DOCUMENT RESUME

ED 389 735 TM 024 362

AUTHOR	Flexer, Roberta J.; And Others
TITLE	How "Messing About" with Performance Assessment in Mathematics Affects What Happens in Classrooms.
INSTITUTION	National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CSE-TR-396
PUB DATE	Feb 95
CONTRACT	R117G10027
NOTE	54p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Standards; *Educational Assessment; Educational Improvement; Elementary Education; Elementary School Students; *Elementary School Teachers; Grade 3; *Mathematics; *Teaching Methods; *Test Construction
IDENTIFIERS	NCTM Curriculum and Evaluation Standards; *Performance Based Evaluation

ABSTRACT

This paper reviews a year's work with third-grade teachers who introduced performance assessments in the hope of improving both instruction and assessment in mathematics. The 14 participating teachers in 3 schools tried many changes in their educational and assessment practices. Patterns of stability and change that resulted from their efforts were examined, focusing in-depth on six teachers. The main finding was that the teachers indeed adopted many changes with respect to course content and pedagogy and assessment. Changes in assessment and instruction were mutually reinforcing for most of the teachers. By the end of the year, many were using more hands-on and problem-based activities that were closely aligned with the "Standards" of the National Council of Teachers of Mathematics. The introduction of performance assessment raised teachers' expectations of what their students could accomplish. Change resulted, not from what teachers were told to do, but from what they experienced as they attempted to change. An appendix provides examples of math tasks provided by the project. (Contains 2 figures, 1 table, and 21 references.) (SLD)

CRESST

on Evaluation, Standards, and Student Testing

How "Messing About" With Performance Assessment in Mathematics Affects What Happens in Classrooms

CSE Technical Report 396

Roberta J. Flexer, Kate Cumbo, Hilda Borko, Vicky Mayfield, and Scott F. Marion

CRESST/University of Colorado at Boulder

► UCLA Center for the Study of Evaluation

in collaboration with:

► University of Colorado
► NORC, University of Chicago
► LRDC, University of Pittsburgh
► University of California, Santa Barbara
► University of Southern California
► The RAND Corporation

How "Messing About" With Performance Assessment in Mathematics Affects What Happens in Classrooms

CSE Technical Report 396

Roberta J. Flexer, Kate Cumbo, Hilda Borko,
Vicky Mayfield, and Scott F. Marion

CRESST/University of Colorado at Boulder

February 1995

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Center Program, cooperative agreement number R117G10027 and CFDA catalog number 84.117G, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

PREFACE

The current intense interest in alternative forms of assessment is based on a number of assumptions that are as yet untested. In particular, the claim that authentic assessments will improve instruction and student learning is supported only by negative evidence from research on the effects of traditional multiple-choice tests. Because it has been shown that student learning is reduced by teaching to tests of low-level skills, it is theorized that teaching to more curricularly defensible tests will improve student learning (Frederiksen & Collins, 1989; Resnick & Resnick, 1992). In our current research for the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) we are examining the actual effects of introducing new forms of assessment at the classroom level.

Derived from theoretical arguments about the anticipated effects of authentic assessments and from the framework of past empirical studies that examined the effects of standardized tests (Shepard, 1991), our study examines a number of interrelated research questions:

1. What logistical constraints must be respected in developing alternative assessments for classroom purposes? What are the features of assessments that can feasibly be integrated with instruction?

2. What changes occur in teachers' knowledge and beliefs about assessment as a result of the project? What changes occur in classroom assessment practices? Are these changes different in writing, reading, and mathematics, or by type of school?

3. What changes occur in teachers' knowledge and beliefs about instruction as a result of the project? What changes occur in instructional practices? Are these changes different in writing, reading, and mathematics, or by type of school?

4. What is the effect of new assessments on student learning? What picture of student learning is suggested by improvements as measured by the new assessments? Are gains in student achievement corroborated by external measures?

5. What is the impact of new assessments on parents' understandings of the curriculum and their children's progress? Are new forms of assessment credible to parents and other "accountability audiences" such as school boards and accountability committees?

This report is of one of three papers that were presented at the 1994 annual meeting of the American Educational Research Association and summarize current project findings.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27-32.
Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.
Shepard, L. A. (1991). Will national tests improve student learning? *Phi Delta Kappan, 73*, 232-238.

HOW "MESSING ABOUT" WITH PERFORMANCE ASSESSMENT IN MATHEMATICS AFFECTS WHAT HAPPENS IN CLASSROOMS [1,2]

Roberta J. Flexer, Kate Cumbo, Hilda Borko,
Vicky Mayfield, and Scott F. Marion
CRESST/University of Colorado at Boulder

Introduction

This paper reviews a year's work with third-grade teachers who introduced performance assessments in the hope of improving both instruction and assessment in mathematics. Our interest in this effort, and the staff development program we designed, drew upon ideas central to current reform in mathematics education and educational measurement. Participating teachers tried out many changes in their instructional and assessment practices. By year-end, teachers had increased their use of hands-on and problem-based activities, extended the range of mathematical challenges they considered feasible to attempt with third graders, and incorporated performance tasks and observations to replace or supplement computational and chapter tests.

This report also examines teachers' beliefs related to assessment and instruction in mathematics as they experimented with new assessments in their classrooms. More specifically, we examine patterns of stability and change that resulted from teachers' year-long effort to incorporate performance assessments into their instructional programs.

The current reform in mathematics education can be described by three sets of standards produced by the National Council of Teachers of Mathematics (NCTM): *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989), *Professional Standards for Teaching Mathematics* (NCTM, 1991), and

---

[1] Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1994.

[2] We thank Abraham S. Flexer for his support throughout the project and for his editing of this manuscript. We also thank Carribeth Bliem, Kathy Davinroy, and Maurene Flory for their many hours of work on the project, particularly the hours of sitting through meetings with teachers, transcribing tapes, and checking transcripts. We give special thanks also to Pam Geist, a visiting researcher, for her very valuable contributions to the teachers and to the research team.

We are particularly grateful to the teachers who worked so hard for this project and to their district administrators and personnel.

Assessment Standards for School Mathematics—Working Draft (NCTM, 1993). (These sets of standards will be referred to in the rest of this paper as the NCTM *Standards*.) These standards grew out of work done in the late 70s, reported in 1980 in an *Agenda for Action* (NCTM, 1980), that was a reaction to the Back to the Basics Movement of the 70s. The curriculum, assessment, and instruction proposed in these NCTM *Standards* emphasize mathematical thinking, reasoning, problem solving, and communication. Students are expected to understand the mathematics they do and to model and explain their work. The emphasis is no longer on memorization of facts and the mechanical following of procedures. Mathematics is supposed to be relevant and contextualized. The content of the curriculum is supposed to be broader than numeration and computation, and to involve, for example, topics in geometry, probability, and data analysis. Algebraic ideas are to be brought into the elementary schools, giving younger students powerful tools for attacking problems.

Concurrent with this reform in mathematics education, a reform movement is underway in the measurement community. Researchers are investigating the extent to which instruction is influenced by standardized tests (Romberg, Zarinnia, & Williams, 1989; Smith, 1991). The standardized tests, then and now, focus on recall of facts and definitions and demonstration of computational procedures; and many teachers appear to respond by narrowing instruction to what is on the tests and in a format compatible with the tests. Teachers state their sense of responsibility for "preparing" their students for such tests. Their position is often justified by the high stakes some districts place on having their students perform well (Shepard & Cutts-Dougherty, 1991). A prior study by this CRESST-CU research group showed that elementary students in a high-stakes district were able to produce scores on standardized tests that did not hold up when the students were given other tests of the same material (Flexer, 1991; Koretz, Linn, Dunbar, & Shepard, 1991). In addition, the more the format of an alternative task varied from the corresponding standardized-test task, the poorer was students' performance. From these studies it appears that standardized tests in high-stakes contexts are having a deleterious effect on what students are learning in mathematics. The response of many teachers to these tests is to omit or limit instructional time on untested topics and to teach others at the lower levels of thinking that match the tests.

In the late 80s there was a convergence of writings by mathematics educators who encouraged the adoption of the new standards of curriculum, evaluation, and teaching, for example, *Everybody Counts* (Mathematical Sciences Education Board, 1989), on the one hand, and by researchers in the measurement community (e.g., Shepard, 1989; Wiggins, 1989) who argued that standardized tests were having a negative effect on instruction and curriculum and were inadequate for promoting higher order thinking, on the other. Curriculum proposed by the NCTM *Standards* is incompatible with standardized tests, but because standardized tests were in place, they were affecting what and how teachers taught. One approach to bring about the hoped-for changes in curriculum and instruction proposed in the *Standards* was to develop state or national tests that are more compatible with the *Standards*. Several state and one national assessment project took this approach and developed tests that included performance assessment tasks, for instance, Maryland, Kentucky, Massachusetts, Maine, and the New Standards Project. If the new tests require broader thinking, reasoning, and problem solving, then teachers would have to teach in such a way that their students were ready for these kinds of tasks. Here at last was a way to change curriculum and instruction—by adopting an end-of-year test that requires a different kind of performance than the old standardized tests. Support for this "top-down" approach to change comes from Gipps' (1992) report that performance assessment (the UK's Standardized Achievement Tasks, SATs) can have positive effects on instruction. But there are also questions about the effects *any* externally imposed test, even if more authentic, will have on instruction, particularly concerns about narrowing the curriculum (Shepard, 1991).

Another approach to change is a "bottom-up" approach in which teachers are helped to change their assessment program in ways that comply with the *Standards* and are further helped to change their instruction to align it with their assessment, and similarly with the *Standards*. This is the approach taken in the current study, and this paper is a report of the effects of third-grade teachers' work on performance assessment in mathematics on their beliefs and practices about curriculum, instruction, and assessment. It is an account of their struggles and successes during an academic year—and of the ways they changed what they thought was important to teach, how they taught, and how they assessed the performance of children.

In this study we are concerned about the teachers' beliefs and practices with respect to what they value in mathematical performance, what school mathematics should be, how children learn, and how they should teach. Both from our own work with teachers and from that of other researchers (Battista, 1994; Cobb, Wood, Yackel, & McNeal, 1992), it is clear that teachers' beliefs about how children learn mathematics and the nature of school mathematics will very much influence their beliefs and practice about instruction and assessment in mathematics (see Figure 1). We did not intend to confront directly teachers' beliefs but expected beliefs would shift through work on assessment practices and, as it turned out, on instruction practices. We believe that belief and practice can be causally related in both directions, and that it is not only the case that a change in belief causes a change in practice. A shift in practice may lead to a shift in belief which can lead to further shifts in practice (see Figure 2). We know from the literature on teacher change (Borko & Putnam, in press; Nelson, 1993; Richardson, 1990) that making changes in either direction is no easy task.

Research Questions

Because the primary goal of this research project was to help teachers change their assessment practices, the primary set of questions addressed the effect of the staff development intervention on teachers' assessment programs—what did they try; what problems did they encounter; what advantages and disadvantages did they find in performance assessment; and, most importantly, what changes did they make?

Because we see assessment and instruction as inextricably linked, and because we were interested in the effects of changing assessment on instruction, we also examined teachers' beliefs and practice about instruction. A second set of questions asks about these beliefs and practices—what was the effect of the teachers' work on assessment on their instruction; what instructional changes did teachers make; what effect did teachers report the changes had on children's learning; and how did teachers view the new instruction? And the questions that are very much a part of teachers' belief systems ask—what are teachers' beliefs and practice about how children learn; what is important to teach them in mathematics; and were there any changes in these beliefs or practices?

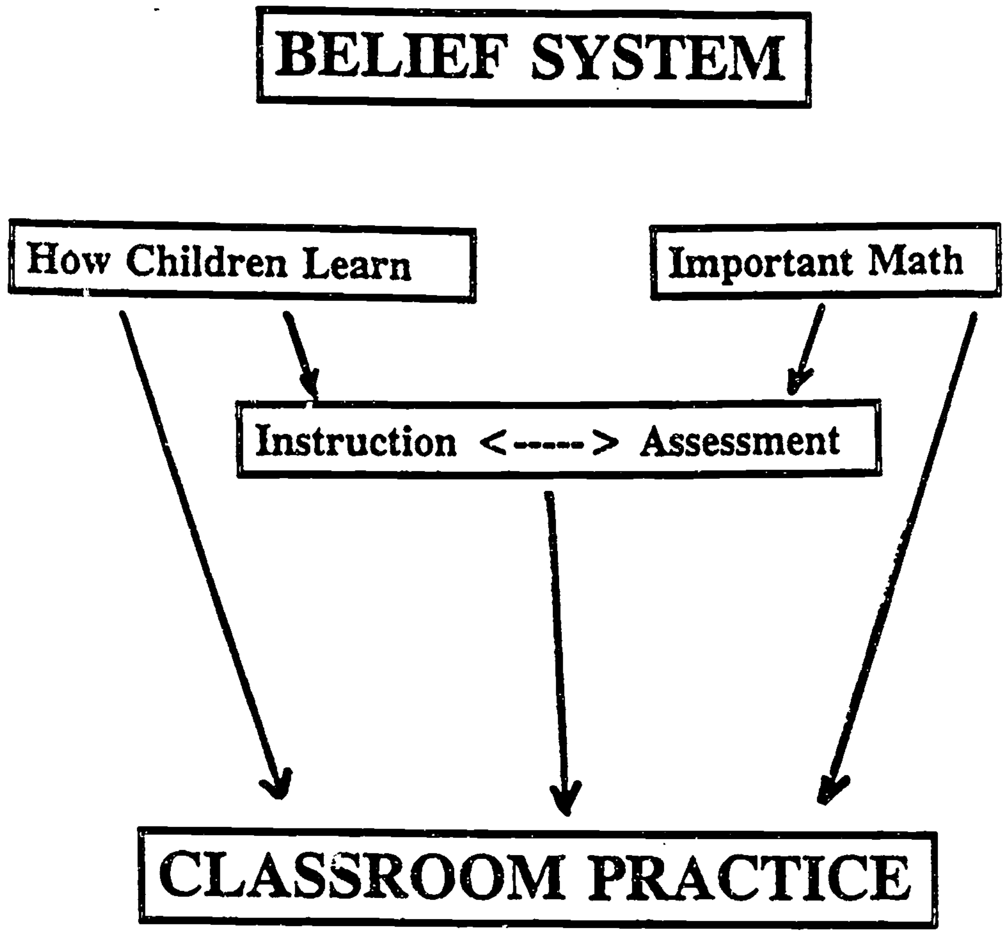


Figure 1. Knowledge and beliefs about how children learn and what mathematics is important to teach affect knowledge and beliefs about instruction and assessment. The three key areas are part of a teacher's belief system and will affect classroom practice.

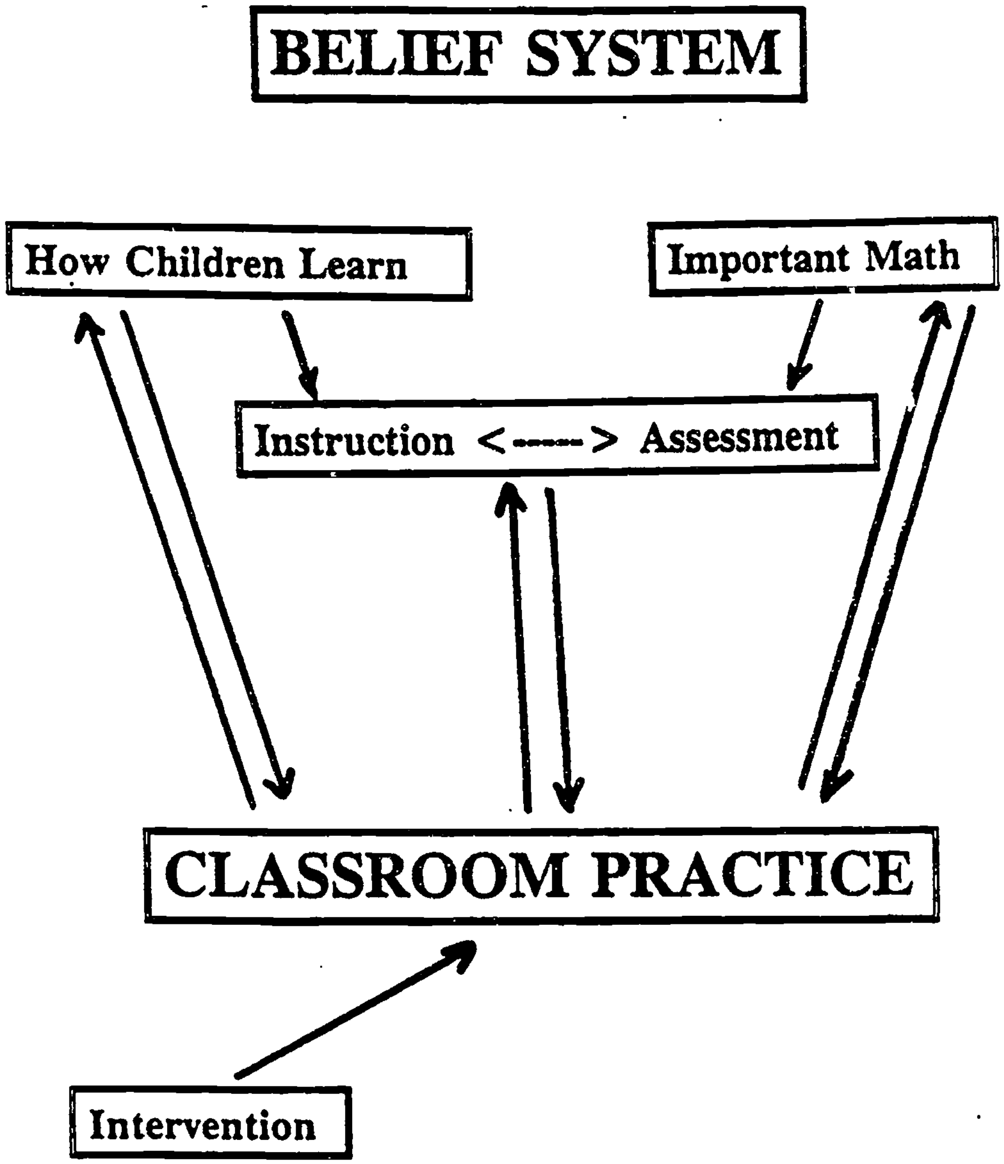


Figure 2. Applying an intervention that changes classroom practice can have an effect on a teacher's belief system.

Method

The Project

This paper is based on data collected during the 1992-93 school year as part of the Alternative Assessments in Reading and Mathematics (AARM) project. The professional development aspect of the project was designed to help third-grade teachers select, develop, and improve classroom-based performance assessments in reading and mathematics that were compatible with their instructional goals. Our overarching research goals were to describe and explain the effects of these professional development activities on the instruction and assessment practices, and knowledge and beliefs of participating teachers, and on student outcomes. This paper describes the effects of staff development efforts in mathematics on several teachers with whom we worked. The team working with the teachers in mathematics throughout the year consisted of a mathematics educator, an expert in assessment, and a specialist in teacher change. The team had the assistance of several doctoral students and a visiting researcher.

Participants and Setting

We sought a school district that had a standardized testing program in place, a large range in student achievement, and considerable ethnic diversity. The district had to be willing to waive standardized tests for two years in the schools in which we worked.

The district selected is on the outskirts of Denver with a population that ranges from lower to middle socioeconomic status. The research team worked with 14 third-grade teachers in three schools (5 in each of two schools and 4 in the third). Each school submitted a letter of application signed by the principal, by the school's parent accountability committee, and by all third-grade teachers in that school.

While all 14 participating teachers were technically volunteers, some were less enthusiastic than others to engage in the project. Some of the original teachers who volunteered changed grade levels or schools and were replaced by other teachers who found themselves involved in a project for which they had *not* volunteered; others may have been "strongly encouraged" to volunteer. Our original assumptions were that all teachers were true volunteers and enthusiastic about the national reforms in reading and mathematics that their district also supported. We later found that these assumptions were incorrect.

Intervention

The intervention was a program of staff development, the primary vehicle for which was a series of weekly workshops between teachers and researchers; reading and mathematics were the focus in alternating weeks. The original intention of the workshops was to help teachers expand their classroom assessment repertoires, for example, by helping them learn to design and select activities, develop scoring rubrics, and make informal assessments "count." A second purpose for the workshops emerged early in the year. Many teachers requested materials for teaching in a way that their district now required and that would match the new assessments, so the scope of the workshops broadened to include more focus on instruction.

It also became clear early in the project that most teachers held fairly traditional views about what mathematics is important to teach, what instruction should look like, and how students should be assessed. Even teachers who were teaching or planning to teach in more activity-oriented, problem-based ways primarily used traditional tests of facts and skills for assessment. Because the instructional and assessment goals of the project matched those of the district (closely aligned with the NCTM *Standards*), we were at odds with the knowledge and belief systems of most of the teachers. Given that we were in the schools to help teachers with assessment, that the teachers had requested help with changing their instruction, and that we had not proposed a project to challenge beliefs, we took the position that teachers, like researchers, would learn from the evidence they accumulated from their classrooms. We worked on assessment (and instruction as teachers requested) in the context of current reforms in measurement and mathematics education, asking teachers to select and use instructional and performance tasks with their students and to bring feedback. We also worked with them on a plan for assessment for the term.

Our discussions in workshops were often about teaching with hands-on, problem-based materials and activities. The project provided tasks (see Appendix A for examples), many of which required problem solving, reasoning, and explaining, that could serve for both instruction and assessment. Because we had agreed to provide tasks that matched teachers' instructional goals and because those goals were primarily computational, most of what we provided the first term focused narrowly on place value, addition, and subtraction. The tasks were also short and structured so that teachers could see the connection between what they were teaching and the assessment task. One might say we were asking them to take small steps. We also selected tasks from sources that are easily available to teachers, so they would be able to make selections independently. We tried to help teachers think about their instructional goals, particularly what they want students to know and why; what it means to know math; how to tell if a student understands mathematics; and how to design and select problem-solving activities to elicit higher order thinking. Dialogue at workshops was about, among other things, selecting, extending, designing, and using activities and materials for instruction *and* assessment; making observations and how to keep track of them; analyzing students' work; and developing rubrics for scoring it. There was major emphasis on helping the teachers see the connection between assessment and instruction, that is, the "embeddedness" of assessment in instruction and curriculum.

The intervention or staff development included several full- or half-day in-service workshops attended by teachers from all three schools, the biweekly workshops within schools, project "assignments" that each teacher did with her class between workshops, demonstration lessons in two of the schools, and consultation on making observations in the third. Three interviews that were part of data collection (see below) are also part of the intervention because they gave teachers a chance to reflect formally on their beliefs and practices.

Sampling

A sample of six teachers, two from each of the three schools, was selected for in-depth study for this paper. The teachers were selected, after an initial analysis of the data, to represent a range of assessment and instructional practices and comfort with mathematics and mathematics teaching and were moderately to strongly engaged in the project. The method of selection, based on the initial analysis frame, ensured that the six cases are representative of 10 of the original 14 teachers. Of the remaining four teachers, one was marginally engaged in the project; the other three had more limited mathematical content knowledge.

Data Sources

The analyses for the present study were based on two sources of data collected from all three schools: semistructured interviews and biweekly workshops. All teachers participated in face-to-face interviews three times during the 1992-93 school year: fall, winter, and spring. The interviews were designed to assess teachers' knowledge, beliefs, and reported practices about mathematics instruction and assessment, as well as the relationship between assessment and instruction. A member of the research team conducted each interview; each interview took place at the participant's school during the day. The interviews were audiotaped and transcribed.

All 15 mathematics workshops from each school were read and coded (see analysis section below for description of the coding scheme). For the second round of analyses we then selected 6 workshops from each school,[3] 2 each from fall, winter, and spring, that addressed our project goals most explicitly and extensively. We decided, based on an initial analysis of the coded transcripts, that this sampling strategy would enable us more easily to search for trends without losing valuable information about patterns in the teachers' knowledge, beliefs, and practices.

Data Analysis

Our analyses began with all five authors reading the same two transcripts (one interview and one workshop) to develop a tentative coding scheme that would take into account issues of learning, instruction, and assessment in mathematics, as well as teachers' background and reactions to the project. This coding scheme went through two more iterations; that is, we coded different workshop and interview transcripts, discussed our codes, and modified the scheme. Our final coding scheme included categories listed in Table 1. Additionally, whenever a teacher talked explicitly about changes, we added a flag for change to the original code (see Appendix B for complete description of the coding scheme). If teachers mentioned change in an interview that did not fall under one of the original codes, for example, if a teacher talked about her growth in confidence, it was given a code for teacher insight or learning (Tlrn).

During the second stage of analysis, we developed "cases" of each of the 6 targeted teachers, that is, summaries of data for each teacher organized according to several key areas. (At this point we focused on the three interviews and the sample of workshops, rather than the entire set.) These key areas were drawn from the original coding scheme by eliminating several less productive codes and expanding key ideas where our data revealed a rich picture about

[3] For one school, 7 workshops were analyzed because each targeted teacher was absent from one or more workshops initially selected for in-depth analyses.

Table 1

Coding Categories for Analysis of Interview and Workshop Transcripts

Background Underlying Instruction and Assessment	
	Beliefs about students' learning
	What it means to know mathematics
Instruction	
	Teachers' goals for mathematics learning and instruction
	Instructional tasks and activities
	Organization and management of instruction
Assessment	
	Roles and purposes of assessment
	Content/substance of assessment tasks
	Scoring of assessment tasks
	How teachers keep track of what students know
	How teachers assign grades in math
	What teachers hoped to learn about assessment through this project
Reactions	
	Dilemmas the teachers faced
	Dilemmas the researchers faced
	Advantages and limitations of performance assessments, including changes in student learning
	Advantages and limitations of the project

changes in beliefs, knowledge, and practices of these teachers. The three key areas were: (a) beliefs and practice about how children learn mathematics; (b) beliefs and practice about what school math is and what is important to learn and assess; and (c) beliefs and practices about instruction and assessment. These areas were augmented by data about variables that we considered important to this study: comfort with mathematics teaching, support for change, and engagement in the project. Because the area of beliefs and practices about instruction and assessment was central to our goals and included extensive data, it was divided into the following four subcategories: general instruction and assessment, problem solving, explanations, and additional assessment. Beliefs and practice varied from a "traditional" conception (e.g., children learn by being told; school math is about facts and computation; instruction is through the text;

assessment is through tests of facts and computation) to a conception aligned with the NCTM *Standards* (1989, 1991, 1993) (e.g., children figure things out themselves; school math is about mathematical thinking, patterns, relationships, and explanations; instruction is through activities that require doing, thinking, reasoning, communicating, and generalizing; assessment is through multiple sources of data that give teachers evidence of student abilities to do, think, reason, communicate, and generalize). The variables of support, comfort with mathematics teaching, and engagement with the project varied along dimensions from limited or low to generous or high. (See Appendix B for more details.)

Our third and final stage of analysis entailed "looking across" these cases for themes that best describe the effect of the intervention on changes in this group of third-grade teachers' beliefs and practices about mathematics instruction and assessment. This final analysis addressed the research questions initially posed for this study.

Results

In this section we present themes that emerged within each of the three key areas from our analysis: beliefs and practice about (a) how children learn mathematics, (b) what school math is, and (c) instruction and assessment in mathematics. Although our primary interest is in the third area, we begin with the first two areas because of their influence on the design of instruction and assessment. We then discuss beliefs and practice about instruction and assessment and how teachers changed in these areas.

To protect their anonymity, teachers' names are not used, and the findings are presented in a way that prevents reconstructing individual cases.

Beliefs and Practice About How Children Learn

We found two major themes in examining teachers' beliefs and practice about how children learn. The first has to do with differences among children and the second with how learning should be structured in mathematics and the importance of children's comfort.

Differences among children. Most teachers believed that some children are more capable of doing mathematics than others. Teachers in this project believed that observed differences among children's mathematical capabilities are the result of either developmental differences at a particular time, or enduring differences in children's native abilities. One teacher compared learning mathematics to the way children learn to speak—at an early stage a child understands more than he or she can say, so the child has *received* concepts and information but is not ready to *transmit* evidence that she or he has them. Some teachers frequently reminded us that their students are only eight years old and may be at too early a developmental level for higher order thinking tasks, or at least that some third-grade students are not ready. Further, at least two teachers in the fall held the position that a few children in each class may *never* reach a developmental level that allows them to understand and should of necessity be taught by rote. For example, early in the year one teacher said:

> . . . a child like that, maybe we're better off just teaching him how to add and subtract on paper the traditional way, because that child may never until he's 30 understand what he's doing. See, I'm not sure that understanding has to come before doing it. I think many times doing it on pencil and paper, later then will help you understand it. See, I'm not sure that understanding has to come first. Because I think some children aren't capable of understanding.

She went on to say that most of the children *will* understand, and that she was talking about only a few. This teacher seemed to soften her position by winter, moving from the view that some children may lack *capacity* to the idea of developmental levels.

> . . . there are children who just developmentally, aren't thinkers yet. And what we feed into them they can spit out, but they're not mature enough to really do a lot of real heavy thinking. . . . I think it can be, you know, developed, but some children are at different developmental stages and some kids just aren't ready for that. I have a couple of them in my classroom that just seem to, you know, if I show them how to do a problem, they can do it. But to really do some thinking about it, it's hard for them.

One teacher thought that some children had more logical ability than others and that would affect their capacity to do mathematics.

> . . . some children think more logically than others when it comes to everything and they are better in math and some children have no logical thinking at all and that is one reason why they just don't do well in math.

Teachers with either of these beliefs would be unlikely to present children with material, either for instruction or for assessment, that required higher order reasoning and problem solving—processes the *Standards* promote for all children. As the year progressed, some teachers were surprised at how much third graders could do and became more willing to increase their expectations. By spring, most had a view of the developmental continuum for third graders that included higher order thinking.

Teaching children in small steps and keeping them comfortable. A second theme involves how teachers believe children learn mathematics and also involves teachers' concerns for the comfort of their students. Most teachers believed that children learn mathematics by having mathematical concepts and procedures explained to them in small steps. Prior to this project, all but one of the six teachers had demonstrated their view of how children learn by telling, explaining, and showing, along with some questioning. They had, prior to this year, depended heavily on their textbooks to guide their instruction, holding the traditional view that children learn by being told and shown and then practicing exercises. Children's comfort was very important to the teachers, and this method of instruction appeared to be the path to comfort. For all but one teacher in the fall this meant presenting material in small bits and modeling carefully what the child was to do. For some this also meant that rote instruction of procedures was appropriate because understanding would follow the doing; that is, children learn "how" before they learn "why."

For several teachers, teaching students to do computations without understanding was also acceptable because doing procedures that others in the room can do would raise the student's self-esteem. Similarly, teachers were reluctant to give children tasks they might find frustrating. Yet, if children were used to being shown how to do everything, then any task requiring them to figure out what to do as well as to do it might cause discomfort. One teacher was ambivalent and was determined to give her students problems to solve and explain (even if, at the beginning of the year, "it made some cry"), *but also* to shape responses to problems to the point of eliminating most of the task's problem-solving character. For example, having selected a task that required students to find two-digit numbers that sum to 25, she gave the students the task with 3 sets of boxes set up as an addition/subtraction exercise.

> Because I really didn't think my kids were going to get two digits. I mean I didn't think they were going to understand the concept of two digits, and so I . . .

All of the teachers believed that experiential learning has some place in instruction, although at the beginning of the year only one teacher's primary mode of instruction was modeled after the position of the NCTM *Standards*. She seemed convinced that children could figure things out for themselves and that part of their work was to solve problems.

> I would see myself as most commonly, or probably the most often as the questioner posing questions, and then letting kids figure out how to work things to get an answer to that question.

Two others expressed a desire early on to move in this direction, although their later frustrations suggest they had not anticipated the full implications of this kind of instruction. Even at the end of the year, two teachers were concerned that children may be confused during hands-on activities and, unless carefully guided, may go through the motions without learning anything. One thought that some children are "dependent" workers and would be unwilling or unable to discover important concepts on their own. Even though she believed children learn from these experiences, she had doubts about using them.

> If they are dependent workers they need somebody to guide them through. They don't learn by the discovery method . . .

The implication for assessment is clear. If students must be told everything in order to learn it, then it is unfair to give them a novel or unfamiliar assessment task. If, however, teachers expect children to use their knowledge to solve unfamiliar problems, then an assessment task can present a problem for which no method of solution was taught. Teachers' reactions to the latter idea coincided with their beliefs about how children learn: from wanting to set problems that are challenging,

> I often look for problems that don't really have a solution. Sometimes I really like problems that have lots of solutions,

to wanting to narrow the tasks until the students knew exactly what they were to do. But even the teacher who wanted to challenge her students used assessment challenges that were within a reasonable expectation of what students could do. For example, when she was shown a missing-digit assessment task that involved regrouping, she modified it to one that did not.

Beliefs and Practice About What Is Important to Teach in School Mathematics

In the fall, we asked teachers what their overall instructional goals for mathematics were for the first quarter of the school year and then, over the year, asked them what they considered important for students to learn specifically about addition and multiplication. We also asked teachers in fall, winter, and spring what they mean when they say a student is "excellent" in math. Two themes emerged from these conversations about goals and questions about what it means to be excellent in math. The first was about computation, the second about problem solving and explanations.

Computation. All teachers talked about the importance of knowing and understanding facts, skills, and computation throughout the year. However, the emphasis was different for different teachers, and the views broadened during the year. In the fall computation was valued predominantly, but several of the teachers also talked about wanting children to be able to see patterns, estimate answers, and think about the reasonableness of answers. For one teacher computation was *not* a final goal, and even in the fall she said:

> . . . the computation that we do is really a means to an end. That [it] is not enough for you to be able to add three three-digit numbers. I mean, we want you to be able to do that, but that's not enough, they need to be able to apply it . . .

Another teacher whose major emphasis was on facts and computation in past years and in the fall was not as concerned about them in the spring. Facts and computation remained a primary focus for the other teachers, although their view of "understanding" a process broadened from expecting students to know that "3 X 4 means three groups of four" to expecting students to be able to explain, to show with models, and to apply the computation.

Problem solving and explanations. The second theme is that, as the year progressed, teachers gave more importance to strategies for problem solving and being able to explain how problems are solved and how procedures are done. Problem solving was mentioned at the beginning of the year as an important instructional goal for most teachers, but given the heavy use of the text, several teachers may have been talking about story problems. Teachers did not mention explanations as a goal in the fall, and one teacher may have expressed the concerns of several colleagues early in the year when she questioned the district's goal of explanation. In winter and spring, teachers talked more about wanting students to be able to solve problems in real contexts. By spring, teachers talked about knowing the difference between "problem solving" and "story problems," and "problem solving" had become an important goal, along with explanations.

Teachers' description of excellence in mathematics mirrored closely their instructional goals: a student who is excellent can do well all of the things a teacher listed as important to learn in mathematics. In the fall that meant he or she knows facts and can do computation accurately and quickly. Teachers also expected excellent students to catch on quickly, to be "good thinkers," and to be enthusiastic about mathematics. Teachers who valued problem solving in the fall included it among descriptors of an excellent student.

One teacher said in winter that there were two different ways a student can be excellent in math—either quick at computation *or* good at thinking and problem solving, but by spring she thought an excellent student would be both. By winter, teachers were also describing excellent students as those who could go beyond what had been taught, who sought challenging problems, and who might even make up their own problems. By winter, teachers also mentioned the evidence they expected to see from such a student—demonstrations of good understanding through explanations, writing, modeling, and problem solving. In the spring, all teachers talked about excellent students being good thinkers and skilled in solving problems and explaining their solutions; several teachers expected them to be able to produce more than one solution to a problem, and at least two teachers talked about students' ability to apply what they know to real world problems. There is evidence from their conversations in workshops that every teacher would have this latter expectation, although she might not have mentioned it specifically in the interview. In other words, just as the teachers' ideas about what is important in mathematics developed over the year, so did their view of what it means to know or be excellent in mathematics. Not only did their comments broaden to include more higher order thinking, problem solving, and explaining, but they showed a keener awareness of the evidence they can collect as proof of these processes.

The implications for assessment and instruction of a teacher's ideas of what is important to include in a school mathematics program and what comprises excellence in mathematics are clear. When the emphasis is on computation (as it was for most of our teachers in the fall), then classroom tasks reflect that. When teachers value mathematical thinking and problem solving (a shift we saw in most teachers to some extent by spring), both instruction and assessment will include activities that require students to think and solve problems.

Instruction

Even though the primary focus of this research project was on assessment, we became interested in instruction for three reasons: (a) We believe instruction and assessment progress in tandem; (b) advocates of performance assessment claim beneficial effects on instruction; and (c) the teachers requested assistance with their instruction.

Teachers were asked specifically about their instruction in interviews in the fall, winter, and spring. They also talked about their instruction frequently in the workshops and shared with the research team classroom activities and methods they were using. Three themes emerged: (a) Teachers changed their instructional practice; (b) teachers perceived that students had learned more; and (c) making instructional changes was difficult.

Shift in instructional practice. There was a shift during the year toward using manipulatives, hands-on small-group activities, problem solving, and explanations; and, for the four teachers who used a text in the fall, a corresponding shift away from it. One of the teachers had been teaching in this way before the project started, so that her shift was not so striking, but by spring she was doing more problem solving and requiring explanations that she had not required before. For the teacher who called the text her "bible" the change was dramatic. The shift away from the text surprised two other teachers who had been convinced that their text was excellent. They initially saw no reason to leave it and supported it vigorously to the research team. But when they compared it to the district's new goals for mathematics, they saw the inadequacies of the book, both in coverage of certain topics, for example, probability, and in the book's approach to teaching. They continued to use the book as a source of exercises but shifted to more activity-based instruction.

> [We] found holes in the text book so we used a variety of resources in order to build a unit around probability and statistics. And we spent a whole, the whole grade level, . . . created centers for probability and statistics, and then we exchanged those and we did it with whole group and the kids were, had a variety of materials, spinners, colored, colored tiles . . . dice and we found that in our book there was only one page on probability and statistics. And that is an important strand.

By spring all teachers reported having students solve more problems, write more explanations, and engage in more hands-on activities and suggested that the set of resources our project had supplied facilitated this change.

An interesting, unplanned curricular development became an influential addition to our intervention. Teachers at all three schools adopted the Marilyn Burns multiplication replacement unit, *Math by All Means: Multiplication, Grade 3* (1991). For one school team the project year was the second year of using the Marilyn Burns unit, but it was a first experience for the other two school teams. In one of those schools, the unit was used by the math specialist at the school; the classroom teachers did some follow-up but only one teacher at the school, one of the two in our sample, was significantly involved. Although all teachers mentioned some use of manipulatives in the fall, for several these were limited or largely nonsubstantive; for example, a child could roll a pair of dice twice to get the two numbers he should add together. The Burns unit gives a teacher complete instructions for a hands-on, manipulatives approach to teaching multiplication that includes solving problems and explaining answers and solutions.

This unit may have had considerable effect on the teachers at the first two schools and the one teacher at the third. Teachers had a model of exemplary nondidactic teaching, and they saw how it engaged students. It showed them a way to use manipulatives that was not routinized, although we had discussions with some of the teachers about whether or not students could go through the activities in a rote and mindless way. This unit used manipulatives as models for computational processes, and some of the models were new to most teachers, for instance, rectangular arrays of tiles to represent the product of two numbers. The multiplication unit seemed to make most of our six teachers more comfortable with substantive, hands-on learning; some, of course, already were.

Beyond the multiplication unit, the areas in which teachers felt most comfortable exchanging the text for hands-on activities seemed to be those that were noncomputational and had not been stressed in their programs in the past. For example, teachers at one school developed their own unit on probability, organized around menus of activities; and all three schools used hands-on activities to teach geometry.

We saw some exciting changes in a teacher who had vigorously resisted many of the project ideas. She talked about changing her instruction because of the assessments, and how using the Marilyn Burns multiplication unit along with the activities provided by the project had made her see

> how you change your instruction so that you're making children think more, more engaged, relating it to their everyday life.

She talked of the project being a "catalyst for change," and said that even though the anxiety it produced was not always comfortable, anxiety is sometimes necessary in order to get change.

A teacher who had taught very traditionally in the fall got lots of positive feedback from seeing how much her students now enjoy math. She said:

> T: I like math better myself.
>
> I: Why do you like it better?
>
> T: I just like the way I'm teaching it. The kids are enthused about it. I make sure I have math everyday. Last year, I can't say that.
>
> . . .
>
> Yeah, last year I'd skip a week or two. But the kids do ask for math; they like math.
>
> . . .
>
> I'm doing a better job this year.

Student learning. Teachers reported that they thought their students were learning more and had better understanding. By the end of the year students could solve problems and give explanations at a level that surprised many of the teachers. Teachers were stressing flexibility in solving problems, and students were responding with multiple approaches to their solutions.

> T1: Well, I just think they understand it more, it is not just rote memorization— that they really know what it means when you say 20 times 80 even if they don't know the answer . . . There is a much deeper understanding.
>
> T2: But I think we have given a lot more challenges this year to our group that we would normally not have given a normal third grader. Don't you think? . . .
>
> I could say that she's been exposed to a lot more problem solving than she would have been in my classroom last year.

T3: Also something I'm really encouraging with my kids is to be flexible, that there isn't one way. Today we solved a problem and we got six different explanations of how you could have possibly solved it. In my mind, math has been, in the past, right or wrong, and I'm really trying to encourage them to think flexibly, to be flexible in their thinking that, well if it didn't work this way I could try this, or if it worked this way could it work another way? Could I look at it from a different avenue?

Difficulties with new instruction. The third and not surprising theme is that some teachers had difficulties with two aspects of this kind of instruction. One aspect involved content. Teachers were concerned, for example, with the Marilyn Burns unit, that students would not come away with knowledge of facts and appropriate skills. While they agreed that students had a better understanding of multiplication and its application, they questioned whether it taught the facts adequately and whether students were learning anything from all the activities.

> . . . how to use—to do menus independently and a lot of them were going through the motions of it but they weren't catching multiplication.
>
> . . .
>
> Yeah, other people liked it. But, I had to make a professional judgment. Now I will do Marilyn Burns again but at the same time I will be working—I will incorporate the multiplication tables at the same time. When we were done with Marilyn Burns I think maybe they did have an understanding of multiplication, what we were looking for . . . [but] they can't do any of their tables, then I had to take four weeks out of my math curriculum to work on the tables.
>
> (Oh, so they didn't know any of their tables?)
>
> They didn't know any tables, but I think they had a basis for—that's why we will go back to it. I do think they had some multiplication understanding of the real world, like they looked at things in multiplication. They looked at egg cartons and they saw that things came in sixes, where before I think I just taught the multiplication tables and they never related it to the real world.

The other aspect involved the organization of instruction alternative to the text. As already discussed, two teachers thought their text excellent and saw no reason to change, particularly when it was all organized; leaving the text requires planning, collecting, and organizing new materials. It is unreasonable to expect teachers to choose to add burdens of curriculum development to those of teaching their classes. Even teachers who had been given materials for hands-on instruction in courses they had taken needed time to organize them.

> I have taken all of the math manipulative courses in the district so I got that [a set of activities] from [a district math specialist]. So I was very familiar with them. But I never—it just takes some time to fit it all in, like when to use it and how much do you run off, and you really need that, and then being able to make a critical viewpoint of how much we need and the variety of levels, being able to read that.

Although most teachers welcomed the resources provided by the project and found them useful, these resources themselves increased the amount of material with which teachers had to cope.

All of the teachers found the additional work in the project burdensome in the fall, and by Thanksgiving, they were feeling overwhelmed. The project director negotiated arrangements to ease the burden, for instance, a half day each month of released time and only one weekly assignment instead of two (one each for math and reading). For many of the teachers these arrangements seemed to remedy the problem. Of course it was also the case that they were becoming more comfortable with the new assessments. A couple of teachers remained frustrated, particularly if they were trying many new practices. For example, one teacher had enthusiastically embraced the kind of instruction and assessment we, her district, and NCTM were advocating and set out to revamp totally her mathematics program. By February, she appeared to be overwhelmed with the magnitude of the changes she expected of herself and was having second thoughts and returning to worksheets.

> I am giving more worksheets at this point in time because I found that I couldn't just do problem solving . . . and there needed to be a point in which I went through the same old steps I had done before.
>
> . . .
>
> I feel that it needs to be a little more structured than I had it in the fall. Because we're doing the new significant learnings I kind of jumped into . . . this manipulative and problem solving and no worksheets. But I find there has to be a balance. You can't throw out all the stuff we used to do. Even for your own sanity you have to have some of those things like that [worksheets] while you're getting used to the new program.

Spring found her proceeding with caution, doing more problem solving, but continuing to present material in small steps for her students.

This teacher was not alone in talking about wanting to keep a balance among facts, computation, and problem solving. The actions of all the teachers and their comments about what they valued in school mathematics suggest this was something they all thought about. The balance was, of course, different for each teacher. The most vocal seemed to be telling us we were trying to pull them toward problem solving to an uncomfortable degree; they were also the teachers whose programs had had the least emphasis on hands-on activities and problem solving.

> I personally, I still feel like I need a balance of both. I don't want to do all problem solving every day, this kind of problem solving. And I don't want them to do all pages out of their books every day. But I do think for them to survive, I think they need a balance, and I want them to be able to do some thinking skills, but I also, if they go to fourth grade next year and the teacher says you need to do page 36, 1 through 25, I don't want them to look at each other and not have a clue on what they would do with something like that . . . not know how to put a heading on their paper or write their numbers so that they can be read by other people. I think they need those things from that kind of practice no matter how well they know their facts from playing cards. I just think there needs to be both. I think they need to be able to write problems on paper and have somebody else be able to read them.

Assessment

A set of themes corresponding to instruction emerged for assessment: (a) By the end of the year, teachers were using more authentic evidence to assess what students know; (b) in spring, teachers reported knowing more about what their students know; and (c) (again, no surprise) teachers encountered many difficulties with performance assessment.

Shift in assessment practice. The first theme is the central goal of this project—to help teachers select and/or design performance assessments that expand the variety and quality of ways in which they assess their students. Because established policy at all three schools required timed tests of facts, all teachers used such tests during the year, but some more frequently than others. One teacher's fall program included daily one-minute tests of facts. All teachers also graded children's work on daily computation during the fall, either from the text or from a set of five problems written on the board. At least one teacher in the fall graded students' daily work for neatness and format as well as for accuracy. The teachers described earlier, who valued their text in the fall, also used its pre and postchapter tests (parallel forms of the same test), although they used them differently. One gave the pretest at the beginning of the chapter's work and the posttest at the end to show both the students and the parents how much the children had learned. The other gave the pretest a few days before the posttest at the end of the work on that chapter, more as an instructional and diagnostic device to help students do well on the posttest. Note that she is one of the teachers who is concerned about the comfort level of her students, and this test preparation probably provided a level of comfort as well as training for the "real" test. But however and whenever these paper-and-pencil assessments were used in the fall, the major focus was on recalling facts and doing computation. The pattern began to change by winter.

The early work in the math workshops was about assessing important mathematical skills, broadly defined, as in the NCTM *Standards*. The research team encouraged teachers to assess more broadly—that, in addition to competence with paper-and-pencil computation, it is important and useful to develop and assess children's ability to model numbers and procedures, make estimates of them, explain them, and solve problems about them. By winter all the teachers were trying to be more systematic in their observations of these abilities and were using problem-oriented computational tasks to assess them. They were requiring children to give explanations, both orally and in writing, of how they were performing procedures. For example, teachers gave students problems with missing digits to solve and to explain their solutions; they also gave them "buggy" problems to do and explain.

(See Appendix A for examples of tasks teachers were given to try; see Appendix C for examples of their assessments.)

The assessment of students' work on these problems in the winter was still at an informal level; that is, they were not scored and recorded in the grade book, merely noted for the information they provided about students. In addition to these more alternative tasks, most teachers continued to use some form of computational tests, either daily pages from the text, examples on the board, or chapter tests, and scores from these *were* recorded in the grade book. It was almost as if the alternative kinds of assessments were interesting activities for children but did not have the same weight for assessment as a computational test. This began to change in the spring.

One focus of the winter and spring math workshops was the scoring of students' explanations, both for explaining procedures and for explaining their methods of solving problems. Teachers developed a variety of general, and very brief, rubrics and applied them to students' work. By spring, all teachers were using students' problem solving and explanations for assessment, although two expressed concern that a child's problems with writing might mask his or her mathematical performance. Even so, all teachers adopted assessments that require written explanations, and they all noted that it was one of the major changes they had made this year. Two teachers tried to deal with the problem of poor communication skills by giving two scores—one for the answer and strategy used and the other for the explanation of the solution.

> And I found that for some, for many kids there are a lot of times [there's] a big discrepancy in whether they had a good strategy and whether they could really explain all of that strategy. And so I have now divided up my marking, a viable strategy and an explanation. Because I thought some kids need credit for their thinking even though they didn't write it out in words, but it's obvious to see the thinking that . . . Because like with [student] now, I mean there was nothing written, but actually after he told me the words I made sense of his picture.

Two teachers talked about giving a daily problem for "experience" but scoring only one each week. One of these teachers required students to write explanations only for the problem to be scored, while the other insisted that students write explanations daily. At least three teachers asked children to score their own and classmates' explanations for the instructional value it provided. As children worked on scoring explanations and saw many examples, they were more likely to internalize the criteria.

Even in the fall, all teachers talked about observing and questioning children, for instance, "Show me five groups of three." They all knew that these observations and exchanges were sources of valuable information about their students' understanding, but seemed not to consider them part of their program of assessment. Only one teacher kept systematic notes; and only one other expressed a desire to systematize her intuitions about what students know, and she placed the highest priority on learning how to make systematic observations. She also felt that she knew what each child knew but wanted to verify her "gut feelings." In fall she said:

> I'd like to be able to have more assessment that will give me some data to go with the gut feeling that I have. So that I could prove an understanding or a lack of understanding.

She also wanted checklists for proof of what children know and to help her plan instruction. In winter, her response to an interviewer's question (Why do you want checklists?) was:

> I think for proof. I think that if someone questioned me, you know if a parent said, well why, why this grade . . . either high or low, that I could say . . . well you know on this date when we were doing this, this is what I saw him do. . . . I think that it would be helpful to me too, to be able to after a lesson, just at a glance, look and see where kids are falling so that, you know, tomorrow I can maybe go to those kids first that are showing a weakness. . . . and one of the things that I find hard in math planning, is planning for a week at a time. Because what we do tomorrow depends on what happened today.

Two teachers were actively opposed to taking notes on these observations. They felt able to keep track mentally of where each student was and saw systematic recording of notes as cumbersome and burdensome.

In order to develop the assessment potential of observations, we made them another focus of our winter and spring workshops, primarily working on developing schemes for keeping systematic notes about students. Teachers developed checklists, used class lists with space for writing, drew grids with children's names in boxes, used spaces in their grade books for checks and other symbols, and even tried to use a copy of the assessment framework for each child to record how they were doing. All expressed frustration and doubts about these attempts. Sometimes a teacher's teaching style affected her ability to keep notes. Those who used direct teaching to the whole class had problems making individual observations. Those who had activity-based classes had difficulty getting around to each child and felt they wanted to give instruction every time they encountered a child with a problem. Some teachers who saw little value in systematic observation notes at the beginning of the year never became convinced of their value but felt they watched children carefully enough each day to know exactly who knew what and what difficulties they were having.

By spring, most of the teachers were trying to use systematic observations, some more successfully than others, but no teacher finished the year with a system for keeping anecdotal records that she felt worked well. The two teachers who tried to take systematic notes while observing children were overwhelmed by the amount of data they had for each child. They realized that anecdotal notes they had made could not be reduced to numbers recorded in a grade book. They thought perhaps that more selective assessment might be a solution for keeping the amount of data manageable. Two teachers seemed equivocal but convinced that they could keep the relevant information mentally.

Also by spring, the two teachers who had been using chapter tests were no longer using them routinely. One used no chapter test all spring, and the other said she used them only after critiquing them and judging them to be relevant.

> (But you also said you used the chapter test or some part of it.)
>
> Yeah, but now I am looking at it more critically. Before it just used to be part of the routine. I look them over and if I feel that they are relevant I use them. If I feel that they are not relevant I just move right on.

These teachers and one other seemed to prefer a balance between traditional and alternative forms of assessment, partially because the alternative assessments the teachers developed had some ambiguities in the directions.

> T: But I still think it needs to be a combination.
>
> R: What combination?
>
> T: Normal assessment and alternative assessments, I would never recommend to a classroom teacher to go with all alternative assessments.
>
> R: That's fine, and what are normal assessments for you, paper-and-pencil, computation?
>
> T: All these were paper-and-pencil.
>
> R: But see I look at, yeah so that's why I'm asking, what's normal? Is normal a chapter test, is normal computation?
>
> T: Like a standardized, a more standardized test because I think as we discover when you make tests there're always glitches in it. You know we've discovered that haven't we?

Also, teachers seemed more comfortable using new forms of assessment in the new instructional units they were trying, such as probability and multiplication. For the latter they were willing to select items from the Marilyn Burns unit and from tasks supplied by the research team; teachers at one school designed an assessment that was similar to the tasks they had developed for a unit on probability. Teachers' willingness to use performance assessments with unfamiliar topics occurred later in the year when they were becoming familiar with this kind of assessment, so it may be that as their comfort level rises, teachers would elect to use alternative assessments even with standard topics.

What is clear about the spring i that teachers were using many more forms of assessment than they had used in the fall, and that the nature of most these assessments had improved. They were focused more on children's thinking and on their performance on higher order skills. Teachers were observing children more carefully, and most were attempting to keep records of what they saw and heard. Most were willing to design their own assessments (with the help of their school team) even if only selecting from a set of tasks supplied by the research team. This was a change from fall when several teachers had been resistant to developing assessments, saying, understandably from their perspective, they did not care to "reinvent the wheel." One teacher was exceptional in her interest in and willingness to design many of her own assessments—some were extensions of those she was shown, and others were original. She also adapted an attitude measure from one she had for reading.

Teachers' knowledge of students. The second theme related to assessment is that teachers knew more about their students from performance assessments. Most teachers claimed performance assessments gave them new and deeper insights into children's thinking and understanding. They saw them providing much more information than whether a student can or cannot do something or whether a student "has it" or not.

T1: . . . Whereas before we were doing all of it but didn't, we didn't have them, the samples of work, we didn't have the collections and I think . . . even our kids have a better understanding of what we expect and what we're looking for that kids previously didn't.

T2: Well, I just don't think I ever really thought about math in terms of writing. It was more a numerical process, and I think being able to see how the kids explain through writing told me a lot about what they know and about their thinking process . . . kind of goes beyond the work sheet . . . be able to explain—not just answer but be able to explain it. It tells me a lot about them as thinkers. . . . Just, I think, getting the picture of a math student as a whole and not just one part of math, can they add on paper and subtract and multiply—it just goes much further than that.

R: Have you learned things about students' knowledge of mathematics that you otherwise might not have learned as a result of these assessment strategies?

T3: Yes, mainly that they can understand and explain to me what they are doing. Otherwise I would I just assume that they knew.

T4: Advantages? Um, I think through the assessments that we've been working with, children can . . . can . . . I mean you can, you can see if they're really understanding the process . . . much more so than just, you know, rote learning and doing what you're supposed to do.

. . .

I think you see how they are thinking . . . and how they problem solve better.

Difficulties with performance assessment. The third theme, that teachers had many difficulties with performance assessment, came as no surprise. The problems teachers faced were understandable and were proportional to the amount of change they attempted. Initially, difficulties had to do with lack of knowledge about what a performance task was, how to use it, and how to score it; and with observation, how to acquire and keep track of information about individual students and teach 25 others at the same time. We discussed above some problems teachers had with systematic observations and with scoring explanations, but they also had problems of a more general nature. For example, there were some initial misunderstandings at one school about teachers' perceptions of "teaching to the test," something they wanted to avoid. The teachers' interpretation was that their assessment tasks had to be very different from the performance tasks they had selected for instruction, and so, after using a wonderful set of instructional activities to teach place value, they chose a set of traditional worksheets for assessment. In addition to their misunderstanding, they believed then that paper-and-pencil computations were the definitive assessment for showing students' understanding of regrouping.

Teachers found it overwhelming to attempt changing their assessment program at the same time that they were changing their instruction in two major curricular areas (mathematics and reading).

> So, I feel like I could do such a better job and I said this thing before, if I was doing all reading this semester and all math next semester. I just think it would make it so much more manageable and I could focus so much more. I find myself going through the folder and I'm looking for what I need to have ready for you on Tuesdays and what I need to have ready for Freddy [the reading expert]. You know, I just, it's been a real management nightmare.

In the fall, many of the teachers saw the new assessments we asked them to try, and the new instructional activities they had requested, as add-ons to their regular instruction and assessment programs. Since they were trying to teach and assess everything as they had been doing, it was difficult to find the time to add the new instruction and assessments. And the assessments themselves took longer: Children take longer to solve a problem and write an explanation than to add some numbers. Scoring was also more difficult and more time-consuming: Rather than merely marking an answer correct or incorrect, each solution and explanation had to be read carefully enough to be scored. Another problem for one teacher was that scoring solutions to problems and explanations was too subjective and lacked the reliability of a standardized or chapter test from the text. Another felt performance tasks did not focus sufficiently on whether students know the facts and have computational skills.

The issue of children's comfort came up as a problem in these assessments, a concern we discussed earlier with respect to instruction. When children are given a problem as an assessment task, and they are not sure of how to solve it, they may be uncomfortable; they may ask many questions; they may whine; they may become unruly; some may cry, particularly if they have never felt the frustration of not being sure how to proceed. By training and selection, a teacher's response is often to want to tell children how to do things and to make them comfortable—just the opposite of what we were asking of teachers. By spring, most of our six teachers had adapted problems to their classes so that the level of difficulty was manageable, and they were rewarded with students who were enjoying the challenges. The early conversations about not giving an assessment task to a student unless you had shown the student how to do it were no longer heard in the spring.

Several teachers mentioned concerns about what parents might say if they did not send home tests of computation and if they used performance assessments instead. Despite the findings of another part of this study (Shepard & Bliem, 1993) that parents were overwhelmingly in favor of performance assessments, teachers feared that that would not be the case. Another teacher expressed surprise when parents were receptive to her including students' performance in solving problems as part of their grade. The resistance of their colleagues in higher grades to their working on mathematics other than facts and computation was also a problem for several of the teachers. Each school had a policy of requiring a certain score on timed tests of facts by the end of each grade, and this requirement seemed to hang heavily as a responsibility on most of the teachers. It is clear that the support of other teachers in the school and parents was important to have, and lack of it, real or perceived, was distressing to teachers.

> It's real frustrating because I know what the thinking is and I know what, pretty much what we're supposed to be doing. But then I was talking to a fifth-grade teacher the day before yesterday and she was saying how the kids don't know their facts and they can't do their computation skills. It's like we're being geared to do problem solving with the kids and all that, and then teachers in upper grades are upset because they're coming into them and not having the computational skills that they think they should have. One teacher does math timed tests and we hear, "No we shouldn't be doing math timed tests, that's not a valid way for kids to learn their facts." It's like being pulled in two different directions. And we can teach the problem solving and, at least we're trying to be able to do that. Not all people believe that that's the way—what we should be doing—and then we send our kids up to them, and it's like, "Could this child do their timed tests when they were in third grade?" Do you know what I mean? Don't you guys feel like that, like you're being pulled in two different directions and then parents come in and say, "I don't understand why my child doesn't bring home 25 addition problems every night to work on, what good is this going to have them do to count the legs on this animal."

It appeared that strong grade-level support was important and helpful to teachers, although even with such support, a teacher could still find the suggested changes too difficult to make. On the other hand, lack of team support did not appear to disturb another of our teachers, as she made significant changes in her instruction and assessment programs.

The difficulties teachers had with performance assessment were similar to those of making any change—not understanding how to do it, not having the time to take it on, thinking they had to add it to what they already used, being overwhelmed by what they were trying to do, doubting whether the change was sound, seeing that the change made their students uncomfortable, and feeling they lacked the support of other teachers and parents.

In summary, the effects of the first year of our project on teachers' practice of instruction and assessment were numerous. Teachers were using more hands-on activities, problem solving, and explanations for both instruction and assessment by spring. They were also trying to use more systematic observations for assessment. All teachers agreed that their students had learned more that year and that they knew more about what their students knew. Every teacher struggled with the revised instruction and new assessments, even those who endorsed them most enthusiastically. Many of the teachers used the word "overwhelmed" in referring to how they felt during the year, but they responded to feedback from their own classes about performance assessment and activity- and problem-based instruction. The feedback they got was generally positive; that is, their students seemed to have more conceptual understanding, could solve problems better, and could explain their solutions. Teachers' response, for the most part, was to attempt further change in their assessment and instruction practices and to become more convinced of the benefits of such changes.

Discussion and Conclusions

This paper reviews a year of work with third-grade teachers during which performance assessments were introduced in order to improve both instruction and assessment in mathematics. The major finding of the study is that participating teachers adopted many changes in their instructional practices (with respect to content and pedagogy) and their assessment practices (with respect to methods and purposes). Moreover, changes in assessment and instruction were, for many, mutually reinforcing. By year's end, many were using more hands-on and problem-based activities more closely aligned with the NCTM *Standards*, as intended by the project, to replace and supplement more traditional practices of text-based work, and they had extended the range of mathematical challenges they thought feasible to attempt with third graders. They used more varied means of assessment, for example, performance tasks and observations, that either replaced or supplemented computational and chapter tests. One teacher whose instructional practices already reflected NCTM *Standards* made even more progress in that direction, and she was able to adopt more authentic assessment practices.

In short, the introduction of performance assessment provided teachers with richer instructional goals than mere computation and raised their expectations of what their students can accomplish in mathematics and what they could learn about their students. There is a certain irony in teachers' concern with their students' comfort and their awareness that solving problems made students less comfortable than learning and performing computational algorithms. One of the goals of the *Standards* is to empower all students mathematically and to make them comfortable with mathematical thinking and problem solving. It appears that to accomplish this long-term goal, students may encounter some initial discomfort.

We list in the results section the many problems teachers reported as they realized the magnitude of the task of revising both reading and mathematics assessment. Then, as most teachers realized they also had to revise their instruction to prepare students for the new assessment tasks, they felt overwhelmed.

It is likely that most teachers also felt uncomfortable with some of the changes, and with being at odds with recommendations of the *Standards*. The teachers, as we would expect, adapted differently to the challenge of change. We can use a Piagetian model of assimilation and accommodation to describe teachers' reactions. Those changes in practice that fit a teachers' system of knowledge and beliefs were assimilated into that system. So a teacher whose belief system corresponded to the district goals was able to assimilate new practices without discomfort, for instance, making anecdotal notes about students. She was comfortable with the task and had to deal only with the amount of work it implied (still a chore, but not an onerous one).

Other teachers also assimilated practices into their belief systems, even when those practices appeared to be discrepant with their systems. They simply adapted the practice to fit their system; for instance, a teacher who believed children learn by being told would show children how to use base ten blocks in a directive manner. These teachers also felt little discomfort, but had the work (again, no small amount) of selecting and adapting the practices that could fit. For some of these teachers the discomfort came with having tried to make too many changes.

The teacher quoted above, who said, "I know pretty much *what* we're supposed to be doing . . ." had not incorporated *what* into her knowledge and belief system. It was still something being imposed from the outside, and so when she met resistance from other teachers and had her own doubts as well, she pulled back from that kind of teaching. She could try some things in a superficial way, but if they had no comfortable place in her system, she was not ready to modify her system.

Practices that made teachers uncomfortable were sometimes rejected, for example, letting students cope with a problem they had no idea how to solve. But if there were reasons why the practice continued to be attractive, the teacher was drawn in two directions (the disequilibrium Piaget talks about), and she began to change her system of knowledge and belief (Piaget's accommodation). We saw an example of accommodation in the teacher who talks about the project being a catalyst for change.

While we did not try to change beliefs directly, we know we affected beliefs through changes in practice. There is no doubt that changes in beliefs alter practice, but it is also the case that shifts in practice may lead to shifts in belief, which can, in turn, further affect practice. In this study the changes that teachers made were likely at first to be changes in practice. We saw teachers whose students gained greater understanding of multiplication from many hands-on activities change their belief about how to teach multiplication. As teachers got positive feedback from students about changes they had made in instruction and assessment, they were encouraged to attempt further changes. In other words, changes in beliefs and changes in practices appear to be mutually reinforcing. While this cycle appeared to lead to, for some, a fundamental change in instructional and assessment *practice*, it is not yet clear whether it also changed their *beliefs* about instruction and assessment.

We report many changes that teachers made in this project. What we cannot know is how durable or ephemeral those changes are. We know that some teachers made some changes superficially, adapting them to "fit," but other changes were made at more fundamental belief levels, and those will likely endure. Our work at two of the schools this year gives us confidence that, with continuing support, teachers are making even more changes. But, the question of the stability or persistence of the changes cannot be answered in real time.

What is abundantly clear is that the change that occurred did so not from anything we *told* teachers to do, but from their experiences with the ways performance assessments improved their classrooms. Just as we hope teachers will permit students to construct their own meaning from mathematical experiences, we must permit teachers to construct their own meaning for performance assessment.

It is important to ask if our intervention is a model for others. Not only was that not our intention, but it is most unlikely that the number of personnel (four university faculty, seven graduate students, and one visiting researcher) devoted to work with 14 teachers could be replicated in a school district. Like the teachers, we were also "messing about" with how to help teachers construct new views of assessment, and through that, of instruction and learning. There are things we would do differently and some other things we hope to try next year (the third year with these teachers), for example, administering some larger performance tasks at the end of this year, perhaps from the Maryland assessment, and then discussing student responses with teachers the following fall.

We learned some things about what and what not to do, and perhaps staff developers can benefit from our struggles and experiences. We know that teachers need a lot of support (from experts, administrators, peers, and parents) for changes they are expected to make, and they need to have some reason for wanting to make them. They need permission to go slowly and perhaps make what might seem to be quite small changes, and to be able to make them over a period of time measured in years, not months. Teachers need many chances to try things out with children (to mess about) and help in discussing and interpreting their classroom experiences. They need a lot of encouragement for all the extra time and hard work it takes to make changes. Staff developers must expect to see stops and starts, and even occasional backward motion. They need to remember that all teachers are *not* at the same starting point; that the same intervention will not work for all teachers; and that each teacher will adopt different changes that match her or his existing beliefs and practices. Staff developers need to know that change in instruction and assessment is not an all-or-nothing proposition—that teachers have it or they don't (or even that everyone agrees on what "it" is)—and that teachers can comfortably hold inconsistent views and engage in inconsistent practices for a very long time. Finally, they can also expect to see some teachers who don't want to play and will want to sit this one out, believing about performance assessment that "this too shall pass."

In conclusion, our results are not a clean sweep. They show it is not a matter of "show the assessment tasks, and teachers will use them," nor is it a matter of "have teachers use performance assessment, and they will change their instruction." Nor are we making an argument for high-stakes enforcement of externally mandated performance assessment. It's not about forcing. It's about a lot of slow, often painful, hard work for both teachers and staff developers. It's about the delight when the teacher who argues most vigorously about the changes says,

> I've changed my instruction. . . . I mean I have to; I mean if I'm going to assess kids differently, I have to teach differently.

References

Battista, M. T. (1994). Teacher beliefs and the reform movement in mathematics education. *Phi Delta Kappan*, *75*, 462-470.

Borko, H., & Putnam, R. (in press). Learning to teach. In R. C. Calfee & D. C. Berliner (Eds.), *Handbook of educational psychology*.

Burns, M. (1991). *Math by all means*. White Plains, NY: Math Solutions Publications and Cuisenaire.

Cobb, P., Wood, T. Yackel, E., & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. *American Journal of Educational Research*, *29*, 573-604.

Flexer, R. J. (1991, April). *Comparisons of student mathematics performance on standardized and alternative measures in high-stakes contexts*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Gipps, C. (Ed.). (1992). *Developing assessment for the national curriculum*. London: Kogan Page.

Koretz, D. M., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991, April). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Mathematical Sciences Education Board. (1989). *Everybody counts*. Washington, DC: National Academy Press.

National Council of Teachers of Mathematics. (1980). *An agenda for action*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1993). *Assessment standards for school mathematics—Working draft*. Reston, VA: Author.

Nelson, B. S. (1993, April). *Implications of current research on teacher change in mathematics for the professional development of mathematics teachers*. Paper presented at the annual meeting of the National Council of Teachers of Mathematics, Seattle.

Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational Researcher*, *19*(7), 10-18.

Romberg, T., Zarinnia, E., & Williams, S. (1989). *The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions.* Madison, WI: National Center for Research in Mathematical Science Education.

Shepard, L. A. (1989). Why we need better assessments. *Educational Leadership, 46*(7), 4-9.

Shepard, L. A. (1991). Will national tests improve student learning? *Phi Delta Kappan, 72*, 232-238.

Shepard, L. A., & Bliem, C. L. (1993, April). *Parent opinions about standardized tests, teachers' information and performance assessments.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Shepard, L. A., & Cutts-Dougherty, K. (1991, April). *Effects of high stakes testing on instruction.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 71*, 703-713.

Appendix A

Examples of Math Tasks Provided by the Project

R.J. Flexer
CRESST Project
C.U. - Boulder

Place Value, Borrowing and Carrying

1. Put 4 different one-digit numbers in the brackets to make
   the largest possible answer
   the smallest possible answer
   an answer between 50 and 60

```
  [ ] [ ]
+ [ ] [ ]
```

   How did you know what to choose?

2. Explain carrying to a second grader, using this problem:

```
   247
+  138
```

3. Jeff adds 62 and 73 on his calculator and gets 113. How do you know it's wrong?

4. Which is more and how do you know?

   324 or 432

   643 or 400 + 60 + 3

5. Sia's little sister wants to write two-hundred-forty-three like this: 20043. What would you tell her?

Gr 3

6. Find three two-digit numbers whose sum is 248.
   Is there just one answer?
   About how many answers are there?

7. Jo did a subtraction problem this way

$$\begin{array}{r} 425 \\ -\ 259 \\ \hline 234 \end{array}$$

   Is Jo right or wrong?
   What would you say to Jo?

8. Pick two numbers whose sum is 105 from this list:
   36, 91, 54, 47, 30, 58
   How did you do it?
   Now you make up a problem like this one.

Cryptarithms

9. Replace each letter with one digit to make the example correct. The same letter gets the same digit each time it is used in one problem. Some problems might have more than one answer.

$$\begin{array}{r} TT \\ +\ VV \\ \hline WYW \end{array} \qquad \begin{array}{r} JK \\ +\ \ H \\ \hline LMM \end{array}$$

Tic Tac Toe (Problem solving, estimating, performing addition both mentally and with paper and pencil)

Need: 5 markers in each of two colors
Tic Tac Toe board below.

Take turns with a friend. Each of you chooses a color marker.
Pick the place where you want to put your marker.
Then pick two addends that you think will give you that sum.
You must put your marker on the sum of the addends you pick.
Three markers in a row of one color wins.

Addends: 17 23 45 32 28

49	68	40
55	62	45
73	77	51

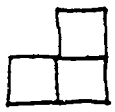
Can you make up a tic tac toe board with different numbers and addends?
Try it to share with a friend.

Buggy Problems

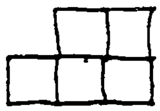
What would you say about each child's work?

Sam	Jeremy	Jill	Jack	Jeff
57	57	57	57	57
+ 26	+ 26	+ 26	+ 26	+ 26
73	713	83	101	31

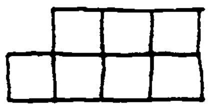
Building Buildings



1st building, 3 squares



2nd building, 5 squares



3rd building, 7 squares

How many squares are in the 8th building?

Building	Squares
1st	3
2nd	5
3rd	7

Explain

Appendix B

Coding Scheme

Tentative Coding Scheme: Revised

know-m	(what does it mean to know math)

instruction codes:

insgoals	(teachers' goals for mathematics learning and instruction)
insorg	(organization and management of instruction)
inswhat	(instructional tasks, activities, & materials; enacted curriculum)

assessment codes:

asgoals	(roles, goals and purposes for assessment)
ashow	(content/substance of assessment tasks; how teachers assess)
asscore	(scoring of assessment tasks)
track	(how to keep track of what students know)
grd	(how to assign grades in math)
aslm	(what do you want to learn about assessment in this project)
tdil	(teacher dilemmas)
rdil	(researcher dilemmas)
student	(student knowledge, beliefs, attitudes, performances in mathematics)

advantages and limitations:

asadv	(advantages of performance assessments)
aslim	(limitations of performance assessments)
projadv	(advantages of the project)
projlim	(limitations of the project)

NOTE: Also indicate instances where teachers talk explicitly about change by using a *delta*. Double code these instances--once with the "regular code" and once with the "delta code" E.g.,

delta-know-m & know-m for teacher's comments about changes in her ideas concerning wr. \t it means to know math

delta-aswhy & aswhy for teacher's reported changes in her ideas about the roles and purposes for assessment

Dimensions for Key Areas
Learning, Curriculum, and Instruction and Assessment in Mathematics

I. Beliefs and practice about how and what children learn

Direct instruction	Constructivist instruction
Kids learn from being told.	Kids figure things out themselves.
Memorizing is knowing.	Being able to use it is knowing.
Only some children can think math'ly.	All children can learn to think mathematically.
Children know their facts, procedures.	In addition, children can reason, solve problems, communicate.

II. Beliefs and practice about what school math is; what's important to learn, assess

Facts, computations, procedures, definitions, copying examples from text	Mathematical thinking, patterns, relationships, explanations
Math as the trivial, mechanical	Math as meaningful; making sense of math
Limited view of understanding	Extended view of understanding
Product	Process

III. Beliefs and practice about instruction and assessment

A. General

Uses textbook pages, worksheets; drill on facts, definitions, and computation	Uses worthwhile mathematical tasks that require thinking, reasoning, generalization, communication
T explains, shows how to do	T poses problems, asks questions, guides, orchestrates
Ss practice what they've been shown; memorize facts, definitions, procedures	Ss work on problems, discuss, report, question others

B. Problem solving

Story problems from text	Authentic, essential problems (everyday & mathematical)
Single answer	Open—multiple approaches, solutions
Well defined, very structured	Not well defined, unstructured
Contrived	Authentic
Only correct answer counts	Use of rubrics (criteria public); process valued

C. Explanations

Not requested	Seen as important—both as a skill and as a window to mathematical thinking
	Ss asked to explain and justify solutions

D. Instruction/assessment materials

Textbook, worksheets	Tasks to demonstrate, solve, discuss
Limited use of manipulatives, calculators	Open use of manipulatives, calculators

E. Additional Assessment Dimensions

Separate from instruction	Could serve as good instruction; enhances instruction
Limited data—timed tests, chapter tests, computation tests	Multiple sources of data—problem solving, observations, alternative paper-and-pencil tasks
Gut feelings about students	Systematic records about students
Assessment of what Ss have been shown	Assessment requires extension and application.
Learned nothing new about students	Learned significant new things about students
Doesn't assess activities, problem solving	Gets assessment information from non-p&p activities

Appendix C

Examples of Teachers' Assessments

Name ____________________ Date (Fall) ____________

Subtraction Test

1.
```
  48
- 19
```

2.
```
  658
- 247
```

3.
```
  428
- 379
```

4.
```
  700
- 236
```

5.
```
  $4.06
- 1.78
```

6.
```
  7,163
- 4,174
```

7.
```
  1,600
-   782
```

Estimate

8.
```
  78
- 63
```

Design a problem that you would have to regroup.

9.
```
  □□
- □□
```

Design a problem that would have an estimated answer of 30.

10.
```
  □□
- □□
```

Name ____________________________ Date ______(Spring)______

Multiplication Assessment

1. Draw a picture that shows 3 X 7.

2. Show all the possible ways you could arrange 24 chairs in rows. Use "x" to symbolize a chair.

3. Use 3, 2, and 5. Make as many combinations that give products under 20. For example: 3 X 2= 6

4. How many legs do 7 cows have?

5. Write a multiplication story that is solved with 4 X 5. The story must end with a question.