DOCUMENT RESUME

ED 378 224 TM 022 593

AUTHOR Flexer, Roberta J.; Gerstner, Eileen A.
TITLE Dilemmas and Issues for Teachers Developing Performance Assessments in Mathematics. A Case Study of the Effects of Alternative Assessment in Instruction, Student Learning and Accountability Practices.
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-364
PUB DATE Oct 93
CONTRACT R117G10027
NOTE 36p.; Paper prepared at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Accountability; Attitude Change; Beliefs; Case Studies; *Educational Assessment; Educational Practices; Elementary School Teachers; Grade 3; *Mathematics; Primary Education; *Teacher Attitudes; *Test Construction
IDENTIFIERS *Alternative Assessment; *Performance Based Evaluation

ABSTRACT

Issues involved in the construction of alternative forms of assessment by mathematics teachers were studied through the case study of assessment development in three elementary schools. Three schools with 14 third-grade teachers were selected and matched with comparison schools. Data are presented about the mathematics part of the study, along with explorations of the dilemmas and issues faced by teachers in all three schools, those unique to one site, and changes observed in teachers in each school. Teachers struggled mainly with issues in the area of beliefs and practical teaching knowledge. The difficulty was not so much in developing performance assessments as it was in believing that it was a worthwhile endeavor. The most disturbing dilemmas were those that focused on what was important to teach and how children learn. As might be expected, there were great differences in the amount of individual change by teachers, but some did adopt the concepts of performance assessment. (Contains 15 references.) (SLD)

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

Dilemmas and Issues for Teachers Developing Performance Assessments in Mathematics

A Case Study of the Effects of Alternative Assessment in Instruction, Student Learning and Accountability Practices

CSE Technical Report 364

Roberta J. Flexer and Eileen A. Gerstner
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

▸ The RAND Corporation

Dilemmas and Issues for Teachers Developing Performance Assessments in Mathematics

A Case Study of the Effects of Alternative Assessment in Instruction, Student Learning and Accountability Practices

CSE Technical Report 364

Roberta J. Flexer and Eileen A. Gerstner
CRESST/University of Colorado at Boulder

October 1993

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

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

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the 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 is one of four reports that document our progress in understanding these questions, based on case studies in three elementary schools.

DILEMMAS AND ISSUES FOR TEACHERS DEVELOPING PERFORMANCE ASSESSMENTS IN MATHEMATICS[1]

Roberta J. Flexer and Eileen A. Gerstner[2]
CRESST/University of Colorado at Boulder

INTRODUCTION

Let me set the stage for this paper by asking you to imagine that you are about to teach a basic statistics course. A team of cultural anthropologists and social constructivists offer to come help you decide how to assess the performance of the students in your class. Furthermore, they tell you they have novel ideas about what you should be teaching and how you should be teaching it. Your dean thinks it's a great idea and convinces you to "volunteer" to work with this team, to "welcome" the group, and to spend half a day each week for a year "discussing" how you should be teaching and assessing your course. The feeling you are now getting in the pit of your stomach might give you some indication of how our volunteer teachers might have felt about getting involved in this project.

The project was prompted by a prior study that showed the negative effects on reading and mathematics understanding of high-stakes standardized testing at the elementary level (Flexer, 1991; Hiebert, 1991; Koretz, Linn, Dunbar, & Shepard, 1991; Shepard & Cutts-Dougherty, 1991). It appeared that teachers focused their mathematics instruction on basic facts and computation; they also engaged in some questionable instructional practice under the pressure of preparing students to take standardized tests. As a result, students'

---

[1] This paper was presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April, 1993.

[2] We thank Abraham S. Flexer for his support throughout the project and for his editing of this manuscript. We also thank the team of graduate students for their many hours of work on the project, particularly the hours of sitting through meetings with teachers, transcribing tapes, and checking transcripts. They are Carribeth Bliem, Kate Cumbo, Kathy Davinroy, Maurene Flory, Bernice Harris, and Vicky Mayfield. 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 district administrators and personnel and of course to the hard-working teachers of Pine, Walnut, and Spruce Schools.

performance on the standardized tests did not generalize to other tests of the same material. Other studies also described wasted class time and energy spent in preparation for the end-of-year standardized tests (e.g., Smith, 1989). At the same time we hear from many teachers that the nature of their instruction in mathematics is dictated by standardized tests and that as long as low-level skills and definitions are the primary focus of testing, that is what they *must* teach. By contrast, there are many claims for the benefits for instruction of performance assessment (Wiggins, 1989).

This project sought not only to free teachers from standardized testing but also to assist them in developing a system of performance assessment. The idea for the mathematics part of the project was to make teachers *free* to instruct as they were being encouraged by the *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989; referred to in the rest of this paper as NCTM *Standards*) and to give them *support* to work on that instruction and on the assessment of their students being taught this way. The hope was that the refocus of assessment on performance, understanding, and higher order thinking would, in turn, have a positive effect on instruction. We have evidence from Gipps (1992) that work with performance assessment (the UK's Standardized Achievement Tasks, SATs) had positive effects on instruction for a significant number of teachers. We were not so naive as to believe this would be an easy task. We know that teachers' initial attempts with performance assessment are fraught with problems. Badger (1992) reports the concerns of teachers involved with the Massachusetts performance assessment.

> The most frequently voiced concern about this type of evaluation has to do with time. "If I am constantly evaluating, when do I find time for teaching?" Another concern refers to the amount of information that is collected. "How do I record all this material and how do I report it? Do I have to make a case study of each student?" (p. 10)

The literature on teacher change in general (e.g., Richardson, 1990) and teacher change specifically in mathematics (e.g., Nelson, 1993) discusses the difficulties of effecting change, the sets of conditions conducive to change, and some success stories.

General issues of teacher change often concern: (a) the organization surrounding the teacher; that is, the culture in which the teacher finds herself or

himself and the amount of support the organization provides; (b) the beliefs and knowledge of the teacher (in this case, about how children learn, about mathematics, and about the teaching and learning of mathematics); and (c) the active involvement of teachers and the acknowledgement of their control in making changes (Nelson, 1993; Richardson, 1990). These three categories are not disjoint and clearly interact; for example, the issues around a change that an organization might make that is not consistent with teacher beliefs belong in both category (a) and category (b), but the categories provide a useful perspective.

METHOD

This project sought to work on the content areas of reading and mathematics regularly (weekly) for a year with all the teachers at one grade level in several schools in one school district. We envisioned working with teachers to help them develop their own performance assessments and to support the changes the teachers wished to make in both instruction and assessment. We were interested in finding a school district willing to participate that had (a) standardized testing in place, (b) a large range in student achievement, and (c) considerable ethnic diversity. Such a district was found near an urban center and was selected for this study. The district population ranges from lower to middle socio-economic status. The research team decided to work with teachers of Grade 3, and only in schools in which a letter of application was signed by the principal, by all teachers at that level, and by the school's parent accountability committee. Three schools, with 14 third-grade teachers, were selected and matched with three comparison schools where data would also be collected. While all teachers were technically volunteers, it is possible that 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 teachers who found themselves involved in a project for which they had not volunteered; others may have been "strongly encouraged" to volunteer. By chance, all school personnel involved in this study are women (teachers, principals, district administrators), as are the researchers (faculty, graduate students, and visiting researcher).

The purpose of this paper is twofold: (a) to discuss some of the dilemmas and issues that arose in our first two terms of work with the participating

teachers, and (b) to give a progress report on how teachers have changed their instruction and assessment as a result of the project. The primary data for this paper are the interactions about mathematics instruction and assessment that the research team had with teachers at each school. Project staff met with teachers from different schools each week to work on mathematics and reading. Schools alternated each week between the two areas. These meetings were audio-taped and transcribed. The data also include classroom assessments discussed at those meetings (student work on those assessments was sampled and copied) and transcripts of some parent-teacher conferences. The research team also visited each classroom once or twice, to observe, to do demonstrations, or to model individual and small group assessment.

The transcripts of the weekly meetings were analyzed, and a coding scheme was developed iteratively as it was applied to successive transcripts.

RESULTS

This section presents data about the mathematics part of the project and our inferences in three areas: (a) the dilemmas and issues with which teachers in all three schools struggled; (b) dilemmas or positions unique to a particular school; and (c) changes observed in the teachers in each of the schools.

Dilemmas Encountered in More Than One School

Many issues, concerns, and problems that arose in all three schools came to be called "dilemmas" by the project staff. We address those as a group and give some indications of how different schools handled them.

The issue of organizational support was clearly crucial to this project, for inadequate organizational support could effectively block any efforts toward change by the participating teachers, whether in assessment or in instruction. In this project, there was very strong organizational support, and it was present at both the building and the district levels. Teachers and the research team had the support of administrators, curriculum specialists, and principals. That support came in both tangible and intangible forms: Teachers were aware that they had this support, and enjoyed many benefits, for example, extra planning time and budgets for materials they might need. The issue of organizational

context was mentioned by teachers as a problem for only one school and will be discussed later.

The most difficult problem for teachers and the first to emerge was the lack of time, a universal problem of teachers dealing with change, whether about assessment, instruction, or other kinds of reform (Badger, 1992). Whether this issue is classified as organizational or teacher-related matters not. It is a serious problem. First, all of these teachers had their regular, ongoing classroom responsibilities to deal with even as they were trying to make changes in assessment and instruction, and several had professional or community projects occurring simultaneously. Second, we asked that each teacher work in both reading and mathematics, although almost every teacher asked if she could focus on just one of those areas at a time. The problem of time was presented in several different forms. Teachers did not have enough time to select and prepare performance assessments outside the classroom; there was not enough time in the curriculum to add new things, and their time seemed fragmented; the performance assessments simply took longer than the traditional assessments, and the teachers did not like to lose instructional time. Teachers in every school found that they were spending more time on the topics we were working on as a project and began to worry about all the other things they were responsible for teaching. They considered instructional and assessment materials from the project to be an *addition* to what they were teaching. They wanted to do what they had always done, and *then* add the new materials if there was time. These comments came from teachers at Spruce, but are similar to comments repeated in each school.

> T1: You know what I'm finding the most difficult in this whole assessment piece that we're going through, is having two subject areas to work on. I think I could be doing a better job if I was focusing on math or reading. . . . You know, it's just been a real management nightmare.
>
> T2: Well, I think we're going to be a little bit slower because we're teaching this in a different frame of mind.
>
> T3: I'm kind of scared that—I mean I don't think there's anything wrong going on, I'm just scared—are we going to get everything in because they have to learn money and they have to learn telling time and we are doing addition and subtraction and fractions and you brought up multiplication and sometimes we try to throw in

division to get all this done when we are doing it so thoroughly. With this, there isn't anything wrong with it, it's just the time element, you know.

We and the district administration helped to address the time required outside the classroom in a number of ways: (a) We arranged to give teachers course credit for their work with us (while not freeing time, this gave teachers an extra benefit for spending their time); (b) the district provided teachers with a half day a month of released time to work on this project; (c) we reduced the amount of work expected of the teachers each week; (d) we staggered assignments for reading and mathematics, so that the teachers could focus on just one in a given week; (e) we supplied each teacher with a lot of resource materials so the teachers did not have to gather these themselves.

The problem of time within the classroom, we felt, arose because teachers viewed the new instruction and assessment as add-ons to their already full curricular schedule. We tried to help them see that instruction should take no longer if they did not try to do double teaching—the old and the new. We heard less about this problem as time went on and as teachers replaced some of their prior instruction and assessment with the new material or, in the case of multiplication, chose to use an innovative unit as their curriculum.

Teachers were learning what performance assessment was all about, and as good students, they expected to be challenged as they learned new things. There were five areas of assessment with which every teacher struggled:

1. what to assess
2. how to assess it
3. how to score the assessment
4. how to keep track of the results
5. how to report the results (grades)

In addition, most of the teachers also worked on instruction because the performance assessments were assessing a new curriculum that they had not been teaching (see below).

At the time of this study, the district was revising its mathematics curriculum and developing a framework consistent with the NCTM *Standards*. The district's "learning outcomes" for Grade 3 included communication,

connections, number sense, geometry, problem solving, probability and statistics, measurement, patterns, algebraic methods, and confidence with mathematics. In order to decide *what* to assess, each team of teachers had to select a subset of learning outcomes from the district's curriculum guide on which they planned to focus. We suggested that they select a small part of the curriculum for our initial work. Since third graders typically learn about place value, addition, and subtraction in the first three or four months of school, each school selected those areas and some combination of the processes of modeling, explaining, performing mentally, performing with paper and pencil, estimating, making connections, and problem solving. Their first task was to build an assessment framework for the selected outcomes. The framework was constructed with levels of proficiency that distinguish students who had limited understanding about the topic from those who knew more, and those students from students who knew even more. The problems the teachers faced in developing an assessment framework were different for each school and will be discussed below in the section on individual schools' dilemmas.

The next task for each school was to decide *how* to assess the learning outcomes they had selected. It had been our initial expectation that the teachers would develop assessment tasks using as models the many examples the research team supplied. Indeed, at the start, most teachers were not ready to develop their own tasks, but they were willing to select tasks to use from those models. (One teacher from Spruce Elementary School designed her own tasks from the beginning.) Our expectation then shifted to having the teachers try out a variety of tasks to see which ones gave good information and matched the new kinds of instruction. *Our* perspective was still on having the teachers develop a custom-tailored assessment program that they would find useful, and, in the process, having them learn techniques that they could apply to developing other assessment programs. The *teachers'* perspective was on assessing *their* particular students *this* year; they did not want to be in the business of designing assessments. Many teachers in the study have moved beyond a willingness only to select among choices to developing their own assessments. However, their purpose, reasonably enough, continues to focus on the information they derive about their particular students and not on which tasks make the best assessment tools by providing that information.

An interesting dilemma that emerged in at least two schools was that of giving pre-tests. Most of the teachers liked the chance to show the children and their parents how much the children had learned, and this could be done by showing huge gains from pre- to post-tests. There seemed no interest in using pre-tests to identify areas in which a child was already proficient and need not repeat material already mastered. The dilemma for some teachers was whether it is fair to a child to assess, using a pre-test, in an area that had not been taught, since many children found taking pre-tests an uncomfortable experience. As part of this discussion we probed whether or not there are some kinds of pre-tests on which a novice could still perform. We have had several discussions about this dilemma, and it continues to be a topic of conversation.

The issues on the topic of scoring assessment tasks were not controversial but still caused some struggle. Teachers did a lot of work on designing scoring rubrics for explanations in problem solving, using student work to help develop the rubrics. They wanted a general form that could be applied to many tasks, not task-specific rubrics. They used their project-related work on developing criteria for a good summary as a basis for establishing their own criteria for a good solution and explanation to a problem. Also, as they had done in the reading side of the project, some teachers asked the children for their ideas about what makes a good explanation, and as a class process, developed a set of criteria. These were then posted in the classroom for students' reference as they worked on problems. One teacher staples a reduced copy of the criteria to each problem a student has solved so that students can understand the basis for their score. Many different rubrics were designed, some of which incorporated both the answer and the explanation into one score and some which gave two separate scores. Teachers are currently applying the rubrics and refining them. (See Figure 1 for some examples.) It is not yet clear if the more elaborate rubrics will prove too unwieldy for regular classroom use; at this point, teachers with simpler schemes seem to be using them more often.

As the teachers in Badger's study (1992) noted, there is a lot of record keeping that goes with performance assessment. All the teachers found record keeping to be difficult because we were asking them to gather evidence about what children can do from sources different from worksheets and end-of-chapter tests. It meant gathering evidence on children's thinking from many sources, for instance, observations as children worked with manipulatives, verbal answers to

A.
4 = Good explanation. Right answer
3 = Good explanation. Wrong answer.
2 = Right answer. Confusing/incomplete explanation.
1 = Wrong answer. No attempt at an explanation.

B.
4 = Correct answer. Clear, correct explanation written out. Correct spelling.
3 = Correct answer. Clear process. Clear but not as involved explanation.
2 = Correct answer. Poor explanation.
1 = Wrong answer. No explanation.

C.
4 = Correct answer. Tells a lot. Well written.
3 = Correct answer. Doesn't tell as much.
2 = Close answer. No sense.
1 = No answer. No explanation.

D.
Answer
4 = Correct.
3 = Almost correct.
2 = Reasonable.
1 = No sense. No correlation to the problem
0 = No answer.

Explanation
4 = Logical. Tells how they got the answer. Thorough.
3 = Solid. Essentially correct. Maybe left out a component.
2 = Vague, unclear. Picture doesn't match (if there's a picture).
1 = No sense.
0 = No explanation.

Figure 1. Examples of scoring guides for problem-solving tasks. A, B, and C use a single score for both answer and explanation. D uses two scores.

specific probes from the teacher, responses as they played instructional games, contributions in discussions, and written explanations to tasks of problem solving and computing. Some of the record keeping meant having a place to put the scores from the rubrics they had developed. It also meant having a way to

keep track of how well a student could model a computation, like adding two-digit numbers with regrouping, or how something he said in a discussion rang a warning bell about his understanding. So it meant having a place for short anecdotal notes.

It is in the job description of elementary teachers that they watch what their children are doing, and most of the project teachers felt they had a pretty good idea of each student's level of performance. We were asking them to keep track of their understandings in a more formal, that is, systematic and written, way. Some teachers felt this was needless because they had it all in their heads, but all were willing to try it. Some responses from teachers are:

> T1: I think it's been, it's good information and it's really helpful, it's just kind of overwhelming with both [reading and mathematics].
>
> T2: It's real time consuming and I also found as I was doing it I found that we needed some sort of a code to match it, so that if they're working on it, or if you really think they've got it or if they're having a lot of difficulty you can mark that somehow.

Many felt the record-keeping process was unmanageable at first, but all have persevered with designing and trying out ways to do it. The teachers who use whole group instruction found the task particularly onerous and management became an issue. What were they to do with the students who were not being observed and assessed? Those who typically used small group instruction had an easier time with the management issues, and their primary task was to find a workable way to keep track of the children.

The 14 teachers in the project seemed to generate dozens of schemes for keeping track, many of which were discarded after the initial trials. Their common characteristic was boxes, small ones for checks or larger ones for anecdotes, and every student's name on one page. Only at the beginning of the project did some teachers try to use a scheme that had one sheet per child with a recording to show what the child had done in each area. That method proved to be too awkward for daily use; instead, there was a move to one sheet for the whole class. In some cases, one assessment activity was recorded at the top, in others two, three, or four, with notes next to each child's name for each activity. Some teachers adapted for mathematics record keeping a form the district had used for Writers' Workshop. It had boxes across the top with days of the week to record the primary activity for the day. Below that was a matrix of boxes, each

with a child's name and each about 2 square inches, for notes about the child's performance. All teachers continue to work on record-keeping plans that will be informative but efficient and reasonable enough to use daily in classrooms where lots of things are going on. Generally, keeping notes about individual students proved a chore for all the teachers, and most of those who persevered with it used abbreviated systems of checks, pluses, and minuses. Teachers who wrote more extensive notes tended to do that as a joint project with a member of the research team.

What can be said about report cards is that all the teachers in the project are dissatisfied with them. (Is that a universal truth for all teachers everywhere?) The district is currently working on developing new forms, and the dilemma is whether or not to use a form in which a child's performance is compared to a standard. Each school has its own form, and while different symbols and descriptors are used, grades are given essentially on a 3-point scale of superior, satisfactory, and needs improvement. Timed-test scores are also reported directly, which, as the teachers point out, imbue those assessments with considerable significance. The teachers have asked for assistance in redesigning their report cards, and that is in the queue for a future discussion.

Dilemmas concerning changes to instruction arose when teachers first looked at models of performance assessment and said that they were not suitable for their students who had not been instructed with that kind of assessment as the goal. Because the examples of performance assessment had been selected to fit the district curriculum, some of the teachers decided to make some changes in their instruction. Others continued to question parts of the district curriculum (see sections on individual schools).

All teachers, even those at odds with aligning assessment to the district curriculum, requested and seemed pleased with ideas for instruction that would be more in keeping with the NCTM *Standards*. There were fewer requests from Walnut Elementary School, but the same materials were prepared and delivered to all three schools. Every teacher used at least some of the ideas that the research team provided. Some of the teachers were already using many of the same kinds of activity. We tried to point out that the same activities could be used both as instruction and as assessment, as Badger (1992) so aptly states:

> [I]n more activity-oriented classrooms, no differentiation exists between the types of tasks that are used for instruction and those used for evaluation. Both should be interesting, challenging activities that reflect important themes in the mathematics curriculum. The difference occurs not in the kind of task but in the role of the teacher. (p. 10)

In many of the conversations we had with teachers we discussed the differences between assessment and instruction and how interrelated and intertwined they are. Some teachers looked at the tasks presented as "fun" things to do in the classroom but saw little value in them as sources of information about how children understand a piece of mathematics.

Individual Schools

This section characterizes each school in turn and presents some dilemmas and issues unique to each school or a position the teachers of that school took on a common problem.

Pine School

Characterization of Pine School. Pine School prides itself on being at the forefront of innovation in curriculum and instruction. Whatever is happening in elementary education, Pine is doing it. The third-grade teachers are aware of their position in trying new curricula and struggling with new ideas, and it is no surprise that the school volunteered for the CRESST project. The teachers talk about how exhausting it is to work on many new programs, but all seem eager to be involved in each new project, and they engage in them with energy and professionalism. Most of the teachers are well seasoned and have taught at the third-grade level for many years. They are a cooperative team and support each other strongly; they particularly support the least experienced teacher. They work well together in spite of divergent ideas about teaching mathematics. They eat lunch together daily and often discuss their classes and plan together during their short lunch break.

Manipulatives are available in the school, for example, base 10 blocks and pattern blocks, and the teachers use them to varying extents. Two teachers have taken the district's version of Marilyn Burns' *Math Solutions* course and use manipulatives regularly. One of these teachers has a goal for the year to work on problem solving and written explanations with her students, and she began to

do that at the start of the academic year. She has a chart of problem-solving strategies in her classroom, and encourages children to select an appropriate strategy to solve each problem she assigns; she then asks children to share the strategies they used to solve each problem. The second of these teachers also stresses problem solving and gives instruction on strategies. By the end of the fall term, the other teachers had incorporated problem solving into their curricula as well.

All the teachers use a popular commercial textbook, along with supplementary material from resource books in the school's collection. Most, though not all, instruction is addressed to the whole class. One of the teachers mentioned above divides her class into two groups so that she can teach one group while the other does practice exercises from the text. Instruction is didactic, clear, and carefully planned. Occasionally, teachers use learning centers which they plan, prepare, and use together, moving all third graders through each center. Pine teachers' team effort in planning and sharing materials is one solution to the time dilemma, a solution already in place in this school.

Their major assessment at the start of the project was multiple-choice chapter tests. Some of the teachers gave the text's pretest at the start of a chapter and the posttest at the end, so that children and parents could see the progress that had been made. The school requires that every teacher give timed tests on addition and subtraction facts and, later, on multiplication facts, so these were also part of assessment. Most of the teachers send drill-and-practice sheets home for students to work on with parents, and at least one teacher also sends sets of timed tests home for parents to use, score, and record. One reason they do this is to have more class time available for instruction of concepts and problem solving, rather than using it for drill and practice.

Dilemmas at Pine School. In the first six months of the project, issues and dilemmas arose for this group of teachers in almost every category. Their manner of dealing with dilemmas was to engage in lively debate with the research team and each other and to argue their positions forcefully. Rarely in these discussions did all the teachers take the same side. The team leader often stated a position that might be unpopular with the research team but did so to represent the views of several teachers. The more experienced teachers were the more vocal in these discussions.

The group used the district curriculum guide as they worked on an assessment framework for the performance assessments of the first semester. Because we had advised them to start on a small section of the third-grade curriculum, they selected place value, addition, and subtraction. The first issue that arose was the place of problem solving, which they saw as a separate topic that comes after children learn basic skills. The research team convinced them that problem solving belongs at every level, along with other processes like modeling and explaining.

While the teachers nominally accepted the district's framework for their curriculum, they interpreted it differently from the research team, and questioned some of the outcomes, for example, the first outcome of "communicating mathematical ideas."

> T1: I still question why kids have to write this all the time. I still question that. I still would like some real rationale for that.

This teacher felt that children who are less verbal and might be brief in their responses might be graded down; she said she didn't see why a verbal explanation wouldn't suffice. A second teacher objected to asking for explanations at all because they might penalize the child who did not have to figure a problem out, who just "got it" and cannot say how. The child's ability to get the correct answer is the important assessment, she asserted.

Under the heading of number sense in the curriculum guide is the outcome that a student "uses mathematical concepts and arithmetic operations with understanding." The teachers and researchers had many discussions about teaching for understanding.

> T1: But do you think, _____, that all students will totally understand?
>
> R: Well, that's what you're pushing for. That's what you're trying to get at.
>
> T1: But don't think you think that some children will only learn it by rote?

Later in that same conversation, after describing a child who must count each time to add, the same teacher says:

> T1: . . . 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 maybe 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 so sure that understanding has to come first. Because I think some children aren't capable of understanding.

But this same teacher then agrees with another who mentions using strategies to add, and says she teaches things like, "if it is adding nines then add ten and subtract one." She goes on to say, "There are very few facts that they really have to just memorize." She also takes the position that one need not mindlessly follow an algorithm, that starting with the tens column in an addition problem is a good idea if you are estimating. Most of her colleagues disagree totally—in adding you *must* start on the right. "Estimating is [a] different, see that's not adding."

As the teachers discuss this point, two recall that children in their classes are able to add two large numbers by first combining tens, for example, 28 + 47, "so they would say 20 plus 40 is 60. And then I know that 8 and 7 is 15. So that makes 75. I was just amazed." As the teachers discuss and reflect, they have a chance to revise their opinions to become more flexible in their thinking about algorithms. Some, though not all of them, do. But the process isn't one of steady progress. Just after this part of the conversation came:

T: But I guess my problem with that is that some kids get confused so easily, that by the time you have shown them five different ways of doing it, they would be totally confused. I mean I think there's a number of children, maybe three in a class, that just don't know.

Another issue arose in October during the selection of instruction and assessment tasks to go with the framework the teachers were using. The team had worked diligently and efficiently with materials they had gathered from the research team, their school collection, and the district's resource center. They had made excellent choices of instructional activities, but misinterpreted what a member of the research team had said about assessment activities. They thought the assessment tasks had to be quite different from the instructional tasks, and so they selected very traditional workbook pages for their assessments. They were satisfied to use these even after they were told of the misinterpretation, because they felt that after children had used manipulatives to gain understanding of a process, then the "bottom line" question was, could they do the process with paper and pencil. After all, they argued, they won't have manipulatives available when they're adults. Their inclination was to go to multiple-choice and paper-and-pencil assessment each time, claiming that the

ability to perform the operations showed a child had passed through the levels of understanding to a more meaningful result. They also felt more comfortable that a child had done his or her own work when it was privately produced on paper, as opposed to displayed on a table with manipulatives that all could see. The issue was not really resolved by discussion but by the researchers requesting that they use tasks similar to the instructional tasks for assessment along with their traditional forms of assessment.

Another issue for Pine was the need to prepare children for each assessment task, that is, to teach them every part of a problem they were asked to do. The teachers seemed to feel that children cannot figure things out if they haven't been shown how to do them first, that they can't hypothesize, conjecture, and test their hypotheses. All the teachers had these concerns in the fall, and they surfaced in discussions about problem solving, when they were encouraged to present some tasks that the children would never have experienced. One teacher said, "All of these are separate things that you would have to teach them." Another suggested that you would have to teach the strategy at some other time in the day. Only recently are they seeing that children can make discoveries for themselves and can investigate problems of their own devising. For one teacher this happened during their unit on multiplication, for which they were using Marilyn Burns' *Math by All Means*, when two children decided to make up a rule for finding prime numbers (numbers that could be represented by only two rectangles, 1 x the number and the number x 1). How interesting that these children had stumbled on a problem that has concerned mathematicians for hundreds of years!

Walnut School

Characterization of Walnut School. Walnut is a fairly new school that is well equipped with manipulatives and resource materials. Its teachers pride themselves on activity-oriented instruction. Much of the classroom work in third grade is done with children seated in groups. During the school year the third-grade students work on several large, integrated projects, and mathematics is taught in the context of these units. The teachers also incorporate mathematics from the daily events of the classroom, like lunch counts and the calendar. Several of the teachers have been teaching at the third grade for seven or more years. Most of the experience of another teacher is at lower grades, and another

teacher was new to third grade this year. Three of the teachers each have more than 17 years' experience. They are an energetic and focused group that sometimes works as a team, but more often in two subgroups. Several of the teachers have been to workshops on teaching mathematics and have large personal libraries of resource materials that they were happy to share with the teacher who had never taught third grade before.

The philosophy of the teachers at Walnut is very close to the newly revised district curriculum, although their assessments in the past have been fairly traditional, for example, worksheets, or a set of four computational examples and one nonroutine computational problem each week.

Dilemmas at Walnut School. Walnut teachers' first dilemma arose when they tried to establish an assessment framework. Their initial goal was to work on "reasonableness" with respect to computation, that is, the ability of students to look at an answer to a computational problem and decide if that answer makes sense. They were quite willing to broaden the application of reasonableness to other mathematical areas and began to build an assessment framework that had elements similar to those at the other schools: the content of place value, addition, and subtraction, and the processes of modeling, translation, explaining, and problem solving. One teacher said, "Things are so intertwined that you can do problem solving with place value." Initially Walnut was the only school that viewed it that way. Teachers tried to put the content on one continuum and "reasonableness" on another, with the processes crossed with both. Their final framework did not have "reasonableness" either as a dimension or a category, but it was their intent to use assessment tasks that probed students' abilities in that area. We had several conversations about the difference between a continuum and a rubric, and they decided to have a rubric for reasonableness.

They worked diligently and probably too intensively on the framework, in that they divided the content dimension into smaller and smaller categories, as if they were defining each task, rather than dealing with a broad area of accomplishment. When they finished with the framework, they felt it was uselessly fragmented and did not fit their teaching, particularly when they were involved in the large, integrated projects. The research team agreed, particularly the mathematics educator who was getting more nervous all the time about all the little boxes they were producing. It is ironic that teachers at

this school were already incorporating processes like modeling, explaining, and problem solving into their curriculum, but when they wrote it out on paper, the framework looked too onerous.

The framework may have influenced them to spend much more time on place value than they might otherwise have done because it had been partitioned into very small bits. But they also felt that spending a long time on place value was going to give their students a sound foundation for addition, subtraction, and multiplication. Of course that decision in September and October caused them panic in April when they realized how much of the third-grade curriculum they had not "covered."

Walnut's dilemma on how to assess had to do with the integrated instruction that they valued. Since they did not know how to assess such instruction, they assessed computation and computational problem solving in a weekly assessment, called Big-5. Students were given a computational exercise to do each day and possibly one involving some problem solving on Friday. They also gave the school-mandated timed tests, but put no emphasis on preparing children to take them. At least two of the teachers reported the results on timed tests during parent conferences. They did not ask children to write very much as part of their mathematics assessments at the start of the year, but began to do more in the winter semester.

Teachers at Walnut also struggled with the pre/posttest problem; they liked the idea of showing growth, but did not want to present their children with a painful pretest.

The issue of how to assess the mathematics in the integrated units remained a problem. We asked what observations they had made of children during their mini-society, a unit in which there were many rich mathematical opportunities. As part of their mini-society, students ran businesses and produced or bought items at wholesale prices that they later sold in a school store. Students had to take loans at 10% interest to buy the items, had to price them, sell them, and calculate their profit. The teachers were frustrated at missing opportunities for assessment, beyond some observation of practice at making change.

Walnut teachers have continued to ask for checklists for observation and for help with how to do it, even for daily work outside of the integrated projects.

The research team was fortunate to have a visiting researcher join it who could work each week with the teachers at Walnut. She has been modeling assessments of individuals and groups of children and holding conversations with the teachers regarding what she is learning about their children.

Spruce School

Characterization of Spruce School. The teachers at Spruce Elementary School have a wide range of teaching experiences—from a few years to about 25. As in the other schools, one of the third-grade teachers is new to that grade this year. The school seems to be fairly traditional. The teachers there value basic skills along with thinking skills, and they are eager to give their students a sound and thorough grounding in the basics. Spruce teachers use primarily whole-class instruction with a popular textbook and additional worksheets; they have a few manipulative materials available, though only limited supplies of each. These teachers' confidence with mathematics and its teaching was not as high as it was in other subject areas. Their assessments in the past have been end-of-chapter tests, worksheets, and timed tests.

As a group, the teachers were quiet in meetings and let the researchers do most of the talking. Perhaps they were overwhelmed by the whole project. Because they rarely objected publicly to the ideas of the project, and chose not to argue about points with which they may have disagreed, it was not always clear how they felt and what their problems were. As at the other schools, several teachers expressed the desire to keep the number of curriculum goals that we would work on to a minimum. One said, "I would really like to keep this narrow until I kinda get the feel for what we're looking at."

Most of the time, these teachers tended to work as individuals, but they always seemed very willing to share instructional or assessment tasks. They had little time together outside of their weekly meetings with us to work on the project as a team, except for the half day each month that they were released from teaching. They all took a class during the first term on using hands-on instruction given at their school by the school's half-time mathematics specialist. This specialist was also available on a regular schedule to demonstrate and teach with them in their classrooms. The presence of the specialist in the school had some positive, but perhaps also some unexpected and problematic, consequences (discussed below). The teachers learned a great deal from her and

received more resource materials to use. She also reinforced with them many of the project's ideas.

Dilemmas at Spruce School. It seems, in retrospect, that two things might have been happening as a result of working with the specialist and the project: (a) They were being pulled in a direction opposite to that of the other teachers in the school and of their students' parents; and (b) the teachers may have been losing control of their instruction. In December one teacher expressed this idea in a powerful way; the others concurred:

> T: 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 time tests and we hear, "So 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."

These teachers were finding it very difficult to reconcile other teachers' expectations, parental demands, school goals, and their own beliefs with the ideas being presented by the researchers and the mathematics specialist. While the teachers were supported by the principal and mathematics specialist to make changes within their school, it appears that the pressure they perceived from other teachers was very much in another direction. Recall that one of the issues in the literature on teacher change is the importance of the surrounding organization—its support and context. The Spruce teachers were having a very difficult time within the culture of their school when nominally it appeared that they had the district's and school's support, but in practice, they felt subjected to countervailing values.

An issue for teachers making major changes is that of control; teachers are in control of their change, and that must be acknowledged (Richardson, 1990). As these conflicting pressures continued, the Spruce teachers seemed to become less decisive about their curriculum, and less in control of their instruction. Perhaps this loss of control had some effect on their interest in making other changes. By the winter semester, they were having a difficult time planning what their curriculum would be for the remainder of the year. The mathematics specialist was scheduled to be in their classrooms several days a week for about a month, and she seemed to take over the unit on multiplication. As in the other schools, they were using the Marilyn Burns' *Mathematics by All Means* unit, but the specialist was introducing it, planning all of the instruction, and teaching the classes with the assistance of the teachers. When asked what they did in mathematics on days that the specialist was not in their rooms, they joked that they didn't do any mathematics.

The dilemma of what to assess (and even what to teach) was very much a problem for the Spruce teachers, although they did not complain or argue about it. One teacher seemed to speak for the group when she said:

> T: 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.

A major concern for the Spruce teachers was the role of the timed test, a topic that came up at many of our meetings. The teachers thought that, according to district policy, all teachers were required to give their students timed facts tests. Each school had a different version of who required these tests, and it was half way through the year before we found out that they were not required by the district, but that each school set its own policy with respect to the administration and standard (e.g., 80% of 100 facts correct in 5 minutes).

The Spruce teachers seemed to take these tests more seriously than teachers at the other schools. They gave each other advice on how to improve scores and made many comments about the tests, including one that pointed to the inability of other kinds of assessment to prepare students for timed tests.

> T1: Try this, try telling them they have to finish it. And just write their time down. Keep track of their time. And they have to finish it. It's amazing how much faster they get.
>
> R: Is the timed test a policy of the school?
>
> T2: Yes. We've talked about it the last couple of years. It has been a topic of discussion but as of right now it's on the report card and our kids are expected to do 100 problems in 5 minutes. . . . So you're thinking that it shouldn't really carry that much weight with the time limit on it?
>
> T2: The fifth-grade teacher, how can I say to her, "We just don't feel that math timed tests are important for kids anymore?"
>
> T2: And if they're supposed to get, I mean if the school's goal is to get 80%, 100 problems right in 5 minutes, how else can you do that except to give them a timed test? How else can you get that data?
>
> T3: And I think the timed test is used to see how they're doing and this stuff is working so they can improve their score. Well, I use the timed test for assessments, is what I'm saying.

When discussing record-keeping for the project, the Spruce teachers were no different from the other participating teachers. All of them complained about the amount of time required to keep records. For example, one teacher, in talking about making individual observations of children working, said:

> I think it's been, it's good information and it's really helpful, it's just kind of overwhelming with both of them [math and reading].

Another said:

> It's real time consuming and I also found as I was doing it I found that we needed some sort of a code to match it, so that if they're working on it, or if you really think they've got it or if they're having a lot of difficulty you can mark that somehow.

They felt the record-keeping process was unmanageable at first, but several of the teachers invented good forms to be used while doing observations, and one of

the teachers was able to use hers effectively. We shared the forms with teachers at other schools.

Although the project dealt with assessment, a great deal of time was spent discussing instruction with all the teachers. The Spruce teachers used traditional teaching methods, but wanted to learn about other ways. They focused on learning new methods and using new materials, not on how instruction is related to assessment. We spent a lot of time gathering materials and discussing how to use them in instruction or as assessment. Calculators, whose use was encouraged in the district curriculum guide as it is in the NCTM *Standards*, were not used because, as one teacher explained:

> T: . . . and that's why we don't give them calculators today until they learn it, learn the process and the understanding of the process because they can take a calculator and punch in the buttons and get the answer anyway.

There was some limited use of manipulatives, and as teachers learned what could be done with manipulatives, they used them more. They seemed to be interested in obtaining more manipulatives, but this did not occur despite the nominal availability of funds; possibly because of reliance on the mathematics specialist who had a supply of materials she used during her instruction.

The dilemmas at Spruce Elementary School seemed not to apply to all teachers equally. One not only used many of the supplied tasks, she began early on to invent her own for both instruction and assessment. For the most part they were exemplary performance assessments, and she was willing to share them with the other teachers at her own school as well as teachers at other schools. She also devised new continua as she moved into new topics in the curriculum, and these reflected the spirit of working on both content and the processes of modeling, communication and problem solving. She designed a Likert-type self-assessment, using happy faces for markers, that the children filled out for her.

Changes

What changes did we see in the teachers' instruction and assessment? There were both school effects and individual changes. The good news is that we saw lots of change (albeit some in small steps), and it was in directions that reflect the NCTM *Standards*. The discouraging but expected non-news is that

change is slow and nonlinear. And the more profound the change, the slower. Had we not gotten the latter results, you would have good reason to doubt our story.

Changes at the Level of the Schools

The Pine School teachers changed their assessment and instruction the most. They came from questioning all of the process strands beyond showing paper-and-pencil computation to a quite different view of assessment. In an April meeting, for example, they were asked what they planned to teach about geometry. The team leader immediately said, "Well, we first have to ask, what do we want kids to know?" Then she and the other teachers made suggestions well beyond the level of the identifications they had worried about in October. Their ideas for assessments also involved more active and engaging tasks, and they just expected that observations of the children working would be a good source of information.

The Pine teachers also developed a unit on probability from a variety of sources—materials the research team brought, books in their resource library, and materials from the district resource collection. They spent one session with the research team discussing ideas about probability, its pedagogy, and its assessment. Pine's instructional activities were solid, interesting, instructive, activity-based, and attractive to the children. Their assessment for this unit could be considered alternative, including assessment of problem solving and communication. As stated earlier, the instruction that followed the probability unit was Marilyn Burns' multiplication unit, and its instructional design called for lots of activities for small groups. It is impossible to say whether our assessment project had anything to do with their selection of this unit. One of the teachers on the team had seen the unit in December, told the others about it, and they had their books in January. They supplemented the suggested assessments in the Burns unit with a set of tasks they developed that was much more open and conceptual than the end-of-chapter tests they had been using, although it was still light on problem solving and communication.

The Pine teachers continue to use the end-of-chapter tests, but they now supplement those with other assessments that involve more conceptual understanding and higher order thinking. They also like representing a concept in many different ways; for example, they decided to include arrays of dots in

their next assessment of multiplication to see if children could extend the idea from the arrays of tiles they had been using.

The Pine teachers were willing to eliminate some topics from their usual curriculum (e.g., division into a three-digit number) in order to do more work on problem solving and to do geometry more thoroughly.

Pine teachers frequently expressed surprise at how well the students were doing, how much they knew, and how much they enjoyed math. The teachers see evidence that the children are getting better at problem solving and at writing explanations for their solutions to the problems. In the first months, these comments were often accompanied by doubts of some kind about the assessments; we don't hear those much any more. (Perhaps they got tired of telling us.)

The conferences the Pine teachers had with their students' parents included a discussion of the new assessments and how important these are and how much they say about what their child knows and can do. We take as good signs that the teachers value the performance assessments sufficiently to discuss them with parents and that the assessments were presented positively and convincingly. It is hard to say at this point whether these changes represent the beginnings of fundamental change in beliefs, or less substantial change in instruction. It is possible that some of the Pine teachers are at the beginning of an epistomological shift. We believe that belief and practice can be causally related in both directions, that is, that the shift in practice may lead to a shift in belief which can lead to further shifts in practice. As these teachers are reinforced by seeing that their students are learning more, for instance, seeing their students talk about the connection of multiplication to geometry, they may change their ideas about how children learn mathematics.

The Walnut teachers did not change their instruction as much as the Pine teachers did, but that was because the Walnut teachers were already using lots of hands-on, small group activities. For example, Walnut also used the Marilyn Burns multiplication unit, but they had used it last year as well. When asked in December what they were doing differently in math, one teacher replied:

> T1: In math, I think for me the biggest thing is to get the kids to verbalize their thinking. Not just to have a strategy and apply it but then, then to regurgitate the strategy and what they were thinking.

R: To recount the strategy.

T2: Well, yeah. I mean and to just, to go back. I think there have been times that we have said, "Tell me how you got that," and the kids explains that and we say, "Great." This year we say, "Tell me how you got that and put it here [on paper]," and I have proof to show parents and to show me.

R: So tell them to write [it] down?

T2: I know that I am doing more writing [by students]. I did some before but I have to say it was very limited.

T1: 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?

Walnut's changes in assessment involved learning more about how to observe and question children one-to-one or in small groups and how to take note of what individuals are doing in large groups. In the multiplication unit, they used the assessment tasks suggested by the text and, in addition, designed an assessment that was alternative in nature. Several of the tasks on the Walnut-designed assessment probed the concept of multiplication and the commutative property; several also required explanations, including a traditional textbook-type problem.

The changes in some of the Walnut teachers seemed more subtle and deeper. They seemed to be struggling further to align their beliefs about pedagogy with their practice. They were interested in and eager to discuss some very fine points, like whether a particular rectangle should be referred to as 2 x 3 or 3 x 2. They were willing to go beyond the idea of "which is correct" to how children could look at it in different ways and what the implications of that might be, and to consider whether one was conceptually more sound than the other. The change for some Walnut teachers was not so much in adopting different beliefs, but in moving further along the lines of the beliefs they already held.

Teachers at Spruce incorporated into their classrooms many of the tasks supplied by the research team. They requested instructional ideas, were willing

to use them, and were pleased with the results. Spruce teachers were aware that their students seemed to understand mathematics better and were enjoying it more. One of the teachers said,

> T: They sure love math I tell you . . . You know they love these games we're doing. They just understand it so much better and I'm doing a better job of teaching it.

The same teacher viewed their role as one of piloting the program. She felt they had learned a lot and they would do much better next year.

One of the teachers at Spruce seems to have made major changes in both her instruction and assessment. She seemed intrigued by the ideas of the project, was willing to try some new things, and produced her own performance assessment tasks, many quite fine. These tasks often incorporated computation, whether they involved problem solving or spatial relations. She also produced assessment frameworks for each new topic and developed a Likert scale for students' self-assessment. It is too early to say if a change in epistomolgy is occurring, but it is clear that she owns the ideas of the project and uses them to extend her own teaching.

The culture of this school may have made it difficult for teachers at one grade to make significant changes. Even if they all agreed that they wanted to change, they feel pressured by teachers at higher grades and perceive that their parents don't want them to.

CONCLUSION

The aim of this project was to help teachers adopt performance assessment in their third-grade reading and mathematics classes and to free them from the constraints of standardized testing. As expected, the preliminary results were mixed, but hopeful.

This paper looked at the problems, issues, and dilemmas teachers encountered as they attempted performance assessment in mathematics for the first time. These issues were primarily in the area of beliefs and practical teaching knowledge. For some of the teachers the real struggle was not in learning how to use performance assessment, but in believing that it was a useful thing to do. The dilemmas most disturbing to them were those that focused on what is important to teach (and therefore assess) and how children learn. For those teachers whose beliefs about those issues were at odds with

those of the district's new curriculum and performance assessment, the issues were more profound and the teachers' initial changes more superficial. Even so, some teachers appeared less disturbed by dilemmas that struck at core values. The changes those teachers made, even if not profound, may be of the kind that will eventually cause their beliefs to change: As they see positive effects in their students from a few changes, they may be reinforced for the changes and continue to do more and more until they eventually shift what they believe. The teachers whose beliefs were consonant with the new views appeared more comfortable with their dilemmas and were able to focus on more fundamental changes.

Was the greatest change seen in schools initially closest in philosophy to the district's curriculum? . . . farthest from the district's philosophy? Although Walnut's philosophy is most closely aligned with the district's curriculum, in the first term this school produced less in the way of performance assessments than either of the other schools. Perhaps they were doing more that did not get assessed, for example, in their integrated unit; or they might have seen the tasks suggested by the research team as an intrusion on an already rich curriculum. In that first term, they tended to assess mainly computation, adding a nonroutine, computational problem-solving task each week. They did little writing in mathematics assessment until the second term. In contrast, Spruce's mathematics folders were packed with assessments they had tried, even though they appeared to be the least enthusiastic about the district's new curriculum and the new assessments presented by the research team.

If change is measured by how well the ideas about performance assessment become integrated into the teachers' repertoires, then we cannot look at schools, but must think about individual teachers. Each school had a teacher or two or three who appeared to be engaged by performance assessment. The teachers in this group fell into two categories: (a) teachers grappling with ideas at deeper levels of belief and using them in their general movement toward teaching from a more constructivist perspective of learning; and (b) teachers who adopted the ideas into their current belief system, revising them to make them their own. Some of the teachers who changed the least may have felt they currently had quite adequate or even superior teaching programs and saw no reason to change. Others found themselves in too much disagreement with these notions; the ideas did not match their current views and therefore proved of little use or value.

Others continued to be overwhelmed by trying out new instructional and assessment ideas in two major areas of the curriculum.

Given that teachers were in very different places in their beliefs about mathematics and about teaching and assessing mathematics before this project began, it was no surprise that there were great individual differences in their dilemmas with performance assessment, their reactions to those dilemmas, and the changes they made as a result of the project and their struggles with the dilemmas.

REFERENCES

Badger, E. (1992). More than testing. *Arithmetic Teacher*, 7-11.

Burns, M. (1991). *Math by all means.* New Rochelle, NY: Math Solutions and Cuisenaire.

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.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18*(9), 27-32.

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

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

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.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* 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.

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.* Boston: Kluwer Academic Publishers.

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

Shepard, L. A. (1991). Will national tests improve student learning? *Phi Delta Kappan*, *73*, 232-238.

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. (1989). *The role of external testing in elementary schools.* Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

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