Teachers' Developing Ideas and Practices About Mathematics Performance Assessment: Successes, Stumbling Blocks, and Implications for Professional Development

CSE Technical Report 423

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

CRESST/University of Colorado, Boulder

January 1997

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 90024-1522
(310) 206-1532

The work reported herein was supported in part 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 policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

TEACHERS' DEVELOPING IDEAS AND PRACTICES ABOUT MATHEMATICS PERFORMANCE ASSESSMENT:

SUCCESSES, STUMBLING BLOCKS, AND IMPLICATIONS FOR PROFESSIONAL DEVELOPMENT[1,2]

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

CRESST/University of Colorado, Boulder

Abstract

This article focuses on the process of change experienced by a group of third-grade teachers and the factors that facilitated and impeded that process during their participation in the University of Colorado Assessment Project. The project's purpose was to help teachers design and implement classroom-based performance assessments compatible with their instructional goals in mathematics and literacy. We examine the change process in mathematics by analyzing conversations between teachers and researchers during workshops conducted throughout the school year and interviews conducted at the beginning, middle, and end of the year. Results are organized around five themes: (a) Situating the change process in the actual contexts where new ideas will be implemented is an effective strategy for helping teachers change their practice; (b) Group discussions can be an effective tool for the social construction of new ideas; (c) Staff development personnel can facilitate change by introducing new ideas based on teachers' current levels of interest, understanding, and skill; (d) When teachers' beliefs are incompatible with the intentions of the staff development team and are not challenged, the teachers are likely to either ignore new ideas or inappropriately assimilate them into existing practice; and (e) Time is a major obstacle to changing classroom practice.

---

[1] The CU Assessment Project is part of a larger research project, Studies in Improving Classroom and Local Assessments. It is supported in part by the Office of Educational Research and Improvement, through the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not reflect the views of other project team members, CRESST, the Office of Educational Research and Improvement, or the U.S. Department of Education.

[2] We would like to thank the CU Assessment Project research team, faculty researchers: Hilda Borko, Roberta Flexer, Elfrieda Hiebert, and Lorrie Shepard. All members of the team contributed to the conceptualization of staff development and data collection for this study, and many provided extensive feedback on drafts of this article. We extend special thanks to the fourteen teachers who spent countless hours on project-related activities. Without their commitment and support, the project would not have been possible. We also wish to acknowledge colleagues outside the project who provided feedback on the article: Margaret Eisenhart, Ralph Putnam, and Maria Timmons.

This article focuses on the process of change experienced by a group of third-grade teachers and the factors that facilitated and impeded that process during their participation in the University of Colorado (CU) Assessment Project. The purpose of the project was to help teachers design and implement classroom-based performance assessments compatible with their instructional goals in mathematics and literacy. In this article, we trace the teachers' change process in mathematics by analyzing conversations about scoring tasks and keeping records between teachers and researchers during workshops conducted throughout the school year and during interviews conducted at the beginning, middle, and end of the year.

Research Background and Purpose

The CU Assessment Project is, in part, a response to the national interest in performance assessment that has grown in the U.S. in conjunction with the current educational reform movement. As one component of that movement, the educational community is calling for new approaches to assessment that better address educational goals such as higher order thinking, reasoning, problem solving, communication, and conceptual understanding of subject matter (e.g., Resnick & Resnick, 1992; Stiggins, 1991; Wiggins, 1989). Similar calls to adopt performance assessments have occurred in other countries (e.g., Black, 1994; Madaus & Kellaghan, 1993).

In mathematics, for example, the *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989) identifies assessment as an integral part of teaching and specifies that the main purpose for assessment is "to help teachers better understand what students know and make meaningful instructional decisions" (p. 189). Assessment is to be aligned with the curriculum outlined in the *Standards* with respect to mathematical content and instructional approaches. Teachers are to use multiple assessment techniques that demand different kinds of mathematical thinking and include a variety of formats. Assessments should yield information about students' "mathematical power," including the knowledge they possess as well as their ability to use, apply, and communicate that knowledge. The mathematics curriculum framework of the district within which the CU Assessment Project was conducted is compatible with the vision of school mathematics reflected in the National Council of Teachers of Mathematics (NCTM) *Standards*. The project

hoped to help participating teachers develop or select classroom-based performance assessments consistent with that vision.

There are a number of unanswered questions about performance assessments ranging from their psychometric properties (e.g., Linn, Baker, & Dunbar, 1991; Shavelson, Baxter, & Pine, 1992) to their impact on teaching and learning (e.g., Resnick & Resnick, 1992), to their implications for professional development (Stiggins, 1991). This article addresses one of these questions—the ramifications of classroom-based performance assessments for professional development. (Findings from the CU Assessment Project related to the impact of performance assessment on teaching and learning are addressed in Borko, Davinroy, Flory, & Hiebert, 1994; Flexer, Cumbo, Borko, Mayfield, & Marion, 1994; and Shepard et al., 1994.)

Implications for Professional Development

The new visions of learning and teaching underlying educational reform are making profound demands on teachers. If they are to move successfully toward these visions, many teachers—experts and novices alike—must make major changes in their teaching practices, as well as in their knowledge and beliefs about teaching, learning, and subject matter (Borko & Putnam, 1996). The transformation teachers are being challenged to make is one they are unlikely to make without support and guidance. As a consequence, a number of researchers studying educational reform are calling for increased attention to support for teachers (e.g., Borko & Putnam, 1996; Carpenter & Fennema, 1992; Cohen & Ball, 1990; Richardson, 1994).

The educational community is not alone in its realization of the importance of assisting teachers in their efforts to comply with the demands of educational reform. For example, in the U.S. the Clinton administration has made professional development a major emphasis in its education agenda, "arguing that teachers need more sustained, intensive training to prepare them to teach for high standards" (Bradley, 1994, p. 20). The administration's commitment is reflected in the new goal that was added to Goals 2000: Educate America Act, which "calls for teachers to have access to programs for the continued improvement of their professional skills" (p. 20) and stipulates that states applying for federal school reform grants include plans for professional development in their proposals. Similarly, Title II of the recently reauthorized Elementary and Secondary

Education Act (Improving America's Schools Act of 1994), intended to improve professional development, emphasizes "more sustained, long-term efforts" rather than "short-term, one-time professional development activities" (U.S. Department of Education, 1994, p. 2).

Given the increasing commitment to professional development on the part of the educational community and federal government, it is important to identify features of staff development efforts that are successful in fostering teacher change. This article addresses the topic of professional development by exploring changes that occurred in the knowledge, beliefs, and practices of teachers participating in the CU Assessment Project. Our analysis focuses on two assessment topics that received attention at all three participating schools—the scoring of open-ended mathematics tasks, and systems for recording information derived from observations of students during mathematics activities. Within the context of these topics, we look closely at issues the teachers faced, the staff development/research team's role in helping them to explore these issues, and the ways in which the teachers embraced and resisted the change process. We speculate about possible reasons for commonalities and differences among the teachers, and about ways in which future staff development efforts might better assist teachers to change their practice.

Framework for Staff Development

Our approach to the mathematics staff development workshops was guided by the overall purpose of the CU Assessment Project—to help teachers design and implement classroom-based performance assessments compatible with their instructional goals—as well as our beliefs about teacher learning. We believed that teachers would learn best by actively constructing new assessment ideas and practices based, in part, on their existing knowledge and beliefs (e.g., Prawat & Floden, 1994), and by sharing ownership of the workshop content and processes. These ideas fit what Richardson (1992) referred to as the new generation of staff development programs—programs that are cognitively framed and based on constructivist views of knowledge and learning.

At the same time, we shared Cobb's (1994; see also Cobb, Yackel, & Wood, 1992) view that the knowledge and practices of the wider community must be considered in the picture of knowledge construction. The constructivist view of learning we hold does not imply that any individual or learning community is free

to construct their own private truths, or that any interpretation is as good as any other. Using mathematics teaching as an example, Cobb and colleagues (1992) argued that "the goal is still that students eventually construct correct or true mathematical understandings with the teacher's guidance. . . . From this perspective, mathematical truth is accounted for in terms of the taken-as-shared mathematical interpretations, meanings, and practices institutionalized by wider society" (p. 16). Applied to staff development, this perspective suggests that staff developers provide the guidance that will enable participating teachers to construct meanings and practices compatible with those advocated by the educational reform community. In mathematics education, this community is represented by the National Council of Teachers of Mathematics and documents such as the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989), *Professional Standards for Teaching Mathematics* (NCTM, 1991), and *Assessment Standards for School Mathematics* (NCTM, 1995). We planned to structure workshop sessions to take into account both the teachers' existing ideas and practices and the mathematics education community's vision of classroom assessment.

Research Design and Methods

Participants and Context

Participants in the study were 14 third-grade teachers from three schools in a lower-to-middle-class school district on the outskirts of Denver, Colorado. Criteria used to select the district included willingness on the part of central office professionals to participate, an ethnically diverse student population, a history of high-stakes standardized testing, and district curriculum guidelines in mathematics and language arts consistent with national standards (Anderson, Hiebert, Scott, & Wilkinson, 1985; NCTM, 1989; see also International Reading Association/National Council of Teachers of English, 1994). In addition, the district had to be willing to waive standardized tests in the participating schools for two years, because we wanted teachers to be free to implement performance assessments and to make concomitant changes in instruction without worrying about how their students would do on standardized tests.

Due to the amount of time and effort that would be required of teachers, volunteer schools were sought. Third-grade teachers had to make a commitment

as a team with the support of their principal and parent accountability committee, and to indicate their commitment by submitting a written application to the district and CU project staff. We accepted the three schools that submitted applications.

Largely because of this application process, we entered into the study with the assumption that all of the teachers were enthusiastic about both the project and the national reforms in mathematics and literacy education. We also assumed they had instructional programs in place that were generally consistent with these reforms and with the district curriculum framework. Based on these assumptions, we negotiated agreements with the teachers that focused on creating assessment practices compatible with their existing mathematics and literacy goals and instructional programs. We did not discuss changing their instructional goals or suggest examining their underlying beliefs about learning and teaching.

We later found that our assumptions were not accurate for all teachers. Although all 14 teachers were technically volunteers, some were less enthusiastic than others about participation in the project. Some teachers who were part of the original application process 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 the principal or other teachers in the school. Further, the teachers' mathematics programs differed in nature and in extent of compatibility with the district curriculum framework and national reform effort. For example, many of the teachers placed more emphasis on computational skills and less on mathematical thinking and problem solving than is suggested in the district framework or NCTM *Standards*.

These differences created a dilemma for us in our staff development efforts. How could we accomplish our goal of helping teachers create performance assessments compatible with the district curriculum framework and national reform agenda, within the boundaries of our negotiated agreements, in cases where their instructional goals were incompatible with these guidelines? Our approach to resolving this dilemma is described in the following section.

Structure of the Project

The staff development program consisted primarily of a year-long series of weekly workshops conducted for the third-grade teachers at each school by a team of researchers from the University of Colorado's School of Education. The focus of weekly workshops alternated between mathematics and literacy. The mathematics educator (Brenda), one other principal investigator (Helene or Lois) and one graduate research assistant were present at each mathematics workshop. Researcher participation in literacy workshops was parallel, with the literacy educator (Francis) present at each session. (Names of schools and teachers are pseudonyms.)

Our initial intention in the weekly workshops was to facilitate changes in the teachers' assessment practices by helping them to think about their instructional goals and the relationships among goals, instruction, and assessment; to develop or select assessment tasks appropriate to their goals; and to create scoring criteria for the assessment tasks. We expected that each team of teachers (i.e., the group of third-grade teachers within a school) would design or select a shared set of assessments that reflected key goals of the school and district in mathematics and literacy, and that individual teachers would adapt assessments to their particular classroom contexts. As Lois explained to teachers at the initial fall meeting, "The project is focused on developing performance assessments, and the developers are the teachers. . . . What kinds of assessments do you need in classrooms, not just to tell outsiders what kids are learning, but to inform instructional decisions?"

As indicated above, we faced a dilemma when teachers' instructional goals were not well matched with the district curriculum framework. We struggled with this dilemma, to some extent, for much of the year. We attempted to resolve it by providing ideas for performance assessment tasks that addressed the teachers' instructional goals (many of which focused on computational skills) but also required problem solving, reasoning, and explaining. At the same time, many of the teachers requested materials and activities for teaching according to the district guidelines. In response to their requests, and because such materials could also be used for assessment purposes, we broadened the focus of workshops to include instruction as well as assessment. We hoped that as the teachers discussed these

tasks in the workshops and used them in their classrooms, they would begin to reconsider their instructional goals.

Workshop sessions took place immediately after school, at the school sites, and were approximately 1 1/2 to 2 hours in length. They addressed topics such as selecting, extending, and designing activities and materials for assessment and instruction; observing students and keeping records of observations; analyzing student work; and developing and using scoring rubrics. The structure of specific workshop sessions varied depending on the topic to be addressed. For example, Brenda often distributed written activities or manipulatives at a workshop and led a discussion on how they might be incorporated into the teachers' mathematics programs. We often asked the teachers to use these materials in their classes and to bring samples of student work to subsequent workshop sessions for examination and discussion. On several occasions, participants worked in small groups to examine student work and assign scores based on rubrics they had developed. Other workshop sessions consisted of more open-ended discussions of issues that arose as the teachers tried out performance assessments in their classrooms.

Data Sources

Data sources for this paper are the biweekly mathematics workshops and the interviews with teachers conducted during the 1992-93 school year, the first year of the project. All 15 workshops at each school were audio-taped. In addition, written notes were taken by the graduate research assistant to record nonverbal communication as well as descriptions of materials to which participants referred. When possible, artifacts (e.g., samples of assessment materials and student work) were collected. The audiotapes, written notes, and artifacts were used to prepare detailed field notes of the workshop sessions. Artifacts were filed with the field notes and served to corroborate reported changes in the teachers' practices.

All teachers participated in face-to-face interviews three times during the year. The first set of interviews was conducted before the weekly workshops began (September 1992); the second set at the beginning of the spring semester (January and February 1993); and the third set after completion of the workshops (May 1993). Semistructured questions focused on teachers' knowledge, beliefs, and reported practices related to mathematics instruction and assessment. Each

interview was conducted by a member of the research team, at the participant's school, during the school day. All interviews were audio-taped and transcribed.

Data Analysis

All workshop sessions were conducted separately for the three schools, with the exception of four meetings attended by teachers from all three schools. Further, the teachers' project-related interactions occurred almost exclusively with colleagues at their own schools. The CU team planned parallel workshop sessions for the three schools. However, the schools began the year in different places with respect to assessment goals and practices, and teacher change—both process and outcomes—evolved differently at the three sites. We therefore conducted our initial analyses for this paper on a school-by-school basis.

Initial data analysis consisted of coding the workshop field notes and interview transcripts for issues related to learning, teaching and assessment in mathematics. Codes included instructional goals, organization of instruction, instructional activities, assessment goals, assessment tasks, scoring, grading practices, teacher dilemmas, researcher dilemmas, advantages and disadvantages of performance assessments, what teachers hoped to learn from their participation in the project, and advantages and disadvantages of the project. At the same time, we created a sequential record of each workshop that summarized both the topics discussed and the flow of the discussions. Based on these initial analyses, we chose two topics to analyze in more depth: the scoring of open-ended mathematics tasks (scoring rubrics), and systems for recording information derived from observations of students during mathematics activities (keeping track). These assessment topics provide examples of different levels of success for the teachers in making changes, and thus provide important contrasts for looking at the change process. In addition, enough time was spent on each topic in workshops throughout the year that an in-depth analysis would likely yield information about the change process. Further, issues related to the development and use of scoring rubrics are among the most important considerations for teachers as they begin to implement performance assessments (Herman, Aschbacher, & Winters, 1992).

In the next phase of analysis we compiled the sections of coded workshop field notes and interview transcripts that addressed the topics of scoring rubrics and keeping track for each school, and we organized these data according to the

changes that occurred, the nature of the change processes, stumbling blocks, and missed opportunities. We then wrote summaries of the change process at each school, drawing upon the compiled data sets and focusing on changes that occurred in the teachers' ideas and practices related to scoring rubrics and keeping track. Next, we conducted a cross-site analysis, looking for commonalities and differences across the schools. Based on that analysis we identified five themes that describe key factors that facilitate and impede teacher change.

By deciding to focus on two topics, we were able to analyze in detail a subset of the large data set in order to identify and trace patterns in the change process. At the same time, this decision meant that our analysis did not capture many of the ways in which the teachers changed. Thus, while this article describes key factors that facilitated and impeded teacher change, it does not portray the scope of changes in ideas and practices made by individual teachers or school teams. For example, it does not address the teachers' increased use of problem-solving activities and emphasis on students' explanations, both of which were evident in interview responses and materials shared in workshop sessions (see Flexer et al., 1994).

Results and Discussion

Despite some teachers' initial ambivalence toward the project, all three teams participated in the mathematics workshops and, through their participation, made changes in their assessment practices. In the first section below, we summarize where the three groups of teachers began the project with respect to their assessment ideas and practices and then briefly describe the changes related to scoring rubrics and keeping track that occurred over the course of the 1992-1993 academic year. We explore these changes more closely in the second section, in the context of our themes regarding supports and impediments to teacher change.

Changes in Assessment Ideas and Practices

Spruce School teachers' instruction and assessment at the beginning of the year emphasized knowledge of facts and mastery of computational skills. They started the year without a clearly stated set of goals for working on new assessment strategies. There was little information in either the workshop or interview data to indicate their reasons for participation in the project or whether

their goals for the project, performance assessments, or mathematics instruction were similar to those held by the CU project team.

Walnut School teachers began their work with us with a desire to make changes in their assessment practices and a positive attitude toward participation in the project. These teachers viewed mathematics as including both skills (mathematics facts and computation) and understanding ("number sense," problem solving, and the ability "to explain and describe how they derive an answer"), and they attempted to assess both of these dimensions. They were using a variety of assessment strategies including worksheets, chapter tests, and observing and listening to students. Two of the five teachers were keeping notes on students during mathematics activities.

Unlike Walnut teachers, Pine School teachers expressed reservations about some of the district and NCTM reforms and questioned their adoption in third-grade classrooms. These teachers also viewed mathematics as including both skills and understanding, although their programs at the beginning of the year stressed the skills component. A common goal for the first grading period was that students "memoriz[e] their facts through 18." The teachers relied primarily on chapter tests and timed tests to assess students' mathematical skills. They began the project with an attitude that their existing mathematics instruction and assessment were fine. Elly, the team leader, questioned why they should put a lot of work into something that didn't need fixing and said she needed to be shown concrete evidence to convince her that alternative assessments are not too subjective.

Scoring rubrics. Several substantive concerns regarding scoring rubrics surfaced during workshop sessions at all three schools. The three most prominent are related to reasonableness of responses, expectations that students explain their answers, and consideration of students' writing ability.

From the initial mathematics workshop, teachers at Walnut wanted "reasonableness" to be a central goal of their mathematics program, and they decided to assign the highest score on problem-solving tasks to solutions with reasonable (even if not fully correct) responses and elaborate explanations. Teachers at Spruce and Pine were more concerned that students solve problems for the correct response. Their scoring criteria required an "exactly correct" answer in order for students to receive the highest score. By the end of the year,

however, many Pine and Spruce teachers accepted the premise that a reasonable answer indicates some understanding, and they made it possible for students to earn a high score with an elaborate explanation and an "almost correct" response.

By February, all of the teachers were incorporating written explanations into their mathematics tasks for both instructional and assessment purposes. Once they adopted this practice, the question of how to score these explanations became a central topic in a number of workshop sessions. One important issue for all three schools was whether their scoring systems should take students' writing ability into account. In general, the teachers felt that having scores be too dependent on written explanations was unfair to students who understood the mathematics but were either unwilling or unable to write a coherent explanation.

Substantive concerns about scoring rubrics often played themselves out as discussions of structural issues. One major focus of these discussions was the number of points to include in a rubric. The teachers were concerned about the trade-off between the information they might receive from an expanded point scale versus the relative ease of scoring with a compressed metric. By the end of the year all teachers were using a general problem-solving rubric that contained either 3 or 4 performance levels.

The other structural issue with which teachers struggled was whether to use a one-dimensional or two-dimensional scoring system. Many teachers seemed to realize that they received more information from two dimensions than one. On the other hand, they were concerned that scoring on two dimensions might be more difficult and time-consuming. The question of whether to use one or two scoring dimensions emerged immediately at Walnut, and it was tied directly to the teachers' commitment to focus on both reasonableness and explanation of responses. The Walnut teachers initially decided to use a single scoring dimension incorporating both reasonableness and explanation, and most finished the year with a similar system. Most teachers at Pine also decided to use a one-dimensional scoring system. In contrast, teachers at Spruce decided upon a two-dimensional system, assigning separate scores for correctness of the answer and quality of the explanation.

Keeping track. All of the teachers reported an increased use of problem-solving activities in their mathematics programs, and materials they shared with us in workshop sessions corroborated this change. We encouraged them to keep

systematic records as they observed students during these activities because we thought that such records would provide additional insights about student understanding. In general, however, we saw less change in the teachers' use of record-keeping systems than we saw in their implementation of rubrics for scoring open-ended problems. There was also much more variability, both across and within schools. Most teachers at Pine did not implement observational record keeping at all; Spruce teachers developed many systems, yet never really seemed to be convinced of the importance of doing so; whereas Walnut teachers appeared to understand that observational records could help them learn different things about their students' mathematical understanding and problem-solving strategies than they learned through other assessment tools.

Two teachers at Walnut were using observational records in their mathematics assessment systems at the beginning of the year. In contrast, other Walnut teachers started the year by questioning whether observing students engaged in mathematical activities "counts" as an assessment task. Abby asked, ". . . if we wanted to pull these students and have them play Place the Digit, and we could watch them and talk to them, then that's an acceptable task?" They appeared to be questioning the validity of such tasks as assessments, rather than their value as instructional activities or tools for providing insights about student understanding. The CU researchers assured them that such events are acceptable assessment tasks, particularly when accompanied by systematic observational records. Soon thereafter, these teachers began experimenting with various record-keeping strategies. The experimentation seemed to increase their understanding of the value of observational data for learning about students' understanding of mathematics.

Three of the Walnut teachers struggled particularly hard with the question of how to make their record-keeping systems less time-consuming. This struggle led them toward systems that combined the efficiency of checklists with the informativeness of anecdotal records. For example, Janet devised a key that provided quick information about the meaning of checks, pluses, and minuses. This approach enabled her to depend more on these symbols and less on written comments about each child.

By midyear, all of the teachers at Spruce and Walnut and one at Pine were experimenting with various record-keeping systems. And their systems continued to evolve as the school year progressed. By the end of the year, some teachers had

systems in place that represented changes from the previous year. In addition to changes in the record keeping itself, a few teachers noted substantive changes in their approaches to learning about students' mathematical understanding. For example, they reported that as they became more interested in knowing how students think (in addition to whether they can get the right answer), they began to ask different types of questions—questions that encouraged students to explore and articulate alternative problem-solving strategies rather than directing them toward finding the correct answer.

Supports and Impediments to Teacher Change

As we analyzed workshop and interview data we noticed a number of factors that facilitated or supported teachers' attempts to change their instruction and assessment practices and a number of factors that interfered with or hindered change. We first viewed these data through the lens of our staff development framework. As the analyses progressed, our knowledge of the teacher change literature—particularly ideas from cognitive psychology, such as the situated nature of cognition and the dual roles of knowledge and beliefs in learning (Borko & Putnam, 1996)—contributed to a growing understanding of the patterns we observed. This understanding helped us to modify the conceptual framework for our analyses and, eventually, to define the themes that became the organizational structure for our presentation of results. Thus, the conceptual framework, data analysis, and writing developed recursively, as we moved from the literature to the data and from the data to the literature. In this section of the paper we discuss our findings about supports and impediments to teacher change, organized around five themes. After presenting each theme, we offer examples to illustrate the factors that influenced teacher change, and we relate these findings to relevant research and theory.

Theme 1: Situating the change process in the actual teaching and learning contexts where the new ideas will be implemented is an effective strategy for helping teachers change their practice.

The CU Assessment Project was able to situate the change process in actual teaching and learning contexts in a number of ways. Perhaps the most direct examples occurred when teachers attempted to implement ideas presented or developed in a workshop in their classrooms. For example, Lena made an

important discovery when she tried out a scoring rubric developed as a result of the project's influence.

> Well, one thing I've noticed is, I've been doing this [problem solving] since day 1. I didn't start scoring until 3 1/2 to 4 weeks ago [a change she attributed to the mathematics workshops]. . . . It's amazing how much better they're getting now that there's a score.

Once Lena discovered the benefits of using a rubric to assign scores to problem-solving papers, she continued to score these papers on a regular basis throughout the year.

One particularly effective version of situated learning occurred when teachers discussed ideas in a workshop session, attempted to implement the ideas in their classrooms, and then reflected or otherwise built upon their experiences in subsequent workshop sessions. For example, one of the problems that Brenda brought to the schools, the Cat Problem (developed in British Columbia for the 1990 Provincial Mathematics Assessment), was the focus of a series of workshops at Spruce. The problem asks students to compute the number of unit squares required to build a pen for 10 cats. Teachers developed rubrics to score the problem at one workshop, deliberating about such topics as the importance of the "right" answer and features of student responses to take into account in assigning scores. They decided to try out the problem and their rubrics in their classrooms during the next two weeks. At the subsequent workshop they reviewed both student responses and their scoring of these responses. As a result of these experiences, several of the teachers spoke about scoring performance assessments with greater understanding than they had previously. For example, after the second workshop they realized the instructional benefits of having the math problems scored along two dimensions and designed a way to have students participate in this process. This sequence—a "hands-on" workshop, related classroom activities, and a follow-up discussion—produced some of the biggest changes we observed in the ways teachers talked about assessment.

Most teachers on the project found it difficult to keep written records of students' performance during independent mathematics activities. To help them develop a workable record-keeping system, Paula (a visiting researcher who arrived midyear and worked with the CU team) offered to come into teachers' classrooms, work with children during independent activities, take notes on her

observations of the children, and then meet with the teachers to share both her insights about the children's mathematical understanding and her ideas about record keeping. The teachers at Walnut accepted Paula's offer. Three of the teachers commented specifically about the value of Paula's input in helping them develop record-keeping systems. They noted that her observations of students and sharing of insights gave them a better sense of what to look for when observing students, what questions to ask, and what information to record. They were also more motivated to keep observational records during the six-week period when Paula visited on a weekly basis. Their records, combined with her notes on the day's mathematics activities, became springboards for discussions of what it means to know a mathematical concept such as multiplication, what teaching strategies might foster student understanding, and what evidence could be used to support a claim of understanding. Often, these discussions ended with Paula and the teachers attempting to design a record-keeping system to provide this kind of evidence. The teachers then tried the system out during subsequent mathematics lessons and brought their records to the next weekly meeting. Abby attributed the improvement in her anecdotal notes, in part, to this work with Paula.

In contrast, Pine and Spruce teachers' experiences with written records of students' performances provide an example of what might happen when teachers' efforts to change are not situated in their actual classroom practice. These teachers declined Paula's offer to spend time in their classrooms. Although teachers at both schools were encouraged to try out various systems for keeping records of students' classroom performances, their anecdotal records were not explicitly shared, discussed or challenged in workshops in the ways that their scoring rubrics were. The majority of time spent discussing record keeping was devoted to systems for recording observations, while relatively little attention was given to helping the teachers figure out what to observe, what questions to ask, or how to make sense of the information. Thus, without classroom-based support, Pine and Spruce teachers missed out on an opportunity to connect their efforts to develop record-keeping systems with their ongoing classroom experiences. This lack of connection may be one reason that there was noticeably less change in their assessment ideas and reported practices related to keeping track, compared to scoring rubrics.

Many current cognitive theorists emphasize the situated nature of cognition (Bredo, 1994; Brown, Collins, & Duguid, 1989; Bruner, 1990; Newman, Griffin, &

Cole, 1989). Rather than considering knowledge to be abstract and detached from the external world, they emphasize the relationship between knowledge as it exists in the mind of the individual and the situations in which that knowledge is acquired and used. They argue, for example, that knowledge is inseparable from the contexts and activities in which it is developed. Similarly, what is learned cannot be separated from how it is learned and used. This view of cognition as situated suggests that if teachers are to learn to teach in new ways, the knowledge they acquire must be grounded in the classroom contexts in which it will be used (Borko & Putnam, 1996). It also offers an explanation for our finding that professional development experiences that provided opportunities for teachers to explore new instructional strategies and ideas in the context of their own classroom practice were among the most effective factors for promoting and supporting teacher change.

Theme 2: Group discussions of instructional and assessment issues—for example, within the context of workshop sessions—can be an effective tool for the social construction of new ideas and practices.

Walnut teachers' discussion of the Cat Problem illustrates how ideas sometimes changed in the course of group conversations. Brenda brought several student responses to the problem that were used at teacher development sessions in Oregon to the Walnut workshop. One response that stimulated an extensive conversation featured a fence drawn around a 2 by 5 array of cats. The teachers initially agreed that they would give this solution a 2 on a 4-point scale. As Jackie explained, "To me [student paper] D is a 2 because they did not pick up on the pattern of the single file." Abby agreed, noting that because the examples provided in the problem showed fences drawn around one cat, a 1 by 2 array of cats, and a 1 by 3 array, she would expect students to extend the pattern and draw a 1 by 10 array: "We would expect it because that's what we teach kids—is to look at the patterns." Teachers and CU researchers studied the directions for the problem—"Show how you find out how many squares are needed to go around 10 cats. Explain your answer and your thinking."—and they examined other sample solutions. After much deliberation, they decided that the solution might be worth more than 2 points. Judy concluded,

> There appear to be two ways to do this. . . . If you continue the cats out this way, you see the pattern that each time you add a cat . . . and two squares. However, if . . . it

> has to be just a rectangle you can make your rectangle either long and thin or you can make it like this [2 by 5]. In which case this is perfectly acceptable and it takes fewer squares.

At Pine, conversations about assessment were more likely to occur outside the workshop setting—for example, when the teachers ate together during their common lunch period. Products they developed and decisions they made provide evidence that the teachers worked on project-related issues when they met outside the workshop setting. Between the 9/22/92 and 10/6/92 workshops, for example, Pine teachers created a continuum for place value that included instructional activities and assessments for each level of the continuum. They shared the continuum with the project team at the 10/6 workshop.

Social constructivism helps us to understand how ideas were developed in these group discussions. Social constructivists accept the premise that knowledge is a social product; knowledge creation is a shared experience (Prawat & Floden, 1994). Group settings such as staff development programs function as learning communities in which individual and social processes of knowledge construction occur concurrently and interactively (Cobb et al., 1992; Simon & Schifter, 1991). Individuals actively construct their own understandings based, in part, on their existing knowledge and beliefs. At the same time, as members of a learning community, they work together to develop shared meanings through questioning, clarification, negotiation, and consensus building. Throughout these processes, the constructions of individuals and of the group mutually contribute to and constrain one another, resulting in the emergence of new ideas and understandings.

Theme 3: Staff development personnel and other persons with specific expertise can facilitate change by introducing new ideas based on teachers' current levels of interest, understanding, and skill.

Workshop discussions were sometimes especially productive when members of the CU team asked questions to keep the conversation focused or made suggestions to enable the teachers to gain new insights. For example, at Spruce, Helene helped Tamara think about the issue of how many dimensions to include in a scoring system:

Tamara: I agree that . . . it's difficult, more difficult when you have two different places but I also agree with [Brenda] saying that we do need to look at the two different things [answer and explanation].

Helene: So, part of it, let me see if this is, if this fits what you're saying, [Tamara] . . . I think it's probably easier to come up with a global score that we could all agree on. . . . But it might be that if what you want to emphasize is, there's getting the answer and there's explaining the answer, then we ought to give two scores—not because that makes the scoring better but because that highlights what we're trying to get at.

Tamara: Yeah, and neither one, and both of them are very important. Because for so long that was what was going on . . . kids could come up with the answers, but they couldn't ever, they really didn't understand the process they, it took them to get there.

Later in the discussion, Helene checked with Tamara to see what she was thinking:

Helene: So now, [Tamara], . . . given that you've talked about, you know, why two scores is better and why one score is better, if you had to make a decision, which would you do?

Tamara: I would probably at this point I'd [give] two: for the answer and for the explanation.

In a conversation at Walnut, Helene and Brenda attempted to encourage the teachers to keep written records—even as simple as checks, pluses, and minuses—as they walked around the room observing students engaged in mathematics activities. Early in the conversation Janet shared her record book with information about students' performance on mathematics skills. She apologized that almost all of the entries were scores on worksheets rather than information recorded as she observed and assisted students during independent work. Helene built on this apology:

Helene: [O]ne of the next steps that you could think about doing, is if you're introducing subtraction with Base-10 Blocks and you have kids working at their desks or working in small groups, whichever, solving problems with the Base-10 Blocks, you could have a column in here for that activity. You could walk around the room and do your checks like that.

Janet: Now, you're right, I don't have that information for addition.

Helene: Is that something you could see yourself starting to do?

Janet: Yes, and that's why . . . there are very [few] grades and it's because we weren't doing things that required a grade.

Helene: Right and I'm suggesting that if you end up with something that looks like this for subtraction . . . or not instead of but in addition to [grades], that could be great. That would be a nice, sort of the next step on how to incorporate some of the ongoing tasks that you have them do with manipulatives that aren't so conducive to a work sheet that you can score.

Janet: Yeah. Because I think that that is the downfall right now because this does not show any of those activities. We did them and I have some sheets but most of it was, you just do it.

Richardson (1992) analyzed the ways in which alternative classroom practices and their theoretical bases were introduced during staff development sessions in the Reading Improvement Study. She concluded that a formal presentation about a practice that grew naturally out of a conversation met the project goal of introducing new ideas in ways that stimulated interest and discussion better than did a lecture planned in advance. The presentations in the above two conversations provide additional support for her conclusions.

The effectiveness of these presentations as strategies for helping teachers to consider new ideas and practices can also be understood using the construct of *instructional scaffolding*. Bruner, Wood, and Ross introduced the metaphor of a scaffold to characterize a pedagogical strategy whereby teachers (in this case, staff developers) assist students' (teachers') learning by providing temporary, adjustable support that can be removed from the situation when it is no longer needed (Bruner, 1978; Wood, Bruner, & Ross, 1976). In scaffolded instruction, the nature and amount of assistance is continually adjusted to take into account the difficulty of the task and the learner's current level of understanding and skill. In the above examples, as in other workshop instances, the CU team was able to introduce ideas into conversations in ways that provided support for the teachers' exploration of new understandings and skills.

Theme 4: When teachers' beliefs are incompatible with the intentions of the staff development team and are not challenged, the teachers

are likely to either ignore new ideas or inappropriately assimilate them into their existing practices.

In some cases, teachers' beliefs about mathematics instruction and assessment that were incompatible with the CU team's intentions for performance assessment were transformed as a result of the work we did together over the year. Lena's beliefs about the value of scoring rubrics and her associated decision to score student work on a regular basis (discussed under Theme 1) provide one example. In general, however, because beliefs were not explicitly questioned or challenged by members of the CU team, they remained unchanged throughout the year. Because beliefs serve as filters through which new ideas are perceived and interpreted, it is not surprising that the teachers adopted practices that were also incompatible, to some extent, with our intentions.

Scoring rubrics adopted by two of the Spruce teachers provide one example of practices that were incompatible with the CU team's intentions. In one class at Spruce, students developed a rubric for scoring mathematical explanations that included grammar, spelling, and punctuation as criteria. Because their teacher wanted to acknowledge the students' work, and because she too thought these criteria important, she and one other teacher incorporated the criteria in the rubrics adopted for classroom use. Thus, to earn the highest score of 4, an explanation in Libby's class had to include "correct spelling, have a picture, have a right answer, neat handwriting and periods and capitals." Although they conformed to the format for performance assessments, these rubrics did not fit with the CU team members' beliefs about criteria for scoring mathematical explanations or the view of the wider mathematics community. When Libby first presented this rubric at a workshop on March 23, Helene and Brenda asked for clarification about the criteria but did not challenge them. The topic was addressed again on April 20, and Libby offered the following explanation for her students' decision to include these construct-irrelevant criteria:

> We talked about that [spelling, punctuation and handwriting] so much with the summaries [used to assess reading comprehension] . . . when they're writing for someone else, the importance of how it should look different than when they're writing for themselves, that it just carried over into what they thought a 4 should be.

Lois raised a concern about what she termed "the fairness issue," an issue that was addressed by several parents during interviews conducted as part of the project. She explained,

> When we move into writing explanations . . . we want kids to be able to explain and to think. But I would also not want to push it so far that, as those parents expressed it, a kid who just had trouble verbally but was good in math could never get a top score.

Lois suggested a two-step scoring process that included an opportunity for revision, as one way to address the fairness of math scores while retaining the spelling and grammar criteria:

> So you might just try to think of . . . ways that we could acknowledge that their thinking was really good and then have them [revise the explanations]. If they revised it, then I think there's no problem.

To the best of our knowledge, based on rubrics the teachers brought to workshop sessions as well as those on display in their classrooms (where the workshops were conducted), the teachers kept their scoring rubric as is and did not incorporate Lois's suggestion.

Pine teachers also held a number of views that differed from those of the CU team. Unlike Spruce teachers, they sometimes questioned tasks that they were asked to do or argued with us about their differing ideas. The following exchange about the question of what students are capable of understanding occurred in the midst of a discussion about the kinds of assessments to use to determine student understanding.

Elly: But do you think, [Brenda,] that all students will totally understand?

Brenda: Well, that's what you're pushing for, that's what you're trying to get at.

Elly: But don't you think that some children will only learn it by rote?

Brenda: No.

Elly: You don't think so.

Brenda: Well, I think we could teach them by rote. We could make them respond correctly to that. And my position is, I don't want them to be able to

> respond correctly to that if they don't understand it. I'd rather that they fail that entirely, if they don't really understand it.

Elly brought up the same topic later in the workshop:

> Elly: See, but then that's I guess, [Brenda,] where I question, 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.

Brenda again responded that we have to try to teach for understanding, because it is not worthwhile for students to learn math by rote. As this example illustrates, Pine teachers were willing to argue for their position and were not easily convinced to change.

Cognitive psychologists' assumptions about the dual roles that knowledge and beliefs play in thinking, acting, and learning help us to understand some of the difficulties that the CU team encountered when our beliefs and the teachers' were incompatible. Teachers' knowledge and beliefs about teaching, learning, and subject matter are critical determinants of whether and how they implement new educational ideas. For teachers to change their pedagogical practices, they must have the knowledge necessary to implement the changes and the beliefs to support them. Thus, efforts to help teachers make significant changes in their teaching practices must also help them to acquire new knowledge and beliefs. At the same time, teachers come to understand new practices through their existing knowledge and beliefs. This dual role of knowledge and beliefs—as both targets of change and filters through which change takes place—can make fundamental changes in teaching practices difficult to achieve (Cohen & Ball, 1990; Putnam, Heaton, Prawat, & Remillard, 1992). For example, when existing beliefs are incompatible with desired practices, it may be necessary to explicitly challenge these beliefs in order for changes in practice to occur.

When the CU team addressed the issues raised by the Pine teachers, and when we commented on practices adopted by the Spruce teachers, we typically acknowledged how our views and recommendations for practice differed from theirs. We did not, however, work with the teachers to examine the theoretical,

empirical or ethical premises underlying their practices (i.e., their practical arguments; Fenstermacher & Richardson, 1993). For example, we did not challenge Elly to consider premises associated with her belief that some children are not capable of learning math with understanding.

A tension developed between commitments to the teachers' goals on the one hand and the reform agenda of the mathematics education community on the other, when we realized that some of the teachers' beliefs and practices were inconsistent with our vision of performance assessment, as indicated earlier. In our efforts to resolve that tension, we were hesitant to challenge the teachers' beliefs because of our initial agreements about the nature and scope of the project and because several teachers were already feeling overwhelmed by project expectations. As a result, the CU team may have been unable to capitalize on a number of opportunities to help the teachers think in new ways about mathematics assessment and instruction.

Theme 5: Time is a major obstacle to changing classroom practice. Competition among priorities for limited classroom time is particularly troublesome.

Time was the factor that the teachers mentioned most frequently as a drawback of the new assessment practices they were attempting to implement. The concerns of teachers at Pine and Spruce focused primarily on the time outside of the classroom that these new assessment practices require. Teachers at Spruce felt that the amount of preparation between workshops expected of them by CU researchers was a burden that interfered with other professional responsibilities such as instructional planning. For teachers at Pine, the time required by activities such as scoring was an obstacle to change. As Lena explained, although performance assessments may enable a teacher to "discover who's a better thinker than somebody else," they are also "very hard to score . . . [and] then you're spending a lot of time trying to figure out" the scores for students' solutions. Despite this concern, because of the value she saw in scoring students' solutions to problems, Lena (unlike some of the teachers) scored them on a regular basis beginning shortly after the winter break.

Teachers at Walnut and Spruce also identified as a problem the time that various assessment activities took away from classroom instruction. At Walnut, in fact, this aspect of time limitations was the only one addressed. Teachers at all

three schools experienced frustration as they attempted to keep records while students were engaged in individual or small-group mathematics activities. The procedures they developed took more time away from their instructional interactions with students than they were comfortable spending. When asked in the winter interview whether she recorded observations as she walked around the classroom, Jackie explained:

> More like a check sheet. Are they displaying an understanding or not? It would be ideal to be able to keep more of an anecdotal record of "this is exactly what I saw and this is what I've learned from it." Unfortunately because of time limitations it's really hard for me to do that.

By the time of her spring interview, Jackie was using check sheets almost exclusively but, in her eyes, the time dilemma had not been satisfactorily resolved:

> So, the check sheets just on my observations. Unfortunately, when it is maybe a class discussion . . . it is so difficult to write down when you are noting that this student has got it or that this student does not. So, that is still a problem area for me . . . to take every moment that you are teaching and really absorbing what the kids know and to make it concrete because, of course, it does tend to fade after a while [if] you don't get it on paper.

Teachers at Spruce expressed similar concerns. As Penny stated in her winter interview:

> It's a little hard to find the time in the day to sit down and make those anecdotal records before it's gone out of your mind. . . . I think the hardest thing we have to do, I have found, is finding the time to do it. Because with third grade . . . you have to be with the children a lot.

Three of the teachers at Walnut also confronted time dilemmas related to students' explanations of problem-solving strategies. Like teachers at Pine and Spruce, they frequently asked students to provide written explanations for their solutions to problems. They were concerned, however, that some students' weak writing ability interfered with the teacher's ability to understand their explanations. These teachers agreed that one solution to the problem was to talk with students when they could not understand their written explanations. Janet recounted an incident with one student that illustrated the importance of such verbal interactions:

> . . . a problem last week, there were 17 legs and a furniture maker and he made some . . . three-legged stools and some four-legged tables, had five pieces of furniture, so figure out [how many stools and how many tables]. And at first with [Vince] I was just totally confused by his picture. . . . [Vince] has a lot of difficulty with written expression and I could tell nothing from what he had written. So I pulled him out today, I said, "Come and explain this to me," and he said, "Well I put out five pieces of furniture and I knew that they at least had three and so I put three on each one and I needed some more legs so I added more legs." . . . After he told me the words, I made sense of his picture.

Although the teachers acknowledged the value of having students explain their solutions orally, they were concerned about the amount of time such explanations demanded. They resolved the dilemma of balancing quality information with time limitations as best they could, by talking with students only for selected assignments, and only in instances where their written explanations were problematic.

These teachers seemed to be genuinely searching for resolutions to their dilemmas. For some issues, however, neither they nor the CU team were able to come up with a satisfactory plan, and the dilemmas remained unresolved by the end of the project year. Possible approaches for managing such dilemmas are explored in the concluding section of this paper.

Conclusions and Implications

Teachers at all three schools made several significant changes in their ideas about and practices of mathematics assessment and instruction as a result of participation in the CU Assessment Project. For example, they incorporated problem-solving activities and student explanations as more central components of their mathematics programs, and they developed scoring rubrics for assessing children's solutions and explanations for open-ended mathematics tasks. These changes are compatible with the mathematics education community's commitment to assessments that yield information about components of students' mathematical power such as problem solving, mathematical reasoning, and the ability to communicate mathematically. Thus, through their participation in the project, the teachers developed a set of ideas and practices that are connected to the wider mathematics education community's reform agenda.

There were also some ways in which the teachers did not shift their ideas or practices, or made changes that were not congruent with the wider community's reform agenda. With respect to the two topics that were the focus of analysis for this paper, for example, most teachers at Pine and Spruce remained committed to the idea of a single correct answer to mathematics problem-solving activities. And only a few teachers developed systems for keeping observational records that they were comfortable using on a regular basis. As we argued in the Results and Discussion section, both sets of patterns can be understood through the lens offered by cognitive theory—specifically, through the constructs of social construction, situated cognition, instructional scaffolding, and the dual role of knowledge and beliefs.

As we step back and take stock of our efforts, and as we look ahead toward future staff development projects, it seems reasonable to ask what it would have taken for the teachers to make more fundamental changes in their beliefs and practices regarding mathematics assessment and instruction. For example, how might the CU team have helped Pine teachers to reassess their belief that some children can only learn mathematics by rote? How might we have helped teachers at Pine and Spruce to embrace more fully the value of multiple solutions and the acceptability of reasonable responses with well-articulated explanations?[3]

If we were to embark on another staff development effort, we would build in explicit attention to beliefs as well as practices. For example, we might draw upon ideas about the elicitation and reconstruction of teachers' practical arguments developed by Fenstermacher (1986; see also Fenstermacher & Richardson, 1993). This approach to helping teachers change their beliefs and practices entails working with them to first describe their actions and provide an explanation or rationale (elicitation), and to then appraise and reconstruct the premises on which their rationale is based (reconstruction). Richardson and Anders (1994; Anders & Richardson, 1992) found that a focus on practical arguments helped teachers in the Reading Improvement Study to change their theories of reading, learning to

[3] The CU Assessment Project continued into the 1993-94 school year on a smaller scale. During that year members of the research/staff development team continued to work with teachers at two of the schools, primarily by observing in their classrooms every four to six weeks and meeting with them to discuss our observations. We incorporated some of the suggestions in this article into our work that year, such as situating staff development efforts more closely in the teachers' practice. We are currently analyzing data from that year. Preliminary results reveal additional changes in the teachers' beliefs and practices, in directions compatible with the CU team's reform agenda.

read, and the teaching of reading; their beliefs about reading comprehension; and their reading instructional practices.

We would also attend more systematically to situating staff development efforts in teachers' practice. Participation in the teachers' classrooms was not part of the original project design, primarily because of limited resources and the fact that changes in instruction were not an explicit project goal. However, our analyses revealed substantial benefits gained through Paula's visits to Walnut teachers' classrooms and through other ways in which we were able to make connections with teachers' classroom practice (for example, by having them bring student work to the workshop sessions). These patterns suggest that building a classroom-based component into staff development efforts may be a valuable use of project resources. This suggestion is supported by other empirical studies of staff development. For example, based on findings from their project to help teachers learn to conduct instructional conversations with their pupils, Goldenberg and Gallimore (1991) suggested that staff development "must be grounded in the mundane but very real details of teachers' daily work lives" (p. 69). When classroom visits are not feasible, they recommended videotaping teachers' lessons and using the tapes as the basis for providing feedback to the teachers (see also Saunders, Goldenberg, & Hamann, 1992). Similarly, Fennema, Carpenter, Franke, and Carey (1992) and Schifter and Simon (1992) found that classroom-based support and supervision provided by members of their research/staff development teams helped teachers as they attempted to incorporate new ideas developed during workshops into their mathematics programs.

Our third recommendation builds directly on successful aspects of the CU Assessment Project. Given the importance of peer support and collaboration, staff development efforts are more likely to be successful if the staff development team works with groups of teachers, preferably at the same school. Research on educational reform conducted within a sociological or anthropological tradition provides additional evidence for the importance of school-based support for teacher learning (McLaughlin & Talbert, 1993; Sarason, 1990). McLaughlin and Talbert (1993) suggested that teachers' ability to successfully adapt their practices to meet the national reform agenda depends upon "participation in a professional community that discusses new teaching materials and strategies and that supports the risk taking and struggle entailed in transforming practice"

(p. 15). Professional communities "offer the most effective unit of intervention and powerful opportunity for reform. . . . The path to change in the classroom core lies within and through teachers' professional communities: learning communities which generate knowledge, craft new norms of practice, and sustain participants in their efforts to reflect, examine, experiment, and change" (p. 18). Our data support these conclusions. Teachers at all three schools changed their assessment ideas and practices, at least to some extent. The degree of change was related to the extent to which the teachers worked together as a learning community. That is, we saw the greatest amount of change at Walnut and the least at Spruce.

Our final recommendations address issues related to time. First, staff development programs must allow sufficient time for teachers to make substantial changes in their practice. Goldenberg and Gallimore (1991) make the point eloquently: "Teachers need one to two years of intensive work to achieve a reasonable mastery of IC (instructional conversations) skills. . . . This insight from our research helps explain why so many past efforts to change teaching practices have not succeeded. Many of the needed changes in teaching are hard to achieve, and there is no tradition in U.S. education for training of the length and intensity required to achieve substantive change" (p. 72). Our experience with the CU Assessment Project, like theirs with IC, suggests that staff development programs be at least one year in length and provide release time for teachers so that they can participate with the intensity needed to make more than superficial changes in their beliefs and practices.

Second, change efforts must take into account the almost universal feeling that classroom time is insufficient for teachers to accomplish all that they wish with their students. In our study, this feeling surfaced most strongly as teachers found that various assessment activities took more time away from instructional interactions with children than they were willing to give. Some of these feelings diminished as teachers realized that they could use the same kinds of activities for assessment and instruction purposes, and as they became more comfortable with the new kinds of activities and scoring systems. In an attempt to address remaining concerns (regarding, for example, teachers' use of anecdotal records and students' oral explanations of problem solutions), we (the authors of this article) found ourselves rethinking our priorities for classroom assessment. For example, although we continue to believe in the value of observational records of students'

participation in mathematics activities, we now think that it may not be feasible for teachers, particularly those in the primary grades, to include systematic anecdotal records in their initial efforts to implement performance assessments. Primary grade children need teacher attention, especially when they are first learning to engage in cognitively complex tasks such as problem solving and mathematical reasoning. As students become more independent learners, and as teachers become more proficient with performance assessment, such record keeping will become more feasible. In the interim, however, we recommend that teachers set aside time at the end of the school day to jot down notes about several students whose participation in mathematics activities stood out for one reason or another.

In conclusion, there are several lessons from our study with implications for current educational policy debates about the effects of a school-based intervention on teachers' beliefs and practices. If new reforms are to succeed, researchers and staff development personnel should see themselves as facilitators and guides, helping teachers to ground new ideas in their existing practice. Even with such support, however, major shifts in assessment and instructional practices are not accomplished overnight. Teachers need time and opportunity during the school day for meeting with experts, conferring with colleagues, experimenting with new techniques, and reflecting on their practice. While we support many of the goals of new reform proposals, we are fearful that simply expecting teachers to teach to "cognitively-complex authentic assessments" without proper support will not lead to desired student performances. Our project's focus on performance assessment is unique within the staff development literature. Our findings and conclusions, however, add to the growing understanding highlighted by Bradley (1994): "Good professional development, researchers have learned, brings teachers together in networks that wrestle, over time, with important issues. Teachers should also receive coaching and follow-up help in using new practices in the classroom" (p. 20).

References

Anders, P., & Richardson, V. (1992). Teacher as game-show host, bookkeeper, or judge? Challenges, contradictions, and consequences of accountability. *Teachers College Record*, *94*, 382-396.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Urbana, IL: Center for the Study of Reading; National Academy of Education, Washington, DC.

Black, P. J. (1994). Performance assessment and accountability: The experience in England and Wales. *Educational Evaluation and Policy Analysis*, *16*, 191-203.

Borko, H., Davinroy, K. H., Flory, M. D., & Hiebert, E. H. (1994). Teachers' knowledge and beliefs about summary as a component of reading. In R. Garner & P. A. Alexander (Eds.), *Beliefs about text and instruction with text* (pp. 155-182). Hillsdale, NJ: Lawrence Erlbaum Associates.

Borko, H., & Putnam, R. (1996). Learning to teach. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 673-708). New York: Simon & Schuster Macmillan.

Bradley, A. (1994). Teacher training a key focus for Administration. *Education Week*, *13*(39), 3, 20.

Bredo, E. (1994). Restructuring educational psychology: Situated cognition and Deweyian pragmatism. *Educational Psychologist*, *29*, 23-35.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32-42.

Bruner, J. S. (1978). The role of dialogue in language acquisition. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language* (pp. 241-256). New York: Springer-Verlag.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Carpenter, T. P., & Fennema, E. (1992). Cognitively Guided Instruction: Building on the knowledge of students and teachers. In W. Secada (Ed.), Curriculum reform: The case of mathematics in the United States [Special issue]. *International Journal of Educational Research, 17*, 457-470.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, *23*(7), 13-20.

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education*, *23*, 2-33.

Cohen, D. K., & Ball, D. L. (1990). Relations between policy and practice: A commentary. *Educational Evaluation and Policy Analysis*, *12*, 249-256.

Fennema, E., Carpenter, T. P., Franke, M. L., & Carey, D. (1992). Learning to use children's mathematics thinking: A case study. In R. Davis & C. Maher (Eds.), *Schools, mathematics, and the world of reality* (pp. 93-117). Needham Heights, MA: Bacon.

Fenstermacher, G. D. (1986). Philosophy of research on teaching: Three aspects. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 37-49). New York: Macmillan.

Fenstermacher, G. D., & Richardson, V. (1993). The elicitation and reconstruction of practical arguments in teaching. *Journal of Curriculum Studies*, *25*, 101-114.

Flexer, R. J., Cumbo, K., Borko, H., Mayfield, V., & Marion, S. F. (1994, April). *How "messing about" with performance assessment in mathematics affects what happens in classrooms*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Goldenberg, C., & Gallimore, R. (1991, November). Changing teaching takes more than a one-shot workshop. *Educational Leadership*, *49*(3), 69-72.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

International Reading Association/National Council of Teachers of English. (1994). *Standards for the assessment of reading and writing*. Newark, DE: Author.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, *20*, 15-21.

Madaus, G. F., & Kellaghan, T. (1993). The British experience with "authentic" testing. *Phi Delta Kappan*, *74*, 458-469.

McLaughlin, M., & Talbert, J. E. (1993). *Contexts that matter for teaching and learning: Strategic opportunities for meeting the nation's educational goals*. Stanford, CA: Stanford University, Center for Research on the Context of Secondary School Teaching.

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. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: Working for cognitive change in schools*. Cambridge, England: Cambridge University Press.

Prawat, R. S., & Floden, R. E. (1994). Philosophical perspectives on constructivist views of learning. *Educational Psychologist*, *29*(1), 37-48.

Putnam, R. T., Heaton, R. M., Prawat, R. S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. *Elementary School Journal*, *93*, 213-228.

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.

Richardson, V. (1992). The agenda-setting dilemma in a constructivist staff development process. *Teaching and Teacher Education*, *8*, 287-300.

Richardson, V. (Ed.). (1994). *Teacher change and the staff development process: A case of reading instruction*. New York: Teachers College Press.

Richardson, V., & Anders, P. (1994). The study of teacher change. In V. Richardson (Ed.), *Teacher change and the staff development process: A case of reading instruction* (pp. 159-180). New York: Teachers College Press.

Sarason, S. (1990). *The predictable failure of educational reform: Can we change course before it's too late?* San Francisco: Jossey-Bass.

Saunders, W., Goldenberg, C., & Hamann, J. (1992). Instructional conversations beget instructional conversations. *Teaching and Teacher Education*, *8*, 199-218.

Schifter, D., & Simon, M. A. (1992). Assessing teachers' development of a constructivist view of mathematics learning. *Teaching and Teacher Education*, *8*, 187-197.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher*, *21*(4), 22-27.

Shepard, L. A., Flexer, R. J., Hiebert, E. H., Marion, S. F., Mayfield, V., & Weston, T. J. (1994, April). *Effects of introducing classroom performance assessments in student learning*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Simon, M. A., & Schifter, D. (1991). Towards a constructivist perspective: An intervention study of mathematics teacher development. *Educational Studies in Mathematics*, *22*, 309-331.

Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan*, *72*, 534-539.

United States Department of Education. (1994). *Improving America's Schools Act of 1994: Reauthorization of the Elementary and Secondary Education Act*. Highlights of the Act. [On line]. Available via gopher: /Other gophers by subject/Education and Research (K-12)/U.S. Department of Education/ Elementary and Secondary Education (OESE) and Early Childhood/ Improving America's Schools Act of 1994. Updated October 20.

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

Wood, P., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89-100.