Building Student Capacity for Mathematical Thinking and Reasoning: An Analysis of Mathematical Tasks Used in Reform Classrooms
Author(s): Mary Kay Stein, Barbara W. Grover, Marjorie Henningsen
Reviewed work(s):
Source: *American Educational Research Journal*, Vol. 33, No. 2 (Summer, 1996), pp. 455-488
Published by: American Educational Research Association
Stable URL: http://www.jstor.org/stable/1163292
Accessed: 13/01/2012 14:28

American Educational Research Journal
Summer 1996, Vol. 33, No. 2, pp. 455–488

Building Student Capacity for Mathematical Thinking and Reasoning: An Analysis of Mathematical Tasks Used in Reform Classrooms

Mary Kay Stein
Learning Research and Development Center, University of Pittsburgh
Barbara W. Grover
Ohio University
Marjorie Henningsen
Learning Research and Development Center, University of Pittsburgh

This article focuses on mathematical tasks as important vehicles for building student capacity for mathematical thinking and reasoning. A stratified random sample of 144 mathematical tasks used during reform-oriented instruction was analyzed in terms of (a) task features (number of solution strategies, number and kind of representations, and communication requirements) and (b) cognitive demands (e.g., memorization, the use of procedures with [and without] connections to concepts, the "doing of mathematics"). The findings suggest that teachers were selecting and setting up the kinds of tasks that reformers argue should lead to the development of students' thinking capacities. During task implementation, the task features tended to remain consistent with how they were set up, but the cognitive demands of high-level tasks had a tendency to decline. The ways in which high-level tasks declined as well as factors associated with task changes from the set-up to implementation phase were explored.

MARY KAY STEIN is a Research Associate at the Learning Research and Development Center, University of Pittsburgh, Room 708, 3939 O'Hara St., Pittsburgh, PA 15260. Her specializations are classroom-based teaching and learning in mathematics and teacher development.

BARBARA W. GROVER is an Associate Professor at Ohio University, 554 Morton Hall, Athens, OH 45701. Her specializations are assessment and teacher preparation.

MARJORIE HENNINGSEN is a Research Specialist at the Learning Research and Development Center, University of Pittsburgh, Room 728, 3939 O'Hara St., Pittsburgh, PA 15260. Her specialization is the study of mathematics classroom processes and teacher development.

The mathematics reform movement posits an ambitious set of outcome goals for student learning. Documents published by the National Council of Teachers of Mathematics (1989,1991), the Mathematical Association of America (1991), and the National Research Council (1989) all point to the importance of students' developing deep and interconnected understandings of mathematical concepts, procedures, and principles, not simply an ability to memorize formulas and apply procedures. Increased emphasis is being placed not only on students' capacity to understand the substance of mathematics but also on their capacity to "do mathematics." In recent years, mathematics educators and philosophers have convincingly argued that full understanding of mathematics consists of more than knowledge of mathematical concepts, principles, and their structure (e.g., Lakatos, 1976; Kitcher, 1984; Schoenfeld, 1992). Complete understanding, they argue, includes the capacity to engage in *the processes* of mathematical thinking, in essence doing what makers and users of mathematics do: framing and solving problems, looking for patterns, making conjectures, examining constraints, making inferences from data, abstracting, inventing, explaining, justifying, challenging, and so on. Students should not view mathematics as a static, bounded system of facts, concepts, and procedures to be absorbed but, rather, as a dynamic process of "gathering, discovering and creating knowledge in the course of some activity having a purpose" (Romberg, 1992, p. 61).

What types of instructional environments might reasonably be expected to produce these kinds of student outcomes? Most reformers agree that "classrooms must be communities in which mathematical sense-making of the kind we hope to have students develop is practiced" (Schoenfeld, 1992, p. 345). According to the *Professional Standards for the Teaching of Mathematics* (NCTM, 1991), classrooms should be environments in which students are encouraged to discuss their ideas with one another, where intellectual risk-taking is nurtured through respect and valuing of student thinking, and where sufficient time and encouragement is provided for exploration of mathematical ideas. One also finds consistent recommendations for the exposure of students to meaningful and worthwhile mathematical tasks, tasks that are truly problematic for students rather than simply a disguised way to have them practice an already-demonstrated algorithm. In such tasks, students need to impose meaning and structure, make decisions about what to do and how to do it, and interpret the reasonableness of their actions and solutions. Such tasks are characterized by features such as having more than one solution strategy, as being able to be represented in multiple ways, and as demanding that students communicate and justify their procedures and understandings in written and/or oral form.

This characterization of instructional environments for the development of mathematical thinking stands in sharp contrast to the ways in which most classrooms are currently organized and run. Most mathematics lessons consist of teacher presentation of a "mathematical problem" along with the algorithm for solving it, followed by the assignment of a similar set of problems for students to work through individually at their seats (Porter, 1989; Stodolosky, 1988). Students' work consists almost entirely of memorizing presented facts or applying formulas, algorithms, or procedures without attention to why or when it makes sense to do so. In this type of setting, "*Doing* mathematics means following rules laid down by the teacher; *knowing* mathematics means remembering and applying the correct rule when the teacher asks a question, and mathematical *truth* is determined when the answer is ratified by the teacher" (Lampert, 1990, p. 31). Classrooms organized in this type of format do not provide the conditions necessary for the development of students' capacity to think and reason mathematically. Over time, students come to expect that there is one right method for solving problems, that the method should be supplied by the teacher, and that, as students, they should not be expected to spend their time figuring out the method or taking the responsibility for determining the accuracy or reasonableness of their work (Schoenfeld, 1992).

Given the numerous calls for the establishment of instructional environments characterized by an increased emphasis on problem solving, sense making, and discourse, a closer examination of the assumptions regarding *how* such environments lead to the desired student outcomes appears to be in order.[1] One model that could be used to examine how instruction relates to student learning outcomes (the *student mediation* model) specifies that teaching does not *directly* influence student learning but, rather, that teaching influences students' cognitive processes or thinking, which, in turn, influences their learning (Carpenter & Fennema, 1988; Wittrock, 1986). From this perspective, a mediating variable that is important to describe and examine is the nature of students' thinking processes in the classroom and how those processes are altered when teachers attempt to create enhanced instructional environments. Are authentic opportunities for students to think and reason created when teachers use tasks that are problematic, that have multiple solution strategies, that demand explanation and justification, and that can be represented in various ways? What kinds of thinking processes do these types of tasks set into motion? If students are not being set on the right cognitive track during classroom lessons, there is little reason to expect that scores on measures of learning outcomes will reflect enhanced understanding or increased ability to think and problem solve.

This article investigates enhanced instruction as a means of building student capacity for mathematical thinking and reasoning. The underlying premise is that students must first be provided with opportunities, encouragement, and assistance to engage in thinking, reasoning, and sense-making in the mathematics classroom. Consistent engagement in such thinking practices should, in turn, lead to a deeper understanding of mathematics as well as the ability to demonstrate complex problem solving, reasoning, and communication skills on assessments of learning outcomes.

The context for the present investigation consists of mathematics classrooms that are participating in the QUASAR Project,[2] a national educational reform project aimed at fostering and studying the development and implementation of enhanced mathematics instructional programs for students attending middle schools in economically disadvantaged communities (Silver & Stein, 1996). The project is based on the premise that prior failures of poor and minority students were due to a lack of opportunity to participate in meaningful and challenging learning experiences rather than to a lack of ability or potential. Beginning in the fall of 1990, mathematics teachers at four geographically dispersed middle schools have been working, in collaboration with resource partners from nearby universities, to provide instruction based on thinking, reasoning, and problem solving. Two additional middle schools were added in the fall of 1991. Two of the QUASAR schools serve student populations that are predominately African American; two serve primarily Hispanic student populations, and the other two schools serve ethnically diverse student populations. In two of the schools, the majority of students speak English as their second language.

QUASAR's reform efforts are targeted at the school level. As such, all teachers who teach mathematics at QUASAR sites are involved in project activities. Although QUASAR teachers have received a broad array of staff development since the inception of the project, the teachers' educational and professional backgrounds prior to joining the project were typical of most middle school mathematics teachers (QUASAR Documentation Team, 1993). The majority is elementary certified and has taken few, if any, mathematics courses beyond high school (other than mathematics teaching methods courses in college). At the time of the final year of instruction represented in the current study, the average years of teaching experience was slightly over 13 (the range was from one year's experience to over 20 years' experience). Compared to the national average, QUASAR teachers are more ethnically diverse.

Conceptual Framework

Investigation of the relationship between instruction and students' thinking in project classrooms was guided by the conceptual framework shown in Figure 1. The framework proposes a set of differentiated task-related variables as leading toward student learning and proposes sets of factors that may influence how the task variables relate to one another. The present investigation focused on the variables that appear in the shaded areas.

Mathematical Tasks

The examination of instruction and thinking processes was framed by the concept of *mathematical tasks*, a close relative of *academic tasks*, a construct that has been extensively employed to study the connections between teaching and learning in classrooms more generally. As the first individual

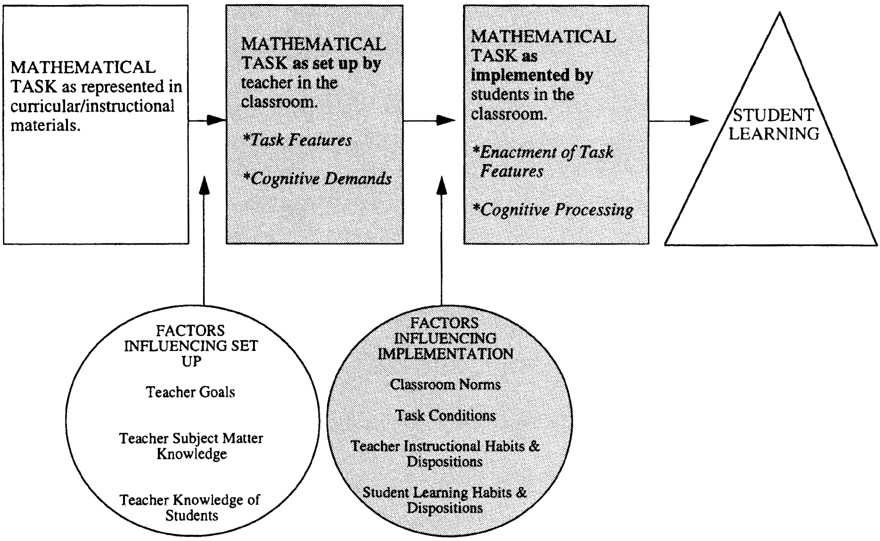

Figure 1. **Relationship among various task-related variables and student learning. Shaded portions represent areas under investigation**

to view the curriculum as a collection of academic tasks, Doyle defined academic tasks as the products that students are expected to produce, the operations that students are expected to use to generate those products, and the resources available to students while they are generating the products (1983, p. 161). Doyle has suggested that an academic-task approach to classroom research constitutes a "treatment theory to account for how students learn from teaching" (p. 167, 1988). Although recognizing that other events and contexts influence student learning, he joins Shavelson, Webb, and Burstein (1986) in proposing that academic tasks serve as the proximal causes of student learning from teaching. "Tasks influence learners by directing their attention to particular aspects of content and by specifying ways of processing information" (Doyle, 1983, p. 161). From this perspective, the mathematical tasks with which students become engaged determine not only what substance they learn but also how they come to think about, develop, use, and make sense of mathematics. Indeed, an important distinction that permeates research on academic tasks is the differences between tasks that engage students at a surface level and tasks that engage students at a deeper level by demanding interpretation, flexibility, the shepherding of resources, and the construction of meaning.

The conception of mathematical task used in the present article is similar to Doyle's notion of academic task in that it includes attention to what students are expected to produce, how they are expected to produce it, and with what resources. It is different from Doyle's, however, in terms of duration or length of the task. In this article, a *mathematical task* is defined as a classroom activity, the purpose of which is to focus students' attention on a particular mathematical idea. An activity is not classified as a different or new task unless the underlying mathematical idea toward which the activity is oriented changes. Thus, a lesson is typically divided into two, three, or four tasks rather than into many more tasks of shorter duration.[3]

A central theme in academic-task research is the extent to which tasks can change their character once unleashed in real classroom settings. As shown by the rectangles in Figure 1, a task can be viewed as passing through three phases: first, as curricular or instructional materials; second, as set up by the teacher in the classroom; and third, as implemented by students during the lesson. The properties of tasks may be different during each of these phases (Doyle & Carter, 1984). This phenomenon is well-known to classroom ethnographers and to sociocultural researchers, as well as to those individuals who have done research on the nature of academic tasks. For example, Newman, Griffin, and Cole (1989) have provided extended ethnographic investigations surrounding the ways in which students' goals and their understanding of the objectives of the task can transform the task to the point that it is no longer the same as what was intended by the teacher at the outset. Teachers also can wittingly (or unwittingly) change the nature of tasks by stressing less- or more-challenging aspects of the tasks or by altering the resources available to students.

As shown in Figure 1, tasks can be potentially transformed between any two successive phases. For example, the circle between the first two boxes suggests factors that may influence how the teacher actually sets up curricular/instructional tasks in the classroom. Although this area of the framework is not the focus of the present investigation, other studies have found differences between the objectives of curricular materials and the ways in which teachers have interpreted and set up the material. For example, studies have shown that teachers' knowledge of subject matter can influence the manner in which they use text materials (e.g., Stein & Baxter, 1989).

Task Set Up and Implementation

This study focused on the relationship between task set up and task implementation. *Task set up* is defined as the task that is announced by the teacher. It can be quite elaborate, including verbal directions, distribution of various materials and tools, and lengthy discussions of what is expected. Task set up can also be as short and simple as telling the students to begin work on a set of problems displayed on the blackboard. *Task implementation*, on the other hand, is defined by the manner in which students actually work on the task. Do they carry out the task as it was set up? Or do they somehow alter the task in the process of working their way through it?

At the task-set-up phase and also at the task-implementation phase, the mathematical tasks were examined in terms of two interrelated dimensions: task features and cognitive demands. The *task features* refer to aspects of tasks that mathematics educators have identified as important considerations for the engagement of student thinking, reasoning, and sense-making: the existence of multiple-solution strategies, the extent to which the task lends itself to multiple representations, and the extent to which the task demands explanations and/or justifications from the students. At the task-set-up phase, task features refer to the extent to which the task as announced by the teacher incorporates or encourages the use of each of these features. At the task-implementation phase, task features refer to *the enactment of the features* by students as they actually go about working on the task. Are multiple-solution strategies actually used? Do students actually produce mathematical explanations and justifications? The *cognitive demands* of the task during the task-set-up phase refer to the kind of thinking processes entailed in solving the task as announced by the teacher. These can range from memorization, to the use of procedures and algorithms (with or without attention to concepts or understanding), to the employment of complex thinking and reasoning strategies that would be typical of "doing mathematics" (e.g., conjecturing, justifying, interpreting, etc.). At the task-implementation phase, cognitive demands are analyzed as the *cognitive processes in which students actually engage* as they go about working on the task. Do students actually memorize facts and formulas? Do students actually engage in high-level thinking and reasoning about mathematics?

The circle between task set up and task implementation in Figure 1 identifies various types of factors that could potentially influence the way in which tasks are actually implemented in the classroom. These include classroom norms, task conditions, and teachers' and students' habits and dispositions. Classroom norms refer to established expectations regarding how academic work gets done, by whom, and with what degree of quality and accountability. Task conditions refer to attributes of tasks as they relate to a particular set of students (e.g., the extent to which tasks build on students' prior knowledge, the appropriateness of the amount of time that is provided for students to complete tasks). Teachers' and students' habits and dispositions refer to relatively enduring features of their pedagogical and learning behaviors that tend to influence how they approach classroom events. Examples include the extent to which a teacher is willing to let a student struggle with a difficult problem, the kinds of assistance that teachers typically provide students that are having difficulties, and the extent to which students are willing to persevere in their struggle to solve difficult problems. These classroom, task, and teacher/student factors are illustrations of the ways in which tasks can be shaped by the ambient classroom culture.

Cognitively Demanding Tasks

Of particular interest to mathematics reform are those instances in which tasks start out as cognitively demanding but, during the course of implementation, *decline* into somewhat less-demanding activities versus those instances in which tasks start out as cognitively demanding and *remain so.* Academic-task researchers have provided descriptions of ways in which high-level academic tasks can decline into less cognitively demanding activities as they are implemented. High-level tasks are often less structured, more complex, and longer than tasks to which students are typically exposed. According to Doyle and others, students often perceive such tasks as ambiguous and/or risky because it is not apparent what they should do and how they should do it. In order to manage this ambiguity and risk, students, it is argued, often urge the teacher to make such tasks more explicit, thereby reducing or eliminating the difficult, sense-making aspects of the task. In addition, task researchers have noted that high-level tasks are not typically associated with quick student entry and engagement, even work production, and smooth, well-ordered classroom management. Thus, they argue, the increased complexity of orchestrating classroom events can lead to teacher imposition of procedures that detract from the challenging aspects of the task.

Tasks that begin as cognitively demanding do not always decline, however, and it is important to understand when and how such tasks remain challenging during the implementation phase. Along these lines, recent work in the cognitive psychology of instruction has begun to outline characteristics of instruction for the development of high-level thinking and reasoning skills (see Anderson, 1989, for a summary of such findings). Examples of factors that have been found to be associated with students' maintenance of steady effort and progress in the face of complex task demands include scaffolding, the modeling of high-level performance by the teacher and/or capable students, the making of conceptual connections, the careful selection of tasks that build on students' prior knowledge, the provision of appropriate amounts of time to explore ideas and make connections, the encouragement of student self-monitoring, and the presence in the environment of a sustained press for explanation, meaning, and understanding. On the whole, however, it is fair to say that less is known about characteristics of instruction that foster high-level thinking than about instruction for the facilitation of basic knowledge and skills (i.e., direct instruction).

In summary, the tasks used in mathematics classrooms highly influence the kinds of thinking processes in which students engage, which, in turn, influences student learning outcomes (as represented in the triangle in Figure 1). When employing the construct of mathematical task, however, one needs to be constantly vigilant about the possibility that the tasks with which students actually engage may or may not be the same task that the teacher announced at the outset. Be they enabling or constraining, one needs to acknowledge the influences of the complex environment of the classroom on the ultimate shape and form of the task.

Purpose of the Study

The overall purpose of the present study was to examine and describe the nature of the mathematical tasks used in project classrooms. Using the conceptual framework as a guide, the study was organized to provide answers to three interrelated questions:
(1) To what extent do tasks *as set up* by teachers include selected features that the mathematics education research and reform communities would view as associated with the development of students' capacity to think and reason mathematically?
(2) To what extent do the tasks *as implemented* remain consistent with the ways in which they were set up?
(3) For those tasks that were set up to place high-level cognitive demands on students, what factors were associated with (a) those instances in which the tasks were implemented in such a way that students *did* engage with the task at that level and (b) those instances in which the tasks were implemented in such a way that students *did not* engage with the task at that level.

Methodology

Data Sources

Narrative summaries. Narrative summaries of classroom observations written by trained and knowledgeable observers formed the basis of the data used for our analysis. Each school year from Fall 1990 to Spring 1993, three 3-day observation sessions (Fall, Winter, and Spring) were conducted in three teachers' mathematics classrooms at four project sites.[4] An observer took detailed field notes focusing on the mathematics instruction and students' reactions to the instruction; simultaneously, a camera operator videotaped the lesson. Following the observations, the observer used both the videotaped lesson and his or her field notes to complete the project's Classroom Observation Instrument (COI).[5] As part of that instrument, the observer provided descriptions and sketches of the physical setting of the room, a chronology of instructional events, and responses to questions associated with five themes: mathematical tasks, classroom discourse, the intellectual environment, management and assessment practices, and group work (if it occurred).

In the COI, a mathematical task is defined as a segment of classroom work that is devoted to learning about a particular mathematical idea. The observers were instructed to segment the instructional time of each observed lesson into the main mathematical tasks with which students were engaged and to append artifacts associated with these tasks to the write-up. The two tasks that occupied the largest percentage of class time were designated as Task A and Task B. In the mathematical tasks section of the COI, the observer described in detail the nature of these two tasks: their mathematical content, the learning goals of the teacher for each task, and the behaviors of the students as they engaged in these tasks. The observer also described the extent to which each task focused students' attention on procedural steps with or without connections to underlying concepts and on "doing mathematics" (e.g., framing problems, making conjectures, justifying, explaining). In the remaining three sections of the COI, the observer considered all activities that occurred during the classroom lesson in their responses, often referring specifically to Task A or Task B. Only Task A of each observation was the focus of the analysis. The entire narrative summary for a classroom observation, however, was reviewed and considered in making coding decisions.

Observer qualifications. The observers were selected on the basis of a set of qualifications that included a strong background in mathematics education, psychology, or a related field; a demonstrated competence in their ability to analyze instructional events from both pedagogical and mathematical content perspectives; prior experience observing classrooms; and their understanding of the ethnic or multicultural nature of the community at the site (many of the observers were residents of those communities). In some instances, Spanish-English bilingual skills were also required because the population included a high percentage of students whose native language was Spanish.

A team of central project staff (including the authors of this article) conducted a 2-day training session for the observers or provided one-on-one on-site training. To ensure the validity of the write-ups, two members of the central project staff independently viewed a videotape of a randomly chosen observation for each observer, each season and independently critiqued the write-up of that observation. The two reviewers provided the observer with jointly constructed detailed feedback on the write-ups prior to the next observation cycle.

Videotapes and artifacts. Videotapes of observations and/or additional artifacts from an observation were used as supplemental data sources on 11 of the tasks (8%). These sources were consulted when a coder determined that the written description of the observation did not provide sufficient information on which to base a decision. Viewing the videotape or reviewing the artifacts provided the necessary supplemental information that assisted the coder in making more accurate coding decisions.

Sampling Procedure

The present investigation used data from the four sites that had participated in the project for a full 3-year period by Spring 1993. That database consisted of 620 tasks (2 tasks x 310 observations). A stratified random sample was selected, using year, site, and teacher as stratification dimensions.[6] Two teachers were selected from each site for each year on the basis of the grade-level classes they taught and whether their classes were represented elsewhere in the sample (e.g., where possible, a teacher whose classes were not represented in the sample for 1990–1991 was selected in 1991–1992 or 1992–1993). Two of the three observations from the fall, two of the three from the winter, and two of the three from the spring were randomly chosen for each of the selected teachers,[7] (see Table 1). Thus, 12 observations were selected from each site for each year, resulting in 144 observations overall (12 observations x 4 sites x 3 years). In each of the observations, the task that accounted for the greatest amount of class time was used as the task for analysis (Task A).[8] These 144 tasks constituted 23% of the entire data pool.

This sampling procedure resulted in an equal distribution of tasks across seasons (48 per season), across sites (36 per site), and across years (48 per year). During the first year of the project (1990–1991), all observations occurred in sixth grade classrooms; during the second year of the project (1991–1992), the emphasis was on seventh grade classrooms and nearly all observations were at that grade level, with a few at the sixth grade level; during the final year of the project, the emphasis was on eighth grade classrooms, and, thus, the majority of observations were at that grade level, with a few in sixth and seventh grade classrooms. This distribution is reflected in our sample as follows: for 1990–1991, 100% of the tasks are from sixth grade classes; for 1991–1992, 96% are from seventh grade classes, and 4% are from sixth grade classes; and, for 1992–1993, 71% of the tasks are from eighth grade classes, 17% from seventh grade classes, and 13% from sixth grade classes. Overall, 39% of the sample tasks are from sixth grade classes, 38% from seventh grade classes and 24% from eighth grade classes.

Coding

Coding procedure. The completed COIs which contained the 144 tasks described above were coded using a system designed specifically for the

Table 1

Example of Selected Observations for a Given Site for a Given Year

	1990-1991 Site 1			
	Teacher 1		Teacher 2	
Fall	Obs. 1	Obs. 3	Obs. 2	Obs. 3
Winter	Obs. 2	Obs. 3	Obs. 1	Obs. 3
Spring	Obs. 1	Obs. 2	Obs. 1	Obs. 3

present study. The coding system was initially developed based on a review of the literature on academic tasks (Bennett & Desforges, 1988; Doyle, 1983; 1988; Marx & Walsh, 1988) and the cognitive psychology of instruction (Anderson, 1989), the literature on mathematical thinking and problem solving (Grouws, 1992; Silver, 1985), and mathematics reform documents (NCTM, 1989, 1991), as well as on our knowledge of the project sites and their goals. The system was modified through the process of actually attempting to code COIs. Hence, the final coding system reflects important features of tasks as suggested by theory and prior research and also salient characteristics of the project and the data set.

Nineteen coding decisions were made for each task. The codes were organized into 4 main categories: task description, task set up, task implementation, and factors associated with decline or maintenance of high-level tasks. The *descriptive* codes included the number of minutes and percentages of class time devoted to the task, the type of resource(s) that served as the basis or idea for the task (i.e., textbook, innovative curricula, teacher-developed material), the type of mathematical topic that the task was about (conventional middle-school topic, reform-inspired topic, focus on mathematical processes more than a particular topic), the context of the task, and whether or not the task was set up as a collaborative venture among students.

The second category of codes was concerned with the *set up* of the task. In this phase of the coding, coders were instructed to refer to the task materials (provided as appendices to the COI write up) and to the task as specified by the teacher, both during her initial announcement of what students were to do and at any subsequent points during the task in which the teacher unilaterally provided additional specifications to guide students' approach to the task. Codes were assigned for task features and for the cognitive demands of the task. The features included the number of possible solution strategies, the number and kind of potential representations that could be used to solve the problem, and the communication requirements of the task (i.e., the extent to which students were required to explain their reasoning and/or justify their answers). The cognitive demands were classified with respect to the following: memorization, the use of formulas, algorithms, or procedures *without connection* to concepts, understanding, or meaning; the use of formulas, algorithms, or procedures *with connection* to concepts, understanding, or meaning; and cognitive activity that can be characterized as "doing mathematics," including complex mathematical thinking and reasoning activities such as making and testing conjectures, framing problems, looking for patterns, and so on. These selections are not necessarily mutually exclusive; when the task appeared to call for more than one type of cognitive activity, coders were instructed to select the code that best described the majority of the task.

The third category of codes represented the task as it was *implemented*. In this phase of the coding, coders were instructed to attend to the ways in which students actually went about working on the task. Once again, codes were assigned for task features (i.e., solution strategies, representations, communication) and cognitive demands. When coding the features of the task as implemented, coders were instructed to infer the extent to which students appeared to be carrying out the various task features as stipulated in the set up (e.g., the extent to which they actually used single- vs. multiple-solution strategies, the extent to which they used and made connections among multiple representations, and the extent to which students actually produced explanations). When coding the cognitive demands of the task as implemented, coders were asked to make judgments about the kinds of cognitive processes in which the majority of the students appeared to be engaged. In other words, was there evidence that students actually employed the cognitive processes that the task set up called for?[9]

The final category of codes included judgments about *factors* associated with task implementation. In this category, only selected tasks were coded: (a) those that were set up to require high levels of cognitive activity and were implemented in such a way that students *did* indeed engage in high levels of cognitive activity and (b) those that were set up to require high levels of cognitive activity but were implemented in such a way that students did *not* engage with the task at high levels. *High level* was defined as tasks that involved "doing mathematics" or the use of formulas, algorithms, or procedures with connection to concepts, understanding, or meaning. All other levels of cognitive demand and activities (i.e., memorization; the use of formulas, algorithms, or procedures without connection to concepts, understanding, or meaning; codes designated as "other") were considered to represent lower levels of cognitive demand. For high-level tasks that remained so during implementation, coders were instructed to select as many as applied from a list of factors that could assist with the maintenance of tasks at high levels (e.g., the modeling of high-level performance by teachers or capable students, sustained press for justification, explanations, and/or meaning through teacher questioning, comments, and feedback; scaffolding [teachers or more capable students simplifying the task so that it could be solved while maintaining task complexity]; and the selection of tasks that build on students' prior knowledge). For high-level tasks that declined, coders selected possible reasons for the decline from a list that included the routinization of problematic aspects of the tasks (students press teacher to reduce task ambiguity or complexity by specifying explicit procedures and/or teacher takes over difficult pieces of the task); the shifting of emphasis from meaning, concepts, or understanding to the accuracy and completeness of answers; the lack of sufficient time for students to wrestle with the demanding aspects of the tasks; and classroom management problems that prevent sustained engagement in high-level cognitive activities.

For most questions, coders had the option of selecting "other" and writing a phrase to specify what was meant by "other." When coding the implementation phase of the cognitive demands question (the cognitive processes in which students actually engaged), coders independently recognized the need for a new code to describe a frequently observed manner of implementing "doing mathematics" tasks (the manner of implementation did not fit into any of the existing codes). The coders agreed to use the string of words, *inadequate implementation of "doing mathematics,"* to describe this mode of implementation, and hence the word string was written with the "other" code when appropriate. The use of this uniform text string allowed the authors to easily identify this form of cognitive activity during the analysis phase.

Coder assignments. A stratified random sampling procedure was used to assign observations to four coders to ensure that each coder had responsibility for coding about the same number of observations from each site and about the same number from each year. Eight to 10 tasks from each site and 12 tasks from each year were assigned to each coder. Circumstances toward the end of the coding period required that six of one of the coder's tasks be equally distributed among the remaining three coders. This resulted in one coder coding 6 1992–1993 tasks (instead of 12) and three coders coding 14 1992–1993 tasks (instead of 12). In addition, this change resulted in one coder coding only 4 tasks from one site while another coded 12 tasks from that same site (instead of the 8–10 originally planned).

Double coded tasks. To ensure a representative subset were double coded, a stratified random sampling procedure was used to identify 36 tasks (25% of the sample) for intercoder reliability purposes. Three tasks from each site for each year (3 tasks x 4 sites x 3 years) were independently coded by two individuals. The 36 tasks included at least one task from each teacher; they were distributed across seasons such that 10 tasks were from the fall, 14 from the winter, and 12 from the spring; they were distributed across grade levels such that 12 were sixth grade tasks, 14 were seventh grade, and 10 were eighth grade. Consensus coding was scheduled systematically over a 5-week period such that each coder independently coded from 2–10 tasks between consensus sessions. Consensus was reached by the two coders on all disagreements.

Intercoder reliability. Intercoder reliability ranged from 53% to 100% with an average of 79%. We consider this percentage of agreement to be sufficiently high enough to warrant confidence in our conclusions. The inferential nature of the coding decisions and the fact that 10–15 pages of written text of the COI were brought to bear on the decision-making process made the coding task complex and intellectually demanding. The frequent consensus sessions interspersed among individual coding minimized opportunities for individual coders to drift away from the shared understanding of the meaning of the codes.

Analysis Procedures

Along with appropriate identification codes (e.g., teacher, site, year, season, observation date, grade level), codes for the four categories described earlier under "Coding Procedure" were entered into a *4th Dimension Version 3.0.5 (4th Dimension,* 1985–1993) computer database. Written text that accompanied codings of "other" were also entered when appropriate. Using *Systat 5 for the Macintosh (Systat,* 1989), initial analysis reported the frequencies and percentages for each possible code across all 144 tasks across all 3 years. These results were reviewed with respect to patterns that emerged, potential explanations for the results, and possible approaches to reporting the data.

In order to analyze the relationships between the set up and implementation of the features and the cognitive demands of the tasks, four matrices were generated: three of the matrices organized information related to the set-up and implementation versions of questions about solution strategies, representations, and communication; one of the matrices organized information related to the set-up- and implementation-versions of the questions about cognitive demands. For each pair of items, the row headings listed the possible responses for task set up; the column headings listed the possible responses for task implementation. In each cell, the appropriate frequencies and percents were recorded. The percentages in the cells along the diagonals of each matrix reflected the extent of agreement between the way in which the task was set up and the way in which it was implemented (e.g., the percentage of tasks in which students used multiple-solution strategies when the task as set up allowed for multiple-solution strategies). The off-diagonal cells reflected the extent to which aspects of the task as set up changed when implemented and what those changes were (e.g., the percentage of tasks in which students used single-solution strategies when the task as set up allowed for multiple-solution strategies). In addition, a listing of text descriptions associated with the code of "other" for the cognitive demands questions was printed out and examined. The four matrices and the text listings were used as the bases for discussions about the identification, meaning, and interpretation of the patterns and the relationships between task set up and implementation that emerged.

Results

The discussion of results is presented in four main sections. First, a descriptive summary of the basic attributes of the mathematical tasks used in project classrooms is provided. The next three sections address each of the research questions. Findings related to the extent to which the tasks *as set up by* teachers included the kinds of features and cognitive demands that are typically associated with the development of students' ability to think and reason mathematically are presented. This is followed by a discussion of the findings related to the extent to which the *implementation* of the mathematical tasks remained consistent with the manner in which the tasks were set up. The section concludes with a discussion of the results of analyses that explored the factors that appeared to be associated with high-level tasks that (a) declined into less demanding classroom activity and (b) remained at high levels during implementation.

Description of Mathematical Tasks

The 144 mathematical tasks that were coded ranged from 10 to 51 minutes, with an average length of 24 minutes. On average, a task comprised 52% of the total instructional time of the given mathematics lesson. Thus, project teachers appeared to be devoting sustained periods of classroom time to academic work for the purpose of developing student facility with a particular mathematical idea.

Resources that served as the basis or idea for the tasks were distributed among basal textbooks, commercial innovative curricula, site- or teacher-developed activities, and commercial supplemental resource books. The resource most often used was material developed by the site or teachers themselves, with 39% of the tasks being created by project participants. Commercial innovative curricula, such as *The Middle Grades Mathematics Project* (Fitzgerald, Winter, Lappan, & Phillips, 1986; Schroyer & Fitzgerald, 1986) were used nearly as often (for 30% of the tasks). It should be noted, that, when the project began in Fall 1990, there were fewer commercial innovative curricula on the market than there are now. Another resource frequently used as a basis for tasks was commercial supplemental resource books; workbooks such as *Make It Simpler* (Meyer & Sallee, 1983) accounted for 19% of the tasks. Finally, 11% of the tasks were based on regular textbook series.

Over half of the task topics were judged to be *reform-inspired,* meaning that they were topics that have not typically enjoyed center stage in more traditional middle-school mathematics programs. Fifty-one percent of the topics fell into the following categories: statistics, algebra, geometry, patterns and functions, and probability. Most of the remaining tasks (42% of the total) were more conventional in character (e.g., fractions, whole number operations, changing from percents to decimals to fractions). Seven percent of the tasks were judged to be focused on mathematical process more so than on a mathematical topic. The majority of these process-focused tasks were driven by an emphasis on problem solving as skill; a few focused on group work skills.

The majority of the tasks (63%) had no "real-life" contexts;[10] rather, they were situated totally in the abstract world of mathematics. Twenty-six percent of the tasks did refer to real-life objects or events, while 12% were found to utilize both abstract and real-world contexts. Finally, the majority of tasks (69%) were set up in such a way that students could use one another as resources. Either students were explicitly encouraged to work with one another, or the classroom norms were such that it was evident that group or pair work was sanctioned and/or expected. In 30% of the tasks, students were either told to work alone, or they participated in whole-class discussions in which students were expected to contribute as individuals. This contrasts sharply with conventional mathematics classrooms in which students typically work independently.

Task Set Up

Figures 2 and 3 illustrate the extent to which the mathematical tasks as set up by project teachers possessed the kinds of features and cognitive demands that mathematics education researchers and reformers would identify as being associated with the development of student capacity to think and reason mathematically.

As shown in Figure 2, the tasks tended to embody the kinds of features that reformers have suggested should be used in classroom instruction if the goal is to produce student learning outcomes such as the ability to understand mathematics and to think and reason in complex ways. Approximately two thirds of the tasks could be solved in multiple ways, whereas about one third lent themselves to only one solution path. The inclusion of many tasks that have multiple-solution strategies would be seen as one way of helping students to develop the view that mathematics involves making decisions

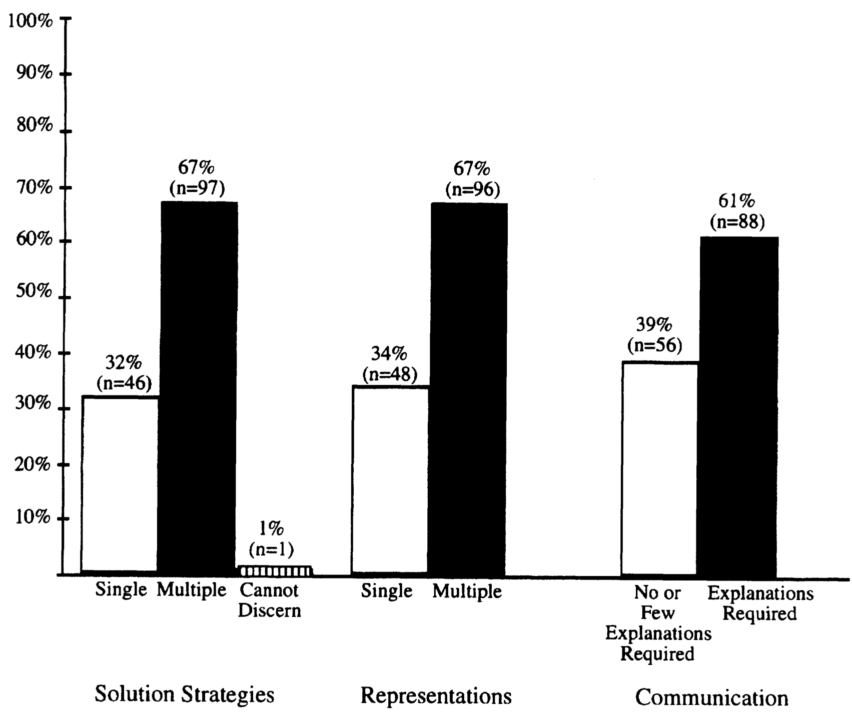

Figure 2. **Task set up: Features**

about how to go about solving problems, not simply employing teacher-supplied procedures. Similarly, two thirds of the tasks were set up to include the use of multiple representations, while about one third were set up as tasks that were to be solved using only one representation. Although more representations do not necessarily lead to greater understanding, numerous cognitive advantages associated with the establishment of links among various ways of representing a problem have been proposed. Finally, Figure 2 shows that the majority of the tasks (61%) were set up to include the requirement that students produce mathematical explanations or justifications. This represents a fairly radical departure from conventional classrooms in which tasks are generally set up to require only an answer. Although teachers in traditional classrooms may ask to "see students' work," this usually means a display of procedural steps used to arrive at an answer, not an explanation or justification.

Figure 3 illustrates the types of cognitive demands that the tasks were set up to require of students.

As shown in the figure, nearly three quarters of the tasks (74%) were set up to demand that the students engage in high-level cognitive processes—either the active "doing of mathematics" (40%) or the use of procedures with connection to concepts, meaning, or understanding (34%). These findings suggest that project teachers were attempting to develop their students' capacities to engage deeply with the mathematics they were learning. Eighteen percent of the tasks demanded the use of procedures without connections to concepts, meaning, or understanding. In this regard, it is important to note that mathematics educators agree that some practice on routine skills, as well as thoughtful inquiry into complex problems, is needed. Consequently, all classroom work on procedures, even if unconnected to concepts or understanding, should not necessarily be viewed as

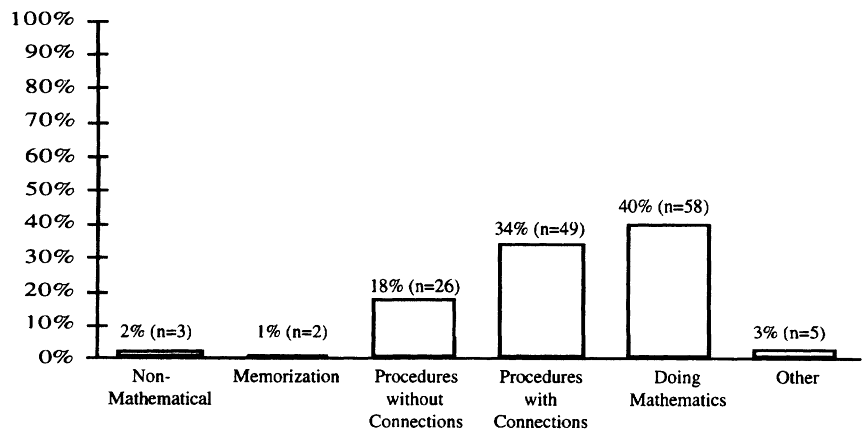

Figure 3. **Task set up: Cognitive demands**

negative. The challenge is to achieve some mixture of routine skill and understanding and, even more difficult, to integrate procedural skills with high-level thinking. Returning to Figure 3, only two tasks out of 144 were classified as demanding memorization, and only three tasks were judged to require no mathematical cognition.

Overall, the above findings suggest that project teachers were selecting and setting up the kinds of tasks that most reformers would argue should lead to the building of students' capacities to think and reason about mathematics in complex ways. The features of the tasks embodied many of the characteristics that should lead students to develop a more interconnected, dynamic, and flexible view of the domain. And the cognitive demands should direct students' attention to complex and meaningful ways of processing mathematical information.

Task Implementation

In this section, discussion focuses on the extent to which the implementation of the mathematical tasks remained consistent with the ways in which the tasks were set up. Tables 2, 3, and 4 illustrate the ways in which the task features (number of solution strategies, number and kind of representations, and communication features) changed or remained the same from set up to implementation. Table 5 reports similar information for the cognitive demands of the task. Percentages in each table are based on row totals (i.e., the sum of row percentages equal 100%).

Solution strategies. Table 2 shows the relationship between the number of solution strategies of tasks as they were set up and the number of solution strategies actually used during implementation. The bold numbers in the diagonal represent the tasks that remained consistent from task set up to task implementation.

As the percentages on the diagonal indicate, there was an overall high consistency between task set up and task implementation. Most notably, 87% of the tasks that were set up as single-method tasks did indeed lead to students' use of a single-solution strategy during the implementation phase.

Table 2
Change in Number of Solution Strategies From Set Up to Implementation

	Implementation		
Set up	Single	Multiple	Cannot discern
Single (n = 46)	***87%*** (40)	11% (5)	2% (1)
Multiple (n = 97)	12% (12)	***69%*** (67)	**19%** (18)
Cannot discern (n = 1)	0	100%	*0*

Tasks that were set up to encourage the use of multiple-solution strategies had a somewhat less consistent relationship to implementation: Students were inferred to have actually used multiple-solution strategies for 69% of tasks that were set up to encourage multiple-solution strategies, suggesting that it may be more difficult to keep multiple approaches alive during the implementation phase.

The percentages representing the number of solution strategies used during implementation need to be interpreted with caution, however, given the relatively high percentage of "cannot discern" (19%—see Table 2, row 2, column 3). It can be confidently stated that at least 12% of the tasks that were set up to encourage multiple strategies ended up being solved using a single-solution strategy. However, more tasks may have dropped to single-solution strategy; equally plausible, however, is that more tasks may have remained at the multiple-strategy level. Given the high number of "cannot discerns," it is impossible to know.

Representations. Table 3 illustrates the relationship between the number and kind of representations encouraged by the task set up and the number and kind of representations that were used during task implementation.

Once again, the bold figures in the diagonal suggest overall high consistency between how the representational aspects of tasks were set up and how they actually were enacted during implementation. Tasks that began as requiring a single representation (symbolic or nonsymbolic)[11] generally were implemented using only a single representation. Similarly, students generally used multiple representations for tasks that were set up to encourage or request the use of more than one representation. Thus there does not appear to be much decline in numbers of representations used from the task-set-up phase to the task-implementation phase. The high degree of consistency for the representational aspects of tasks is especially

Table 3

Change in Number and Kind of Representations From Set Up to Implementation

	Implementation			
Set up	Single-Symbols only	Single-Nonsymbolic	Multiple	Cannot discern
Single-Symbols only[a] (n = 23)	***87%*** (20)	4% (1)	0	9% (2)
Single-Nonsymbolic[b] (n = 25)	0	***88%*** (22)	12% (3)	0
Multiple (n = 96)	1% (1)	4% (4)	***88%*** (84)	7% (7)
Cannot discern (0)	0	0	0	***0***

[a]Representations that are composed entirely of numerals, mathematical symbols, and mathematical notation.

[b]Representations that are either entirely nonsymbolic (most often manipulatives, diagrams, or pictures) or representations that incorporate both symbols and nonsymbols.

noteworthy given the fact that the majority of the 144 sampled tasks started out as multirepresentational. As noted under task set up, two thirds of the sampled tasks required that students use more than one representation.

As stated earlier, however, more representations do not necessarily translate into deeper understanding. The crucial factor is whether and how representations become connected or linked to one another. For the 84 tasks (Table 3, row 3, column 3) that began as multiple-representation tasks and remained so during implementation, 69 (82%) were found to have incorporated some connections between at least two different representations. Thus, evidence exists that students in project classrooms were using or being exposed to not only multiple representations per se but also representations that were linked to one another. Although the meaningfulness of the linkages was not always deep, the fact that a majority of tasks were implemented in such a way to include connected, multiple representations illustrates that the instruction in project classrooms went far beyond total reliance on symbolic manipulations.

Communication. The extent to which the communication requirements of tasks actually translated into the production of explanations and justifications by students is illustrated in Table 4.

As shown in the first cell of the table, tasks that were set up to require no or few explanations and justifications did not change much during implementation: Students produced no or few explanations to the majority (82%) of tasks that did not require them. Requiring explanations as part of the task at set up, however, did not necessarily guarantee that explanations were produced during implementation. In 23% of the cases in which the task set up required students to explain or justify their thinking, no or few explanations were actually produced during the implementation phase. Thus, it appears that requesting explanations as part of the task does not

Table 4

Change in Communication Requirements From Set Up to Implementation

	Implementation		
Set up	No or few explanations/ justifications produced	Mathematical explanations/ justification produced	Explanations, but nonmathematical or nonsupportive of answer
No or few explanations/ justifications required (n = 56)	***82%*** (46)	13% (7)	5% (3)
Mathematical explanation/ justification required (n = 88)	23% (20)	***74%*** (65)	3% (3)

automatically guarantee that student explanations will be produced as the task unfolds.

Across all 144 tasks, 72 tasks (or 50%) were implemented in such a way that it was inferred that the majority of students were explaining and justifying their thinking. Once again, this finding stands in contrast to conventional classrooms in which completing the task most often means supplying the correct answer.

Cognitive demands. The extent to which students' actual cognitive processing during implementation remained consistent with cognitive demands of tasks as they were set up is the focus of Table 5.

An overview of Table 5 reveals a tendency for the cognitive demands of tasks to stay the same or to decline from task set up to task implementation. The cognitive demands of tasks were observed to increase in only two of the 144 cases (row 2, column 5; row 3, column 4). Although it is plausible that a task may become more demanding during implementation than it was originally set up to be, this was found to be highly unusual.

A second general observation is that, the higher the cognitive demands of tasks at the set-up phase, the lower the percentage of tasks that actually remained that way during implementation. For example, the vast majority (96%) of tasks that were set up to require the use of procedures *without* connection to concepts, meaning, or understanding were implemented in such a way that students used algorithms, formulas, or procedures without attention to underlying concepts or rationales. On the other hand, tasks that were set up to require higher level thinking, such as the use of procedures with connection to concepts, meaning, or understanding, were more likely to decline during implementation. Over half (53%) of the tasks that were set up to require the use of procedures with meaningful connections failed to keep the connection to meaning alive during implementation. Similarly, there was a decline during the implementation phase of tasks that were set up to require that students engage in sustained thinking, reasoning, and the "doing of mathematics." Students were observed to actually engage in these types of cognitive processes during implementation in only 38% of these tasks. Hence, it appears as though follow-through during the implementation phase is most difficult for those kinds of tasks that reformers, philosophers, and scholars have identified as essential to building students' capacities to engage in the processes of mathematical thinking.

The manner in which high-level tasks declined is worthy of comment. The majority of tasks that were set up to require the use of procedures with meaningful connections declined to tasks in which procedures were used, but without connection to concepts or meaning (53%—row 4, column 3). Thus, it appears fairly easy for students to slip into the rote application of formulas and algorithms as they actually work their way through these types of tasks. Tasks that were set up to encourage the "doing of mathematics" (i.e., as requesting that students engage in such mathematical processes as framing and solving problems, looking for patterns, and making and testing

Table 5

Change in Cognitive Demands of Task From Set Up to Implementation

	Implementation						
Set up	No mathematical activity	Memorization	Procedures without connections	Procedures with connections	"Doing mathematics"	Other	Cannot discern
No mathematical activity required (n = 3)	***100%*** (3)						
Memorization (n = 2)		***50%*** (1)			50% (1)		
Procedures without connections (n = 26)			***96%*** (25)	4% (1)			
Procedures with connections (n = 49)	2% (1)	2% (1)	53% (26)	***43%*** (21)			
"Doing mathematics" (n = 58)	17% (10)		14% (8)	3% (2)	***38%*** (22)	26% (15)	2% (1)
Other (n = 5)						***100%*** (5)	
Cannot discern (n = 1)			100% (1)				

conjectures), however, declined in a *variety* of ways. Fourteen percent of these tasks declined into the use of procedures without meaningful connections; 17% declined into nonmathematical activity; 26% declined into "other" (see row 5).

The decline of complex, open-ended tasks into proceduralized routines has been well documented in the academic-task literature; our findings bear this tendency out, although not to the extent that one might have expected (only 14% of our sample declined in this way). The finding that 17% of the "doing mathematics" tasks ended up as tasks in which no or very little mathematical cognition occurred is surprising, however, and worthy of continued investigation. During the implementation phase, the total number of tasks that involved little or no mathematical cognitive activity was only 14. Of these, 10 began as tasks that required students to think and reason in complex ways.

The decline of "doing mathematics" tasks into "other" is more informative than it may at first appear. Of the 15 tasks that were classified as "other" during the implementation phase, 13 fell into the emergent code, "inadequate implementation of doing mathematics." This emergent code represents a characteristic pattern of decline that was independently noted (and subsequently discussed) by the coders/authors (see p. 14). Such declines were marked by motivated student engagement, well-intentioned teacher goals for complex work, and well-managed work flow. The cognitive activity, however, was not at a high enough level to be characterized as engagement in complex mathematical thinking and reasoning. Students explored, discussed, and attempted to make connections, but they missed the important and central mathematical substance. Overall, students failed to adequately engage in the process of mathematical thinking, and, perhaps more important, their teachers were unable to assist them to perfom at these higher levels. We came to label this type of decline *unsystematic exploration.*

Overall, the above findings suggest that students were able to implement the task features in a fairly consistent manner but that the cognitive demands of tasks often declined from set up to implementation. The coding for cognitive demands of tasks represented a comprehensive judgment regarding the *entire* task, while the coding for each task feature was more componential in nature (i.e., each task feature was one of several features that, when considered together, composed a worthwhile and meaningful task). As such, the differential levels of success (with respect to implementation of task features vs. implementation of cognitive demands) were understandable. Implementing tasks such that their overall demands remain high appears to be more difficult than faithful implementation of any one selected feature of a task.

Factors Associated With How High-Level Tasks Were Implemented

In this section, a general overview is presented of the kinds of factors that were found to be associated with high-level tasks that declined into less demanding cognitive classroom activity and high-level tasks that remained at high levels during implementation.

Figure 4 presents an overview of the kinds of factors associated with those tasks that declined. If a task was judged to have required high-level cognitive thinking at the set-up phase but to have been implemented in such a way that students did not actually engage in high-level thinking and reasoning, coders were instructed to select as many factors as applied from a list of possible classroom, teacher, and student factors that may have been associated with the decline. The bars in Figure 4 identify the percentage of tasks in which each particular factor was judged to be an influence in the decline.

Overall, 61 tasks exhibited a decline from the set-up phase to the implementation phase. A total of 155 factors was selected across these 61 tasks, meaning that, on average, approximately 2.5 factors were selected as influencing the decline for each task. As shown in Figure 4, the factor most often chosen (i.e., in 64% of the tasks) was that the problematic aspects of the task somehow became routinized, either through students' pressing the teacher to reduce task ambiguity and complexity by specifying explicit procedures or steps to perform or by teachers' taking over the challenging aspects of the task and either performing them for the students or telling

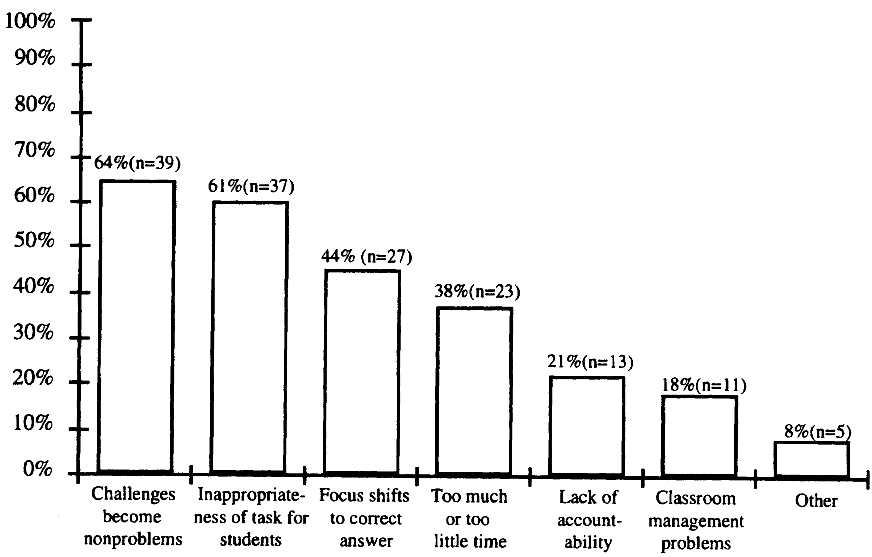

Figure 4. **Percentage of tasks in which factor was judged to be an influence in the decline (total number of tasks = 61). Percentages total to more than 100 because more than one factor was typically selected for each task**

them how to do them. In many instances, teachers appeared to find it difficult to stand by and watch students struggle, and they would step in prematurely to relieve them of their uncertainty and (sometimes) emotional distress at not being able to make headway. All too often, however, teachers would do too much for the students, taking away students' opportunities to discover and make progress on their own.

Another factor that was often selected as contributing to decline (for 61% of the tasks) was student failure to engage in high-level activities due to lack of interest, motivation, or prior knowledge. Although this factor spans a variety of reasons, the reasons all relate to the appropriateness of the task for a given group of students. The preponderance of this factor points to the importance of teachers' knowing their students well and making intelligent choices regarding the motivational appeal, appropriate difficulty level, as well as the degree of task explicitness needed to move their students into the right cognitive space so that they can actually make progress on the task.

Figure 4 also suggests that teachers had tendencies to shift the focus during the implementation phase from solution processes, meaning, concepts, and understanding to the correctness or completeness of the answer (for 44% of the tasks that declined) and to provide either too little or too much time (for 38% of the tasks that declined). Most often, a quick pace was observed to create havoc with opportunities for sustained thinking and exploration of mathematical ideas.

Perhaps the most surprising finding in this area is that classroom management problems were judged to be an influencing factor in only 18% of the tasks that declined. The literature on academic tasks has suggested that high-level tasks are often associated with problems in work flow and classroom management (Doyle, 1988). Moreover, conventional wisdom suggests that using complex thinking tasks with students in urban schools (where behavior problems and inadequate preparation in basic skills are presumed to be prevalent features of the student body) is a prescription for out-of-control classrooms. On the contrary, our findings suggest that students attending urban schools in disadvantaged neighborhoods can work on high-level tasks without becoming disruptive and nonproductive.

Figure 5 presents an overview of the kinds of factors associated with those tasks that were set up to place high-level cognitive demands on students and that remained that way during the implementation phase. If a task was judged to require high-level cognitive thinking at the set-up phase and to have been implemented in such a way that the majority of students actually engaged in high-level cognitive processing, coders were instructed to select as many factors as applied from a list of possible factors that may have been associated with assisting students to perform at high levels during the implementation phase. The bars in Figure 5 identify the percentage of tasks in which each particular factor was judged to be an influence in maintaining high levels of cognitive activity.

Overall, 45 tasks were judged to have been set up as high level and to

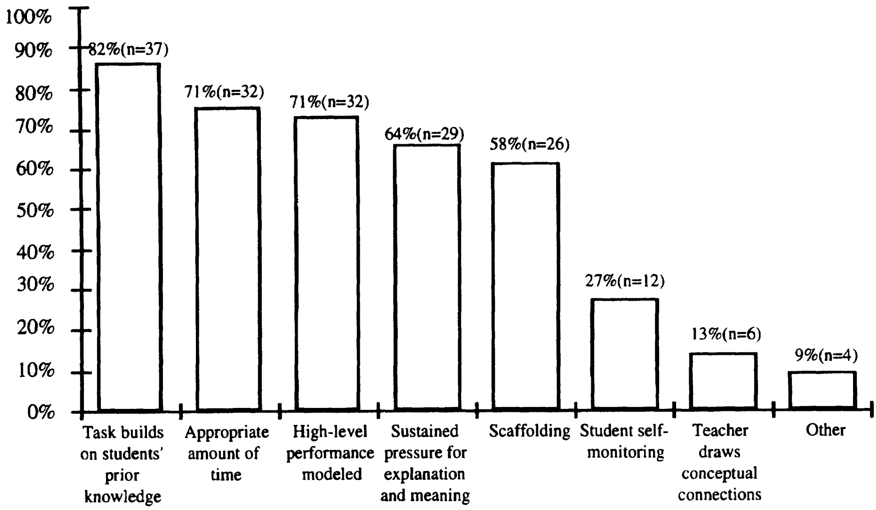

Figure 5. **Percentage of tasks in which factor was judged to be an influence in assisting students to engage at high levels (total number of tasks = 45). Percentages total more than 100 because more than one factor was typically selected for each task**

have remained so during implementation. A total of 178 factors was selected across these 45 tasks, meaning that, on average, approximately four factors were selected as assisting the maintenance of high-level cognitive activity for each task. The factor most often chosen (in 82% of the tasks that remained at high levels) was the task built on students' prior knowledge. Apparently, pitching tasks to appropriate difficulty levels so that they allow students to utilize prior relevant knowledge is an extremely important feature for helping to ensure that tasks will be implemented in the way in which they were intended. For 71% of the tasks that remained at a high level, providing the appropriate amount of time was judged to be an important feature in the success of the task. This factor most often was chosen to indicate that the teacher gave students sufficient time to explore, emphasizing thoughtful inquiry over speed and quantity of work.

Another factor that was judged to be present in many of the tasks that remained at high levels (in 71% of the tasks) was that competent performance was modeled by the teacher or by a capable student. This often came in the form of students' presenting their solutions on the overhead projector. In many of the cases where complex thinking and reasoning were occurring, such presentations modeled the use of multiple representations, meaningful exploration, and appropriate mathematical justifications; often successive presentations illustrated multiple ways of approaching a problem. In 64% of the tasks that remained at high levels, a sustained press for justifications, explanations, and meaning was evident through teacher questioning, comments, and feedback. Clear and consistent messages were sent to students that explanations and justifications were as much a part of classroom mathematical activity as were correct answers. Finally, evidence of scaffolding was found in 58% of the tasks that remained at high levels. These involved cases in which the teacher (and sometimes a more capable student) provided assistance so students could successfully complete a task; however, the assistance was not such that it took away from the overall challenge or complexity of the task. Rather, the complexity was maintained and assistance was gauged to be just enough to allow the students to maintain forward progress.

In summary, Figures 4 and 5 begin to provide a picture of the ways in which the instructional environments of project classrooms shaped the manner in which high-level tasks were implemented. The findings about the process of decline point to the importance of teachers maintaining a focus on mathematical thinking processes, all the while assisting (but not overassisting) their students. The findings about classroom supports for the maintenance of high-level activity are noteworthy from the standpoint of the number of them that appear to be operating in those tasks that were judged to truly engage students in mathematical thinking and reasoning (i.e., on average, 4 per task). In most of these tasks, the classroom environments were characterized by a wide variety of supports that enabled students to accept and take on challenges in productive ways.

Discussion

Instruction in Project Classrooms: Implications for Reform

The present study's findings strongly suggest that project teachers have been successful in selecting and setting up the kinds of mathematical tasks that have been viewed as leading to high-level student learning outcomes. The vast majority of tasks that were used in this representative sample of project classrooms possessed characteristics that set them apart from typical mathematics classrooms as described in the literature (e.g., Porter, 1989; Stodolsky, 1988). Students were more apt to be working from innovative materials and/or from teacher-developed materials than from a textbook series; they were likely to be engaged with statistics, geometry, or some other reform-inspired topic, and they were very often found to be working in pairs or groups. Moreover, the majority of the tasks that were set before them encouraged the use of multiple-solution strategies, multiple representations, and required that they explain or justify how they arrived at their answers. Finally, three quarters of the tasks were set up to demand that students engage in rather sophisticated mathematical thinking and reasoning—either connecting procedures to underlying concepts and meaning or tackling complex mathematical problems in novel ways.

With respect to task implementation, the findings have revealed that project teachers and students have also experienced fairly high levels of success in maintaining certain features of tasks that have been seen as crucial to building students' thinking and reasoning capacities. When preceded with the appropriate set up, students were found to actually use multiple-solution strategies and multiple representations and to produce explanations and mathematical justifications in the majority of cases. On the other hand, success was not as forthcoming with respect to maintaining overall high-levels of cognitive processing. While the overall consistency between task set up and task implementation was 69%, 88%, and 74% for multiple-solution strategies, multiple representations, and the production of explanations respectively, the consistency for high-level cognitive demands was 42%. Moreover, as noted in the results section, the higher the demands that a task placed on students at the task-set up phase, the less likely it was for the task to have been carried out faithfully during the implementation phase. Indeed, the kinds of tasks that scholars and reformers have suggested as most essential for building students' capacities to think and reason mathematically are the very tasks that students had the most difficulty carrying out in a consistent manner. These findings are not startling; one would expect higher level tasks to be most vulnerable to decline. The findings are noteworthy, however, because they provide an empirical basis for these expectations and assumptions. As such, the findings may form the basis for justifying staff development efforts that will help teachers to implement tasks that encourage mathematical reasoning and sense-making.

The findings related to factors associated with the decline and maintenance of high-level demands from the task-set-up phase to the task-implementation phase are ripe for further exploration. In the present article, factors associated with *all types of* decline were reported together. Other analyses have explored the relative prevalence of factors in different types of decline. For example, Henningsen and Stein (in press) examined the kinds of factors associated with three different ways in which the "doing mathematics" tasks were observed to decline (i.e., into proceduralized thinking, into no mathematical activity, and into "unstystemized exploration"). Each of these types of decline was associated with a fairly distinct profile of factors. Henningsen and Stein (in press) used these factor profiles to identify specific mathematics lessons from the project's database. These lessons were then written up as cases to illustrate prototypical types of decline; a case illustrating factors related to the maintenance of high-level activity was also identified and written up. Such cases will provide much-needed, empirically generated details regarding specific ways in which implementing instructional reform may be thwarted and specific ways in which students and teachers can be assisted to develop classroom environments in which teachers and students work together on challenging and worthwhile mathematical tasks.

Finally, follow-up studies to the present investigation have extended the shaded area of Figure 1 to include the assessment of student learning outcomes. For a description of the extent to which various levels of mathematical tasks used in the classroom appear to be associated with differential degrees of student learning, readers are referred to Stein and Lane (in press) and Stein, Lane, and Silver (1996).

Implications for Research

With respect to research issues, it seems appropriate to revisit the conceptual framework. Overall, the framework served as a useful guide for navigating through the thick mazes of substantial quantities of qualitative data. Classrooms are complex environments in which most features are deeply interrelated. The framework used in the present study simplified the environment so that the investigators could take hold of it and learn something about what was happening. In this regard, the framework served as a heuristic, not as an indelible account of classroom events. In particular, the construct of mathematical task was found to be a useful focusing device—one that served to highlight mathematical content and processes and how they were being dealt with in project classrooms. The distinction between task set up and task implementation was also found to be useful because it provided a way to separate different phases of classroom activity—phases that most experienced classroom observers have always known have the potential to vary widely. The explication of factors that influence task implementation has also been useful, although the list needs to be expanded and made more conceptually distinct, especially with respect to ways in which those factors overlap with and also remain differentiated from task implementation.

The data on which this study was based were generated as part of the documentation effort of the QUASAR Project. Although that effort collected a wide range of data, the COI write-ups (as described in the methodology section of this article) formed the heart of the project's efforts to systematically describe and examine instruction in project classrooms. Because project staff were committed to a view of teaching and learning as highly contextualized, the instrumentation for classroom observations was designed to provide rich accounts of teaching and learning. The COI database has been extremely productive for a variety of detailed analyses of instruction in project classrooms, ranging from a study of discourse patterns (e.g., Williams & Baxter, in press) to case studies of groups of teachers and their instruction during the early years of the project (e.g., Brown & Rothschild, 1993; Grover & Saulis, 1993; Smith & Seeley, 1993; Stein & Henningsen, 1993).

However, this qualitative database presented challenges to efforts aimed at producing a systematic overview of large numbers of project classrooms; words simply cannot be aggregated in the same way as numbers or check marks. The present investigation represents an initial effort to gain a big-picture view of project data. Through the use of a coding system, large amounts of qualitatively rich accounts of instruction were distilled into quantitative data. This process of distillation has enabled us to step back from the rich, but cumbersome, data and to begin to examine the types of patterns that emerged. These patterns have, in turn, provided useful, empirically generated insights regarding the successes and challenges of implementing mathematics reform in the classroom.

Notes

Preparation of this manuscript was supported by a grant from the Ford Foundation (Grant Number 890-0572) for the QUASAR Project. It is a revised version of a paper that was presented as part of a symposium, entitled "Mathematics Instructional Innovation in Urban Middle Schools: Implementation and Impact in the QUASAR Project, 1990–1993," at the Annual Meeting of the American Educational Research Association, New Orleans, 1994. Any opinions expressed herein are those of the authors and do not necessarily represent the views of the Ford Foundation. The authors would like to thank Ray St. Pierre for his help in coding the data and Gay Kowal, QUASAR data manager, for her assistance with the data management and analysis aspects of the study. The authors also acknowledge the helpful comments of Edward Silver on an earlier draft of this manuscript.

[1]It is important to recognize that current recommendations for the reform of mathematics instruction are not built on a strong base of empirical evidence linking such instruction with the desired student outcomes. The mathematics education research community has few descriptions of reform mathematics classrooms that include information on both instruction and learning outcomes (Hiebert & Wearne, 1993).

[2]QUASAR (Quantitative Understanding: Amplifying Student Achievement and Reasoning) is based at the Learning Research and Development Center at the University of Pittsburgh and is directed by Edward A. Silver.

[3]The academic task literature, to our knowledge, does not include a discussion of how to determine boundaries between tasks. From reading the work, however, we infer that researchers in this area segment the lesson into many more, smaller tasks than we do. It should be noted that our conception of mathematical task is definitely different from how the term *task* is used in the area of mathematics performance assessment. In that area, the term *task* is used to refer to a *single* mathematical problem.

[4]The instruction in the three teachers' classrooms who were selected for observation was representative of reform instruction in general at that site in the following ways: (a) only teachers of heterogeneously grouped students were selected; (b) the selected teachers were instructed to identify for observation a class that was typical with respect to the perceived ability levels of the students; and (c) each year, the majority of teachers selected for observation taught at the target grade level—i.e., the grade at which most of the project's staff development and program development efforts were being placed. Thus, during the first project year, sixth grade teachers were observed; during the second year, mostly seventh grade teachers were observed; and, during the third year, mostly eighth grade teachers were observed.

[5] The initial draft of the COI drew from two main sources: NCTM's *Professional Standards for Teaching School Mathematics* (1991) and a classroom observation system used for the state of California study of elementary mathematics (Cohen, Peterson, Wilson, Ball, Putnam, Prawat, Heaton, Remillard, & Wiemers, 1990). The COI was pilot-tested in several middle-school mathematics classrooms and underwent several rounds of critique and revision.

[6]The purpose of selecting these stratification dimensions was to maximize the representativeness and independence of the selected tasks. The nature and characteristics of the tasks and the way in which they were implemented might change from the first to the third year of the project, might differ from site to site, and from teacher to teacher. In addition, the limited human resources available to code the data required that the sample size be constrained to about one fourth the full database. Consequently, it was important that the sample include tasks that would provide equal representation on each

of the three dimensions and that approximately one fourth of the database be selected in a manner that would preserve representativeness and independence of tasks. The decision to maintain task independence was necessary, given that the overall purpose of the present study was to characterize instruction in project classrooms by using a representative sample of tasks. The authors acknowledge that, in other contexts, thinking about the ways in which tasks relate to and build on one another may be desirable.

[7]Since it was likely that there would be some coherence to the instructional activities implemented during the three consecutive observed lessons of any one teacher, coding tasks in only two of the three lessons would minimize bias.

[8]It would be likely that Tasks A and B within a given lesson would not be totally independent of each other; hence, we decided to sample only one task from each lesson.

[9]The basis for these inferences varied. For questions about multiple solutions and multiple representations, we looked for evidence across the entire class that students used a variety of solution strategies/representations or that multiple strategies/representations were publicly displayed (e.g., in successive presentations of solutions strategies at the overhead projector). For the communication and cognitive demand questions, coders were instructed to make inferences based on what the majority of students appeared to be doing. Were the majority of students producing explanations? Were the majority of students engaging with the task at a high level?

[10]By context of the task, we refer to whether an attempt was made to have the task relate to "real-life" situations (e.g., using football fields to discuss measurement concepts) or whether the task dealt solely with abstract mathematics. We are aware that the distinction between real-life and abstract contexts is complicated and is currently the object of some debate within the mathematics education community. Our distinction does not pretend to attend to some of the more subtle dimensions involved in deciding whether or not a task is real life but, rather, takes at face value whether or not real-life objects are part of the task.

[11]*Single-Symbols only* refers to representations that are composed entirely of numerals, mathematical symbols, and mathematical notation. *Single-Non-Symbolic* refers to representations that are either entirely nonsymbolic (most often manipulatives, diagrams, or pictures) or to representations that incorporate both symbols and nonsymbols.

References

Anderson, L. M. (1989). Classroom instruction. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 101–115). Washington, DC: American Association of Colleges for Teacher Education.

Bennett, N., & Deforges, C. (1988). Matching classroom tasks to students' attainments. *Elementary School Journal, 88* (3), 221–234.

Brown, C. A., & Rothschild, L. W. (1993, April). *Teachers constructing meaning in school-based mathematics reform: The case of technology.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta.

Carpenter, T., & Fennema, E. (1988). Research and cognitively guided instruction. In E. Fennema, T. P. Carpenter, & S. J. Lamon (Eds.), *Integrating research on teaching and learning mathematics* (pp. 2–19). Madison: University of Wisconsin, National Center for Research in Mathematical Sciences Education.

Cohen, D. K., Peterson, P. L., Wilson, S., Ball, D., Putnam, R., Prawat, R., Heaton, R., Remillard, J., & Wiemers, M. (1990). *Effects of state level reform of elementary-school mathematics curriculum on classroom practice* (Elementary Subject Center Series No. 25). E. Lansing: Michigan State University, Center for the Learning and Teaching of Elementary Subjects & the National Center for Research on Teacher Education.

Doyle, W. (1983). Academic work. *Review of Educational Research, 53*, 159–199.

Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist, 23* (2), 167–180.

Doyle, W., & Carter, K. (1984). Academic tasks in classrooms. *Curriculum Inquiry, 14,* 129–149.

Fitzgerald, W., Winter, M. J., Lappan, G., & Phillips, E. (1986). *Middle grades mathematics project: Factors and multiples.* Menlo Park, CA: Addison Wesley.

4th Dimension Version 3.0.5. [Computer program]. (1985–1993). Cupertino, CA: AC IUS.

Grouws, D. A. (Ed.). (1992). *Handbook of research on mathematics teaching and learning.* New York: Macmillan.

Grover, B. W., & Saulis, M. G. (1993, April). *Teachers constructing meaning in school-based mathematics reform: The case of manipulatives.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta.

Henningsen, M., & Stein, M. K. (in press). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education.*

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30,* 393–425.

Kitcher, P. (1984). *The nature of mathematical knowledge.* New York: Oxford University Press.

Lakatos, J. (1976). Proofs and refutations: *The logic of mathematical discovery.* New York: Cambridge University Press.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27,* 29–63.

Marx, R. W., & Walsh, J. (1988). Learning from academic tasks. *Elementary School Journal, 88* (3), 207–219.

Mathematical Association of America. (1991). *A call for change: Recommendations for the mathematical preparation of teachers of mathematics.* Washington, DC: Author.

Meyer, C., & Sallee, T. (1983). *Make it simpler: A practical guide to problem solving in mathematics.* Menlo Park, CA: Addison Wesley.

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 school mathematics* (Working draft). Reston, VA: Author.

National Research Council. (1989). *Everybody counts.* Washington, DC: National Academy of Sciences.

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

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

QUASAR Documentation Team. (1993, March). *The QUASAR teacher background questionnaire summary report on selected items.* Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Romberg, T. A. (1992). Perspectives on scholarship and research methods. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 49–64). New York: Macmillan.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–371). New York: Macmillan

Schroyer, J., & Fitzgerald, W. (1986). *Middle grades mathematics project: Mouse and elephant.* Menlo Park, CA: Addison Wesley.

Shavelson, R., Webb, N., & Burstein, L. (1986). Measurement of teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 50–91). New York: Macmillan.

Silver, E. A. (Ed.). (1985). *Teaching and learning mathematical problem solving: Multiple research perspectives.* Hillsdale, NJ: Erlbaum.

Silver, E. A., & Stein, M. K. (1996). The QUASAR project: The "revolution of the possible" in mathematics: instructional reform in urban middle schools. *Urban Education, 30,* 476–521.

Smith, M. S., & Seeley, M. M. (1993, April). *Teachers constructing meaning in school-based mathematics reform: The case of problem solving.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta.

Stein, M. K., & Baxter, J. A. (1989). *Teachers' use of textbooks.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Stein, M. K., & Henningsen, M. (1993, April). *Teachers constructing meaning in school-based mathematics reform: The case of cooperative learning groups.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta.

Stein, M. K., & Lane, S. (in press). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation.*

Stein, M. K., Lane, S., & Silver, E. A. (1996, April). *Classrooms in which students successfully acquire mathematical proficiency: What are the critical features of teachers' instructional practice?* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Stodolsky, S. (1988). *The subject matters: Classroom activity in mathematics and social studies.* Chicago: University of Chicago Press.

Systat 5 for the Macintosh. [Computer program]. (1989). Evanston, IL: Systat.

Williams, S., & Baxter, J. A. (in press). Reconstructing constructivism. *Elementary School Journal.*

Wittrock, M. C. (1986). Students' thought processes. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 297–314). New York: Macmillan.

Manuscript received August 10, 1994
Accepted February 3, 1995