Conduct during the first 20 minutes each Monday
   b. Focus on skills and concepts covered during the previous week
2. Monthly review/maintenance
   a. Conduct every fourth Monday
   b. Focus on skills and concepts covered since the last monthly review

From T. Good, D. Grouws, and H. Ebmeier, ***Active Mathematics Teaching*** (New York: Longman, 1983), p.32.

in place of their memory for computational procedures. They differ, though, in their presumed path of influence. The two studies in group 3 provide teachers with some theory about student learning and then move to a recommended set of teaching strategies and a recommended curriculum that is justified by that knowledge of student learning. The one study in group 4, on the other hand (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989), focuses on the particular mathematical content that students will learn and on the particular kinds of difficulties they are likely to have in learning this content. Carpenter and Fennema and their colleagues first examined the research literature on children's learning of early mathematics and then used these findings to create a taxonomy of types of arithmetic problems and types of student learning difficulties associated with each. This analysis of the curriculum and of students' responses to it constituted the basis for the inservice program. Teachers were not provided with a set of invariant teaching strategies, but the researchers encouraged teachers to think about instructional implications of these findings, and they engaged teachers in discussions about different ways of teaching different types of problems to children.

Table 3 shows the size of program effects on student achievement outcomes in each of several mathematics outcomes in each study. Each number indicates the size of treatment effect in standardized units relative to a comparison group. With the exception of group 1 programs, all studies involved teachers who volunteered to participate, but who were randomly assigned to experimental and comparison groups. Table 3 indicates two important findings.

First, studies in group 1 tend to examine outcomes mainly in the basic-skill side of mathematics, whereas those in groups 3 and 4 tend to examine outcomes in both basic skills and in more advanced reasoning and problem solving areas. These differences reflect, in part, the substantive differences I mentioned earlier. Researchers in group 1 do not recognize mathematics as a subject with any unique teaching requirements, but they measure mathematics outcomes because the content is already present on standardized achievement tests. In contrast, the researchers in groups 3 and 4 are more cognitively oriented, interested in how students come to understand mathematical ideas, and interested in student reasoning, analysis, and problem solving in mathematics. These researchers tend to include both traditional achievement test scores as well as evidence of students' mathematical reasoning and problem solving among their outcomes.

Second, Table 3 indicates noticeable differences in effect sizes across these different program groupings. The smallest program effects on student learning appear in group 1 and the largest appear in group 4. The group 4 study is especially remarkable in that it demonstrates strong effects across all types of outcomes–both basic skills and more advanced reasoning and problem solving.

This pattern of outcomes suggests that the content of inservice programs does indeed make a difference, and that programs that focus on subject matter knowledge and on student learning of particular subject matter are likely to have larger positive effects on student learning than are programs that focus mainly on teaching behaviors.

Why is this the case? In a review of Good et al's group 2 work, Lampert (1988) stressed the significance of one particular finding: the activities that were most difficult for teachers to

Table 3

Standardized Effect Sizes Attained in Each Mathematics Study

Study	Basic Skills		Reasoning, Problem Solving		Attitudes toward Mathematics
	Test	**Effect**	**Test**	**Effect**	
Group 1—Focus on Teaching Behaviors Applying Generically to All School Subjects					
Stallings & Krasavage (1986)					
2nd grade	not specified	-.41			
3rd grade	not specified	-.08			
4th grade	not specified	-.34			
Stevens & Slavin (1995)	CAT Computations	.29	CAT Applications	.10	
GROUP 1 AVERAGE		**-.14**		**.10**	
Group 2—Focus on Teaching Behaviors Applying to a Particular Subject					
Good et al. (1983)					
4th grade	SRA	.15			
6th grade	SRA	.16	Problem solving	-.19	
Mason & Good (1993)					
Two-group format	Computation	.03	Concepts	-.07	
			Problem Solving	-.05	
			Estimation	.09	
Whole-group Format	Computation	.32	Concepts	.16	
			Problem Solving	.13	
			Estimation	.28	
GROUP 2 AVERAGE		**.17**		**.05**	
Group 3—Focus on Curriculum or Pedagogy Justified by how Students Learn					
Cobb et al. (1991)	ISTEP	-.01	ISTEP	.30	.43
	Instrumental Arithmetic	-.06	Relational Arithmetic	1.06	
Wood & Sellers (1996)					
2 yr. vs. 1	ISTEP	.45	ISTEP	.31	.01, -.006
	Instrumental Arithmetic	.16	Relational Arithmetic	1.14	
2 yr. vs. 0	ISTEP	.35	ISTEP	.25	
1 yr. vs. 0	ISTEP	-.09	ISTEP	-.05	
GROUP 3 AVERAGE		**.13**		**.50**	**.13**
Group 4—Focus on how Students Learn and how to Assess Student Learning					
Carpenter et al. (1989)	Interview	.70	Interview	.70	
	ITBS Number fact	.42	ITBS Complex	.43	
	ITBS Simple Computations	.44	ITBS Advanced	.10	
GROUP 4 AVERAGE		**.52**		**.40**	

Note: All effect sizes were derived from comparisons between treatment and control group. All use pooled within-group standard deviations to standardize the metric. Because Wood and Sellers used classes as their unit of analysis, rather than individual students, I multiplied their between-classes standard deviation by 19, the average class size, to make their metric comparable to the others.

implement were those in the “development” category. Good et al. also noticed that development was the most difficult part of their instructional model and offered several possible hypotheses to account for it: Teachers had too many new things to learn; teachers were not motivated; or the researchers had difficulty in defining the development portion of the lesson. Lampert, though, suggests that the development portion of Good et al's lesson format requires clear and compelling explanations, and there may not be an adequate behavioral definition of a clear and compelling explanation. When Good and others describe the development portion of their ideal lesson, they tell teachers to “focus on meaning” and to use “lively explanations, demonstrations,” etc., without offering any guidance on what the meaning is or what a good explanation or demonstration might be. Lampert suggests that the reason this portion of the Missouri Mathematics Model was difficult for teachers was that they did not have adequate subject matter knowledge or knowledge of how students learn subject matter. If Lampert is right, then teachers’ need for specific subject matter knowledge and for knowledge of how children learn subject matter, could account for the greater impact of studies in groups 3 and 4.

Another possible explanation for differences in impact across these program groups is that instructional models advocated by researchers in groups 1 and 2 are simply more boring to implement, for they prescribe an almost invariant daily routine. Stallings and Krasavage (1986) provide some provocative data to this effect. In their study, teachers implemented the instructional model quite well for the first two years, but implementation fell off dramatically in the third year, when the press for compliance was lifted. Models such as the Hunter model being tested by Stallings and the Missouri Mathematics Model being tested by Good et al. do not give teachers much leeway in how they manage their classrooms from day to day, and it may be that both teachers and students need more variety than these models allow. If this is the case, perhaps the greater effectiveness of programs in groups 3 and 4 results from the discretion they permit teachers. The Carpenter et al. study, which demonstrates greater across-the-board effects on student outcomes than any other study, is also the least prescriptive in its approach to inservice teacher education. In fact, that inservice program provided teachers with the ***least*** amount of specific information about what they should do in their classrooms and with the ***most*** specific information about the mathematics content to be taught and on how students learn that content.

Programs Aimed at Improving Student Learning in Science

For reasons outlined above, science studies are not as various as mathematics studies. There were no examples of group 1 studies-that is, studies focusing on generic teaching behaviors that include evidence of student learning in science in their portfolios of outcomes. And there were no examples of groups 3 or 4 studies relying on evidence of how students learn particular science content. Such studies exist, but they tend to examine intermediate outcomes such as teacher learning or teaching practices rather than examining student learning. The science studies that met my criteria fell entirely into group 2. Like mathematics studies in group 2, these science studies all claim to offer teaching techniques that are uniquely suited to the subject matter, but the techniques themselves are still generic within that subject. Consider, for instance, the behaviors outlined by Rubin and Norman (1992). These science researchers wanted teachers to model discrete science processes such as generating hypotheses, identifying and controlling variables, defining things operationally, and so forth. During their inservice program, the researchers used generic lesson formats to train teachers in how to model each of these skills. For instance, modeling the skill of “identifying and controlling variables” consists of asking aloud such questions as, “What is the manipulated variable in this experimental situation?” Teachers were taught to model five specific science process skills.

Unlike the mathematics studies in group 2, though, the science studies in group 2 are various. Whereas the group 2 studies in mathematics all derived from one particular model of teaching, the group 2 studies in science reflect two different models of teaching. I therefore sorted these group 2 studies according to the particular model of teaching that these researchers examined.

Table 4 shows the student learning outcomes that were obtained from these science studies. It shows us three important points. First, no science study included an outcome that might conceivably be considered a “basic skills” outcome. The science studies all focused exclusively on scientific reasoning and problem solving. I suspect that this difference reflects the non-3 R status of science, relative to mathematics, in the elementary school curriculum.

Second, almost all of the effects shown in Table 4 are larger than their counterparts in Table 3. It might be tempting to speculate that researchers in science have developed better content for their inservice programs, but I don’t think that is the case. Instead, I suspect that this difference reflects the status differences I mentioned earlier between mathematics and science. That is, science researchers are more likely to devise their curriculum materials, and they are also more likely to devise their own outcome measures. Consequently, there is likely to be a much greater articulation between the content taught in participating “treatment” classrooms and the content assessed by the science researchers than is the case in mathematics programs. Moreover, there is likely to be almost no articulation between the content taught and content tested in the comparison classrooms used in science studies.

The third important finding shown in Table 4 is that programs that taught teachers to model scientific reasoning seem to have had a greater influence on student achievement than did programs that taught teachers to use the learning cycle. That such a difference is visible again suggests that the content of the inservice program makes a difference to later student achievement. As an aside, notice that the modeling techniques presented in these programs are similar to those that have been recently dubbed as a “cognitive apprenticeship” approach to teaching, and the evidence shown here may lend further support for that idea.

In both mathematics and science teacher education, then, the content of the program makes a difference. Inservice teacher education programs that teach different content also differ in their eventual effect on student learning. Yet the content of inservice programs is rarely mentioned in discussions of how to improve the value of inservice teacher education. Instead, these discussions tend to focus on such issues as the total contact time spent with teachers, whether that time is concentrated or distributed, and so forth. In the next section of this paper, I use the data presented in Tables 1, 3, and 4 to examine some of these other features of inservice programs.

Table 4

Standardized Effect Sizes Attained in Each Science Study
(All from group 2)

Study	**Basic Skills**		**Reasoning, Problem Solving**		**Attitudes**
	Test	**Effect**	**Test**	**Effect**	
(a) Modeling as a Teaching Strategy					
Rubin & Norman (1992)			Integrated Processes	.69	
			Logical Reasoning	.00	
Otto & Schuck (1983)			Circulation Unit	1.08	
			Respiration Unit	1.07	
Part (a) Average				**.71**	
(b) Learning Cycle as a Teaching Strategy					
Marek & Methven (1991)					
Kindergarten			Conservation	-.23	
1st grade			Conservation	.27	
2nd grade			Conservation	.28	
3rd grade			Conservation	.60	
5th grade			Conservation	-.36	
Lawrenz & McCreath (1988)					
4th grade			NAEP	.36	.09
7th grade			NAEP	.87	.20
Rubin & Norman (1992)			Integrated Processes	.34	
			Logical Reasoning	.54	
Part (b) Average				**.43**	**.15**

Note: All effect sizes were derived from comparisons between treatment and control groups. All use pooled within-group standard deviations as denominators. Because Marek and Methven reported percents of students who had achieved criterion, I used a hypothetical standard deviation of .25 to convert their data to standardized effect sizes.

The Relevance of Other Features of Inservice Programs

The two sections above have focused on the content actually provided to teachers during their inservice programs. I was interested in this dimension in part because it seems self-evident that content would make a difference, and in part because content has received so little attention from other researchers and research reviewers interested in teachers' professional development. However, the inservice programs examined in this small body of research differed on several other dimensions as well, and many of these dimensions have been hypothesized to be important to successful inservice. Because the studies included in this review vary on many different dimensions, it is possible to use these studies to examine the merits of several hypotheses about critical features of inservice teacher education. In particular, these studies allow us to examine the following dimensions of program variations:

- Program intensity, as measured by total contact time with teachers
- Whether the time was concentrated or was interspersed with teaching experiences
- Whether the program included classroom visits for consultation or coaching, or was entirely outside the teachers' classrooms
- Whether the program worked with whole schools of teachers, in an effort to create schoolwide reform, or worked with individual teachers who signed up on their own

Total Contact Time

Criticisms of the one-shot workshop often claim that the time spent in these workshops is not nearly enough to promote serious changes in teaching practices. Many inservice programs in the studies I have reviewed involved large numbers of contact hours. But some programs involved only small amounts of time with teachers. The total contact time with teachers in these programs ranged from a minimum of 2.5 hours in the Otto & Schuck (1983) study of modeling to 150 contact hours in several other studies. The differences in total contact hours, however, suggest that this variable by itself is not the most important predictor of effects on student achievement. All of the very brief mathematics inservice programs in group 2, for instance, demonstrated greater influences on student learning than did the very time-intensive program studied by Stallings and Krasavage. Moreover, the Carpenter, Fennema, and others study in group 4 used less contact time than the studies in group 3 did, with no obvious detriment to student learning. In terms of effects on student learning, then, total contact time is not as important a dimension of teacher inservice as is the content that is actually taught.

Concentrated or Distributed Contact Hours

Another popular argument in the literature on inservice teacher education is that programs will have a greater impact on teacher learning, and ultimately on student achievement, if they are distributed over the school year, rather than offered in single blocks of time. The idea here is that teachers will have more opportunities to connect the new ideas to their own classrooms and their own students if they move regularly back and forth between these two environments.

The science studies provide some evidence for an advantage to distributed time. Among these studies, only the Marek and Methvan (1991) study provided a concentrated summer institute, while the other three all provided their program in distributed sessions during the academic year. And the Marek and Methvan study does appear to demonstrate a smaller influence on student achievement than do the other three studies.

The studies in mathematics, on the other hand, do not support this hypothesis, for the mathematics program with the most substantial influences on student learning, the Carpenter et al. study, consisted of a summer institute with no seminars distributed during the next academic year. Conversely, the one program that demonstrated negative effects on student learning, the program studied by Stallings and Krasavage, provided both seminars and in-class visitations throughout the school year.

It is possible that distributed time makes a difference if the content of the program is worthwhile to begin with. That is, perhaps the influence of distributed time was apparent in the science studies because their content was more similar. With less variation in content, the influence of variations in time distribution was more apparent. But in the mathematics studies, distributed time appeared less effective because the variations in content had a greater influence on eventual student learning.

In-class Visitations

The hypothesis that in-class visitations will promote teacher learning, and hence student learning, is based on an argument similar to the argument for distributed time. If the inservice providers actually visit teachers while they are teaching, and provide feedback on their practices or suggestions for change, they enhance the likelihood that the teacher will make connections between the ideas promoted in seminars and the practices they engage in within their own classrooms.

Four of the programs included in this review provided teachers with in-class visitations: The Madeline Hunter model studied by Stallings and Krasavage, the cooperative school model studied by Stevens and Slavin, and the constructivist model studied by Cobb et al. (1991) and by Wood and Sellers (1996). None of these programs produced noticeably greater influences on student learning. In fact, the Hunter model was least successful of all the programs included in this study, and, although the group 3 programs had greater gains than the group 2 programs, they did not do as well as their group 4 counterpart, which provided no in-class assistance.

The effects that these programs demonstrated on student learning, then, do not fall into a pattern that would justify in-class visitations as necessarily a key ingredient in inservice teacher education.

Schoolwide or Individual Programs

Some researchers argue that inservice programs should be targeted toward school buildings rather than toward individual teachers. The reasoning behind this proposal is that schoolwide programs have more likelihood of influencing a critical mass of teachers and that teachers thus influenced might be more likely to encourage one another toward new teaching practices.

The relative merits of working with schoolwide groups versus individual teachers is a difficult issue to sort out, in part because it is confounded with whether or not teachers volunteer to participate in inservice programs. Most programs that serve individual teachers also serve volunteers, while most programs that serve whole schools are likely to have some teachers who are interested in the program and others who are not. Moreover, there is some evidence that people who volunteer for programs are already sympathetic with program goals, even before participating, so that they may be more inclined to adopt the program's ideas than nonvolunteers would be (Kennedy, 1998). The differences inservice arrangements employed by these various programs are shown in Table 1.

The two studies in group 1 provided their programs to whole schools rather than to individual teachers. The remaining studies all involved individual teachers who voluntarily enrolled in the programs. Those programs that worked with volunteers also randomly assigned their volunteers to treatment and comparison conditions, so that the effect sizes we see in Tables 3 and 4 compare volunteer participants with volunteers who were assigned to placebo programs. Studies of whole school programs tend to use nonvolunteers for their comparison groups, a fact that could increase the apparent effectiveness of their programs. However, the fact that the group 1 programs-those working with whole schools-demonstrated the smallest influences on student learning among these studies suggests that providing services to whole schools may not the most important feature of inservice teacher education.

Summary and Conclusion

The widespread distaste for one-shot workshops in education has led to a plethora of proposals for alternative approaches to inservice teacher education. Surprisingly, none of these proposals addresses the content of inservice teacher education. Instead, most focus on such structural or organizational arrangements as the total contact hours, the distribution of contact hours, whether the program includes in-class visits and coaching, and so forth. My aim in this paper has been to examine the importance of program content relative to some of these other variables.

This review is limited in at least two important ways. First, it includes mainly programs that were devised and examined by university professors. None of the programs included in this review were devised or sponsored by national, state, or local bureaucracies, as far as I could tell. Thus, the findings may not apply to programs sponsored by these agencies. The second major limitation is related to the first, and that is that this review is limited to studies that examined the effects of teacher inservice programs on student achievement. It therefore omits scores of studies that examined the effects of teacher inservice programs on teacher knowledge, teacher attitudes, or teacher behaviors. As a result of this constraint, the population of studies included in this review is quite small.

Based on the studies I was able to review, however, it looks as if a strong case can be made for attending more to the content of inset-vice teacher education and for attending less to its structural and organizational features. In the studies reviewed here, programs whose content focused mainly on teachers' behaviors demonstrated smaller influences on student learning than did programs whose content focused on teachers' knowledge of the subject, on the curriculum, or on how students learn the subject. Moreover, the knowledge that these more successful programs provided tended ***not*** to be purely about the subject matter-that is, they were not courses in mathematics-but instead were about ***how students learn*** that subject matter. The programs in groups 3 and 4 were very specific in their focus. They did not address generic learning, but instead addressed the learning of particular mathematical ideas.

I suspect this type of program content benefits teachers in two ways. First, in order to understand how students understand particular content, teachers also have to understand the content itself, so that subject matter understanding is likely to be a by-product of any program that focuses on how students understand subject matter. Second, by focusing on how students learn subject matter, inservice programs help teachers learn both what students should be learning and how to recognize signs of learning and signs of confusion. So teachers leave these programs with very. specific ideas about what the subject matter they will teach consists of, what students should be learning about that subject matter, and how to tell whether students are learning or not. This content makes the greatest difference in student learning.

On the other side, one could argue from a cost-effectiveness point of view that the program sponsored by Tom Good and his colleagues is the most beneficial, precisely because it is so inexpensive to run. Though it yielded smaller average effect sizes, it also cost substantially less to operate. The relative merits of more expensive investments such as those tested in groups 3 and 4 cannot be evaluated without much longer-term studies that can examine the cost and benefits over multiple years.

An equally important finding from this review is the ***lack*** of clear relationship between several other features of these programs and gains in student learning. These programs differed in the total number of contact hours with teachers, in whether or how that time was distributed, in whether that time included in-class visitations, and in whether teachers participated as members of whole schools or as individuals. Arguments have been made in the teacher education inservice literature for all of these dimensions of inservice teacher education. That is, advocates have argued for more contact hours, for more distributed time, for focusing on whole schools rather than on individual teachers, and so forth. Yet the studies reviewed here do not support the merits of any of these dimensions of inservice teacher education, particularly relative to the merits of the content of the program.

While the findings reported here cast serious doubt on much of the professional development reform literature, they cannot be taken as definitive, for there simply have been too few studies of inservice teacher education that randomly assign participants to treatment and comparison groups and that follow the learning of participants' students after the program. If anything, these findings suggest a need to test more carefully the many claims being made about inservice teacher education.

References

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. ***American Educational Research Journal, 26,499-53 1.***

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. ***Journal for Research in Mathematics Education, 22, 13-29.***

Good, T. L., & Grouws, D. A. (1979). The Missouri Mathematics Effectiveness Project: An experimental study in fourth-grade classrooms. ***Journal of Educational Psychology. 71, 355-362.***

Good, T., Grouws, D., & Ebmeier, H. (1983). ***Active mathematics teaching. New*** York: Longman.

Kennedy, M. M. (1998). ***Learning to teach writing: Does teacher education make a difference?*** New York: Teachers College Press.

Lampert, M. (1988). What can research on teacher education tell us about improving the quality of mathematics education? ***Teaching and Teacher Education, 4, 157-170.***

Lawrenz, F., & McCreath, H. (1988). Integrating quantitative and qualitative evaluation methods to compare two inservice training programs. ***Journal of Research in Science Teaching, 25, 397-407.***

Marek, E. A., & Methven, S. B. (1991). Effects of the learning cycle upon student and classroom teacher performance. ***Journal of Research in Science Teaching, 28, 41-53.***

Mason, D. A., & Good, T. L. (1993). Effects of two-group and whole-class teaching on regrouped elementary students' mathematics achievement. ***American Educational Research Journal, 30, 328-360.***

Otto, P. B., & Schuck, R. F. (1983). The effect of a teacher questioning strategy training program on teaching behavior, student achievement, and retention. ***Journal of Research in Science Teaching, 20, 521-528.***

Porter, A. C. (1989). A curriculum out of balance: The case of elementary school mathematics. ***Educational Researcher, 18(5), 9-15.***

Rubin, R. L., & Norman, J. T. (1992). Systematic modeling versus the learning cycle: Comparative effects of integrated science process skill achievement. ***Journal of Research in Science Teaching, 29, 715-27.***

Schwille, J., Porter, A., Belli, G., Floden, R., Freeman, D., Knappen, L., Kuhs, T., & Schmidt, W. (1983). Teachers as policy brokers in the content of elementary school mathematics. In L. S. Shulman & G. Sykes (Eds.), ***Handbook of teaching and policy*** (pp. 370-391). New York: Longman.

Smylie, M. A. (1989). Teachers' views of the effectiveness of sources of learning to teach. The ***Elementary School Journal, 89, 543-558.***

Stallings, J., & Krasavage, E. M. (1986). Program implementation and student achievement in a four-year Madeline Hunter Follow Through project. ***The Elementary School Journal, 87,*** 117-138.

Stevens, R. J., & Slavin, R. E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. ***American Educational Research Journal, 32,*** 321-351.

Wood, T., & Sellers, P. (1996). Assessment of a problem-centered mathematics program: Third grade. ***Journal for Research in Mathematics Education, 27, 337-353.***

Single copy price is $6.00. To order copies contact:

CENTER DOCUMENT SERVICE
Wisconsin Center for Education Research
1025 W. Johnson St., Room 242
Madison, WI 53706-l796
608/265-9698

NO PHONE ORDERS. PREPAYMENT REQUIRED FOR ORDERS UNDER $20.00.

Price is subject to change without notice.