NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

Transforming Secondary Mathematics Teaching: Increasing the Cognitive Demands of Instructional Tasks Used in Teachers' Classrooms
Author(s): Melissa D. Boston and Margaret S. Smith
Source: *Journal for Research in Mathematics Education*, Vol. 40, No. 2 (Mar., 2009), pp. 119-156
Published by: National Council of Teachers of Mathematics
Stable URL: http://www.jstor.org/stable/40539329
Accessed: 02-04-2015 14:03 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

National Council of Teachers of Mathematics is collaborating with JSTOR to digitize, preserve and extend access to *Journal for Research in Mathematics Education*.

http://www.jstor.org

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Journal for Research in Mathematics Education
2009, Vol. 40, No. 2, 119–156

Transforming Secondary Mathematics Teaching: Increasing the Cognitive Demands of Instructional Tasks Used in Teachers' Classrooms

Melissa D. Boston
Duquesne University

Margaret S. Smith
University of Pittsburgh

Mathematics teachers' selection and implementation of instructional tasks were analyzed before, during, and after their participation in a professional development initiative that focused on selecting and enacting cognitively challenging mathematical tasks. Data collected from 18 secondary mathematics teacher participants included tasks and student work from teachers' classrooms, lesson observations, and interviews. Ten secondary mathematics teachers who did not participate in the professional development initiative served as the contrast group and participated in 1 lesson observation each. Analysis of the data indicated that, following their participation in the professional development initiative, project teachers more frequently selected high-level tasks as the main instructional tasks in their classrooms and had improved the maintenance of high-level cognitive demands. Significant differences existed between project teachers and the contrast group in task selection and implementation. These differences were not influenced by the use of Standards-based or conventional curricula in project teachers' classrooms.

Key words: Higher order thinking; Professional development; Secondary mathematics; Teacher knowledge; Teaching practice

Raising the overall level of mathematical proficiency of students in the United States can be seen "as both a matter of national interest and a moral imperative" (RAND, 2003, p. xi). Recent international comparisons make salient that U.S. students do not perform at the level of U.S. economic competitors (Lemke et al., 2004; Mullis, Martin, Gonzalez, & Chrostowski, 2004). A recent report from the American Institutes for Research (Ginsburg, Cooke, Leinwand, & Pollock, 2005),

This article was prepared under the auspices of the ESP Project, under a grant from the National Science Foundation (DUE-0301962), and is based on dissertation research conducted at the University of Pittsburgh by Melissa D. Boston under the directorship of Margaret S. Smith. Any opinions expressed herein are those of the authors and do not necessarily represent the views of the Foundation. An earlier version of this article was presented at the 2007 Research Pre-Session of the Annual Meeting of the National Council of Teachers of Mathematics.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

in which the performance of students in 12 countries who participated in both Trends in International Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA) in 2003 was analyzed, suggests that students in the United States need to improve their ability to use cognitively demanding skills (e.g., reasoning) and apply facts and procedures in order to achieve at the level of higher performing countries.

Although engaging U.S. students in more cognitively demanding activity may well be the key to improved performance, the recent TIMMS video study (NCES, 2003; Stigler & Hiebert, 2004) highlights the difficulty that U.S. teachers have in maintaining the rigor of a task during instruction. In that study, a random sample of 100 eighth-grade mathematics classes from each of seven countries was videotaped during the 1999 school year. Although 17% of the tasks used by teachers in the United States were coded as high level (making connections), *none* of these tasks was implemented as intended. Instead, most of the making-connections problems were transformed into procedural exercises. The authors concluded that U.S. eighth-grade students spend most of their time in mathematics classrooms practicing procedures, regardless of the nature of the tasks they are given.

It is clear that providing students with more rigorous classroom instruction will require improvements in current teaching practices. And such improvements will depend on providing teachers with professional development opportunities that are *transformative* in nature— catalyzing changes in teachers' long-held, underlying beliefs about effective teaching and learning of mathematics (Thompson & Zeuli, 1999). Specifically, teachers will need opportunities to thoroughly overhaul their thinking about what it means to know and understand mathematics, the kinds of mathematical tasks in which their students should engage, and how they can supports their students' learning without taking over students' opportunities for high-level thinking and reasoning.

The purpose of the study reported herein was to determine the impact of a professional development program, specifically focused on the selection and implementation of cognitively challenging tasks, on the instructional practices of the secondary mathematics teacher participants. The research questions that framed the analysis in this investigation are the following:

1. Does teachers' use of mathematical tasks with high-level cognitive demands change during and following teachers' participation in professional development experiences specifically focused on the selection and implementation of cognitively challenging mathematical tasks?
2. Does teachers' implementation of mathematical tasks change during and following teachers' participation in professional development experiences specifically focused on the selection and implementation of cognitively challenging mathematical tasks?
3. Does the type of curriculum used in a teacher's classroom (Standards-based vs. conventional) influence the teacher's selection and/or implementation of cognitively challenging instructional tasks?

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Based on classroom artifacts (i.e., instructional tasks and student work) and observations, changes in the selection and implementation of cognitively challenging instructional tasks in teachers' classrooms were considered evidence of instructional change. We begin with a review of literature in two areas that are central to this study: mathematical tasks and professional development.

BACKGROUND

The professional development experiences that collectively form the "treatment" in this study specifically focused on the selection and implementation of cognitively challenging mathematical tasks. In this section, we establish the importance of cognitively challenging tasks to the teaching and learning of mathematics. We then utilize successful professional development studies for teachers of mathematics to describe the type of professional development experiences that can catalyze change in teachers' classroom practices.

The Importance of Cognitively Challenging Mathematical Tasks

A *mathematical task* is defined as a set of problems or a single complex problem that focuses students' attention on a particular mathematical idea (Stein, Grover, & Henningsen, 1996). The construct of task also encompasses the intellectual and physical products that are expected of students, the operations that students are to use to obtain the desired products, and the resources that are available to assist students in completing these products (Doyle, 1983). On a practical level, tasks influence student learning because working on mathematical tasks constitutes what students *do* during the majority of their time in mathematics class. Students in all seven countries analyzed in the TIMSS Video Study (NCES, 2003) (i.e., Australia, Czech Republic, Hong Kong, Japan, Netherlands, Switzerland, and the United States) spent over 80% of their time in mathematics class working on mathematical tasks.

On a theoretical level, Doyle offered two premises for why "tasks form the basic treatment unit in classrooms" (1983, p. 162). First, a mathematical task draws students' attention toward a particular mathematical concept and provides certain information surrounding that concept (Doyle, 1983). Students are exposed to (and thus have an opportunity to learn) the concepts embedded in the tasks they complete. Students are not exposed to (and thus have much less of an opportunity to learn) content that is not represented in the tasks they complete. Second, tasks influence student learning by setting parameters for the ways in which information about the mathematical concept can be operated on or processed (Doyle, 1983). Students will become skilled at what they have an opportunity to actually do in mathematics class. According to Doyle (1988), "the work students do, defined in large measure by the tasks teachers assign, determines how they think about a curricular domain and come to understand its meaning" (p.167). Hence, different kinds of tasks lead to different types of instruction, which subsequently lead to different opportunities for students' learning (Doyle, 1988; Stein, Smith, Henningsen, & Silver, 2000).

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Stein and her colleagues differentiated mathematical tasks based on the level of cognitive demand of the task (i.e., the level and type of thinking that a task has the potential to elicit) (Stein et al., 1996). The Task Analysis Guide (Figure 1) identifies two categories of mathematical tasks with high-level cognitive demands ("doing mathematics" and "procedures with connections") and two categories of

Low-Level Cognitive Demands	High-Level Cognitive Demands
Memorization Tasks • Involve either producing previously learned facts, rules, formulae, or definitions *or* committing facts, rules, formulae, or definitions to memory. • Cannot be solved using procedures because a procedure does not exist or because the time frame in which the task is being completed is too short to use a procedure. • Are not ambiguous—such tasks involve exact reproduction of previously seen material and what is to be reproduced is clearly and directly stated. • Have no connection to the concepts or meaning that underlay the facts, rules, formulae, or definitions being learned or reproduced. *Procedures Without Connections Tasks* • Are algorithmic. Use of the procedure is either specifically called for or its use is evident based on prior instruction, experience, or placement of the task. • Require limited cognitive demand for successful completion. There is little ambiguity about what needs to be done and how to do it. • Have no connection to the concepts or meaning that underlie the procedure being used. • Are focused on producing correct answers rather than developing mathematical understanding. • Require no explanations or explanations that focus solely on describing the procedure that was used.	*Procedures With Connections Tasks* • Focus students' attention on the use of procedures for the purpose of developing deeper levels of understanding of mathematical concepts and ideas. • Suggest pathways to follow (explicitly or implicitly) that are broad general procedures that have close connections to underlying conceptual ideas as opposed to narrow algorithms that are opaque with respect to underlying concepts. • Usually are represented in multiple ways (e.g., visual diagrams, manipulatives, symbols, problem situations). Making connections among multiple representations helps to develop meaning. • Require some degree of cognitive effort. Although general procedures may be followed, they cannot be followed mindlessly. Students need to engage with the conceptual ideas that underlie the procedures in order to successfully complete the task and develop understanding. *Doing Mathematics Tasks* • Require complex and non-algorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example). • Require students to explore and to understand the nature of mathematical concepts, processes, or relationships. • Demand self-monitoring or self-regulation of one's own cognitive processes. • Require students to access relevant knowledge in working through the task. • Require students to analyze the task and actively examine task constraints that may limit possible solution strategies and solutions. • Require considerable cognitive effort and may involve some level of anxiety for the student due to the unpredictable nature of the solution process required.

Figure 1. The Task Analysis Guide (Stein et al., 2000).

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

tasks with low-level cognitive demands ("procedures without connections" and "memorization"). Research has shown, however, that teachers typically do not analyze tasks in terms of cognitive demands; that is, in terms of the type or level of thinking that the task can elicit from students (Arbaugh & Brown, 2002; Stein, Baxter, & Leinhardt, 1990). Studies conducted by Stein et al. (1990) and Arbaugh and Brown (2002) found that teachers tend to categorize tasks with respect to similarities in mathematical content or surface-level features, such as "word problems" or "uses a graph." Research consistently indicates that teachers' selection of instructional tasks is largely based on lists of skills and concepts they need to cover (Hiebert et al., 1997) or by adhering to the tasks in their textbooks (Doyle, 1983; Remillard, 1999). In fact, recent national studies indicate that curricular materials or textbooks provide the main source of tasks in teachers' classrooms (Grouws, Smith, & Sztajn, 2004; NCES, 2004). Since curricular materials vary widely with respect to the nature of the mathematical tasks they contain, teachers need to be aware of how different types of tasks influence students' opportunities for learning. If the role of teachers is to facilitate conceptual understanding, then the first step in the process is for the teacher to select cognitively challenging mathematical tasks as the basis of instruction in their classrooms.

The Importance of Task Implementation

Selecting high-level tasks for use in mathematics classrooms, however, does not guarantee that students will have opportunities to engage in high-level thinking. Research by Stein and colleagues determined that the cognitive demands of mathematical tasks can change as tasks are introduced to students or as tasks are enacted during instruction (Stein et al., 1996). The Mathematical Tasks Framework (MTF) (Figure 2) models the progression of a mathematical task from its original form

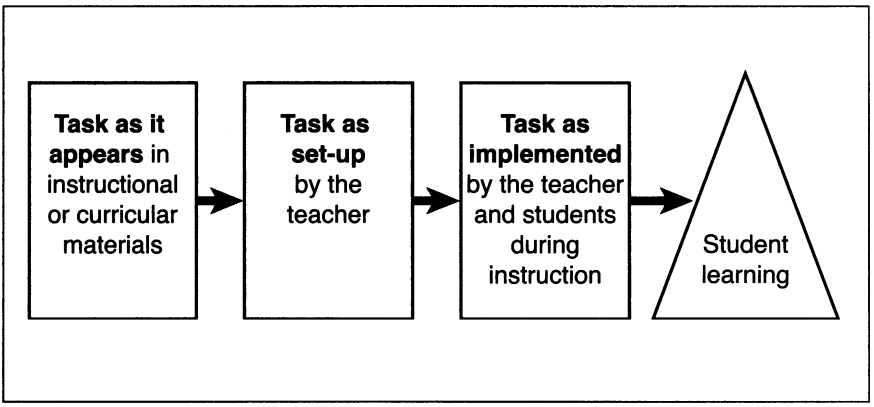


Figure 2. The Mathematical Tasks Framework (Stein et al., 2000).

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

throughout an instructional episode and identifies phases at which the cognitive demands of the task are likely to be altered.

Implementing cognitively challenging tasks in ways that maintain students' opportunities to engage in high-level cognitive processes is not a trivial endeavor (Henningsen & Stein, 1997; NCES, 2003; Stein & Lane, 1996). Teachers can lower the demands of a task by breaking it down into smaller subtasks (Smith, 2000), by focusing on correct answers and procedures (Doyle, 1988; Henningsen & Stein, 1997; Romanagno, 1994), or by adapting the tasks or teaching suggestions to be consistent with their personal notions of effective teaching and learning (Arbaugh, Lannin, Jones, Park-Rogers, 2006; Clarke, 1997; Lloyd & Wilson, 1998; Remillard, 1999). Conversely, teachers' "normative practices" in the classroom (Carpenter & Lehrer, 1999, p. 26), such as the teachers' use of questioning (Boaler & Staples, 2008) or cooperative learning groups (Schoen, Cebulla, Finn, & Fi, 2003), can maintain students' opportunities to engage in high-level thinking and reasoning. From their research, Stein and colleagues have identified specific classroom-based factors that serve to maintain or reduce the cognitive demands of mathematical tasks during implementation, that is, as tasks are enacted by the teacher and students during an instructional episode. These factors are presented in Figure 3.

Consistent results from over a decade of research indicate that the greatest student learning gains occur in classrooms in which the high-level cognitive demands of mathematical tasks are consistently maintained throughout the instructional episode (Boaler & Staples, 2008; Hiebert & Wearne, 1993; Stigler & Hiebert, 2004; Stein & Lane, 1996; Tarr et al., 2008). For example, in a longitudinal comparison of three high schools over a 5-year period, Boaler and Staples (2008) determined that the highest student achievement occurred at the school in which students were supported to engage in high-level thinking and reasoning. Boaler and Staples attributed students' success to the teachers' ability to maintain high-level cognitive demands during instruction, specifically, to the teachers' use of preplanned questions that elicited and supported students' thinking. Tarr and colleagues (Tarr et al., 2008) and Stein and Lane (1996) have both determined that learning environments in which teachers encourage multiple strategies and ways of thinking and support students to make conjectures and explain their reasoning were associated with higher student performance on measures of thinking, reasoning, and problem solving.

Results from the TIMSS 1999 Video Study confirm the connection between higher student achievement and students' engagement with high-level cognitive demands (Stigler & Hiebert, 2004). Teachers in countries whose students outperformed U.S students implemented cognitively challenging tasks in ways that maintained students' opportunities to engage in high-level thinking and reasoning (Stigler & Hiebert, 2004). Although teachers in the United States used percentages of high-level tasks consistent with the percentages of high-level tasks used in most higher performing countries, the most striking and significant difference between the United States and higher performing countries in the study was the inability of

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Factors Associated With the Decline of High-Level Cognitive Demands	Factors Associated With the Maintenance of High-Level Cognitive Demands
1. Problematic aspects of the task become routinized (e.g., students press the teacher to reduce the complexity of the task by specifying explicit procedures or steps to perform; the teacher "takes over" the thinking and reasoning and tells students how to do the problem).	1. Scaffolding of students' thinking and reasoning.
2. The teacher shifts the emphasis from meaning, concepts, or understanding to the correctness or completeness of the answer.	2. Students are provided with means of monitoring their own progress.
3. Not enough time is provided to wrestle with the demanding aspects of the task or too much time is allowed and students drift into off-task behavior.	3. Teacher or capable students model high-level performance.
4. Classroom management problems prevent sustained engagement in high-level cognitive activities.	4. Sustained press for justifications, explanations, and/or meaning through teacher questioning, comments, and/or feedback.
5. Inappropriateness of tasks for a given group of students (e.g., students do not engage in high-level cognitive activities due to lack of interest, motivation or prior knowledge needed to perform; task expectations not clear enough to put students in the right cognitive space).	5. Tasks build on students' prior knowledge.
6. Students are not held accountable for high-level products or processes (e.g., although asked to explain their thinking, unclear or incorrect student explanations are accepted; students are given the impression that their work will not "count" toward a grade).	6. Teacher draws frequent conceptual connections.
	7. Sufficient time to explore (not too little, not too much).

Figure 3. Factors associated with the decline and maintenance of high-level cognitive demands (Stein, Grover, & Henningsen, 1996).

U.S. teachers to maintain the high-level cognitive demands of mathematical tasks during instruction. Similarly, in the sample of mathematics classrooms analyzed by Horizon Research (Weiss & Pasley, 2004; Weiss, Pasley, Smith, Banilower, & Heck, 2003), only 15% of observed lessons were classified as providing opportunities for thinking, reasoning, and sense-making.

Improving students' opportunities to learn mathematics with understanding will require sustained opportunities for students to engage with cognitively challenging mathematical tasks. To provide such opportunities, mathematics teachers will need to (a) select high-level tasks for instruction and (b) implement high-level tasks in

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

ways that maintain the cognitive demands. Hence, increasing students' exposure to and sustained engagement with high-level tasks will require changes in the knowledge and instructional practices of mathematics teachers. Can learning experiences for mathematics teachers transform their selection and implementation of instructional tasks? What would such learning experiences look like? These questions will be explored in the section that follows.

Transformative Professional Development for Teachers of Mathematics

Professional development studies conducted over the past 20 years have documented that teachers' knowledge, beliefs, and instructional practices can be transformed by challenging teachers' conceptions of mathematics (e.g., Simon & Shifter, 1991; Swafford, Jones, & Thornton, 1997), by introducing teachers to students' ways of thinking and learning (e.g., Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Cobb, Yackel, & Wood, 1991; Kazemi & Franke, 2004), by engaging teachers in analyzing inquiry-based, reform-oriented mathematics instruction (e.g., Barnett, 1998; Frykholm, 2005; Wallen & Williams, 2000), by assisting teachers' implementation of reform-oriented mathematics pedagogy and/or curricula (e.g., Arbaugh et al., 2006; Remillard, 1999; Smith, 2000), or by organizing teachers into communities of learners that work together to improve their practice (Arbaugh, 2003; Fernandez, 2005; Stein, Silver, & Smith, 1998). All these studies have contributed to our growing understanding of how to design, conduct, and study professional development and its impact on teachers' knowledge, beliefs, and habits of practice. One of the most influential professional development efforts of the last 2 decades is the work of the Cognitively Guided Instruction (CGI) Project. CGI is based on the belief that instruction should facilitate children's construction of knowledge (Carpenter et al., 1989). CGI researchers designed a professional development workshop that engaged teachers in analyzing children's thinking and provided teachers with research findings on children's strategies for solving addition and subtraction story problems. The researchers contended that these experiences would enable teachers to employ new instructional practices that built on children's existing knowledge, to change their beliefs about the nature of mathematics teaching and the roles of the teacher and student in that process, and to become more aware of the mathematical thinking of students in their classrooms.

In order to test the validity of these assertions, 40 first-grade teachers (with an average of 11 years of teaching experience) were recruited to participate as members of either treatment or control groups (Fennema et al., 1996). The 20 teachers in the treatment group attended an intensive 4-week summer workshop designed to share the findings of research on children's development of addition and subtraction concepts and to provide teachers with an opportunity to think about and to use this research-based knowledge in planning for instruction. Teachers and researchers discussed broad principles that should guide instructional decision-making and how different curricular materials could be adapted for use in light of

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

these principles. In addition, teachers were shown videotapes of children's problem solving and had the opportunity to interview children. The teachers in the control group, by contrast, attended two 2-hour workshops on problem solving. Over the course of the school year following the summer workshop, data were collected to provide information on classroom practices, teachers' knowledge of students as learners, teacher beliefs, student achievement, and student confidence and beliefs. Overall, the findings from the study provide evidence that exposing teachers to research on children's thinking influenced the teachers' knowledge of children's development of addition and subtraction strategies, their beliefs about teaching and learning addition and subtraction, and their instructional practices in ways consistent with the principles of CGI. Most important, the changes in instructional practices of CGI teachers were related to improvements in student achievement.

Several features of the CGI professional development effort and associated research initiative are particularly noteworthy. First, the use of a framework to guide the design of the professional development and the enactment of new instructional practices by teachers in their own classrooms provided a focus for teachers' efforts at reforming or enhancing their instructional practice (Ball & Cohen, 1999). Second, encouraging teachers to flexibly apply general principles in designing instruction for their students made it possible for teachers to adapt and use materials already available in their local settings. Third, the CGI researchers situated professional development in the everyday work of teaching (i.e., adapting curricular materials, analyzing student thinking, conducting interviews). In a seminal paper on this topic, Ball and Cohen (1999) argued that "teachers' everyday work could become a source for constructive professional development" (p. 6) through the development of a curriculum for professional learning that is grounded in the tasks, questions, and problems of practice.

Several other professional development initiatives have utilized this "practice-based" approach, using materials taken from real classrooms to provide opportunities for critique, inquiry, and investigation into the work of teaching. For example, Barnett (1998) and Wallen and Williams (2000) reported that engaging teachers in the analysis and discussion of instructional cases of mathematics teaching and learning helped teachers to identify and begin to solve problems they had been wrestling with in their own classrooms. Other researchers such as Little (2004), Kazemi and Franke (2004), and Jacobs and Philipp (2004) have used student-work examples to help teachers understand children's thinking and use it as the basis for planning subsequent instruction.

The professional development intervention that is foundational to the current investigation shares several features with the studies discussed in this section. Specifically, a research framework (i.e., the MTF in Figure 2) was used to guide the design of the professional learning experiences and to support the development of teachers' instructional practices. Learning experiences for teachers were "authentic" to the daily work of teaching, with opportunities to analyze instructional episodes presented in narrative and video form and to analyze written work produced by students. Finally, the framework and tools served to provide guiding principles

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

rather than prescriptions for designing and enacting instruction in teachers' own classrooms. Additional information about the professional development intervention is provided in the next section.

METHODS

The intent of this investigation was to determine whether mathematics teachers' selection and implementation of cognitively challenging mathematical tasks increased throughout their participation in a professional development workshop. The analysis of teachers' instructional practices assessed the specific focus and intended outcome of the professional development experiences—improving teachers' selection and implementation of cognitively challenging mathematical tasks in their own classrooms. The study thus maintained a clear focus on the cognitive demands of mathematical tasks in the research questions, the professional development intervention, and the collection and analysis of data. In this way, the research design provides credible explanations and inferences to appropriately answer the research questions (Janesick, 2000). To set the context for the study, we begin with a description of the professional development intervention.

Context and Content of the Professional Development Intervention

This investigation focuses on secondary mathematics teachers participating in the NSF-funded project titled Enhancing Secondary Mathematics Teacher Preparation (ESP). The goal of the ESP project was to improve the quality of field experiences for preservice mathematics teachers by ensuring that the classrooms of their mentor teachers provided a "reinforcing culture" of quality mathematics instruction (Sykes & Byrd, 1992). According to Sykes and Byrd, "If the aim of teacher education is a reformed practice that is not readily available, and if there is no reinforcing culture to support such practice, then the basic imagery of apprenticeship seems to break down" (p. 50). Toward this end, mentor teachers participated in 2 years of ESP professional development workshops designed to support improvements in their instructional practices (6 full days in year 1); develop their capacity to mentor beginning teachers (1 week at the end of year 1); and help mentor teachers and the preservice teachers assigned to their classrooms develop a shared vision of effective mathematics teaching (5 half days in year 2). The study described herein focused on a cohort of 18 mentor teachers who participated in the ESP project from September 2004 through March 2006.

The team who designed and conducted the ESP professional development workshops consisted of the second author (who was the principal investigator of the project) and three graduate students (including the first author) who collectively had a range of experiences teaching mathematics in Grades 7–12 environments and working with preservice and practicing secondary mathematics teachers. The research component of the project was directed by the first author, and another graduate student on the team assisted with data collection.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

The MTF (Figure 2), developed by QUASAR researchers (Stein et al., 1996), served as the underlying theoretical framework that guided the design and selection of the professional development activities. Other tools that were central to the professional learning experiences were the Task Analysis Guide (Figure 1), the factors that influence the cognitive demands of mathematical task (Figure 3), and a lesson-planning tool designed to support teachers' implementation of cognitively challenging tasks (Smith, Bill, & Hughes, 2008). *Implementing Standards-Based Instruction: A Casebook for Professional Development* (Stein et al., 2000) served as the "textbook" for the professional development sessions. This book provides a summary of the QUASAR research and includes a set of narrative cases that highlight the empirical patterns of task set-up and implementation that emerged from that research.

At the heart of the ESP professional development sessions were ongoing opportunities for teachers to solve mathematical tasks, to assess the cognitive demands of mathematical tasks, and to analyze the implementation of mathematical tasks during instructional episodes. In general, activities used within the professional development sessions were selected and designed to represent "samples of authentic practice" (Smith, 2001, p. 7) that focused teachers' learning on the cognitive demands of mathematical tasks at each phase of the MTF. These opportunities were intended to influence teachers' ideas of how students learn mathematics and, in turn, of how they should endeavor to teach mathematics. General theories or "lessons learned" that could serve to catalyze changes in teachers' own instructional practices were abstracted from the examination of specific classroom artifacts (e.g., student work, narrative, and video cases of instruction).

To encourage the application of lessons learned in professional development to teachers' own classroom practices, teachers engaged in scaffolded field experiences (SFEs) (Borasi & Fonzi, 2002). The first of these SFEs engaged teachers in selecting a task from their curricular materials and analyzing the cognitive demands of the task using the Task Analysis Guide (Figure 1). Over time, the SFEs expanded to include the full cycle of teaching activities (e.g., planning, teaching, and reflecting on a lesson based on a high-level task). Teachers had opportunities at each professional development session to share the results of the SFE with their peers and receive feedback (see Hughes, Smith, Boston, & Hogel, in press, for additional details on this process).

Consistent with recommendations by researchers and teacher educators (e.g., Ball & Cohen, 1999; Borko & Putnam, 1995; NCTM, 2000; Simon & Schifter, 1991), the ESP facilitators modeled the type of instructional strategies that they intended for project teachers to begin to incorporate into their own classrooms. Based on a social-constructivist view of teacher learning (Borko & Putnam, 1995; Simon & Schifter, 1991), these instructional strategies provided a supportive and collaborative environment in which project teachers were challenged to wrestle with new ideas, to openly express disagreements, and to identify aspects of their own instructional practices that they would like to change. Three components of social constructivism that have significant implications for teachers' learning were foundational

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

to this professional development experience: (a) the importance of building professional development experiences on teachers' prior knowledge and beliefs, (b) the assertion that change occurs as new conceptions of mathematics teaching and learning conflict with the teachers' prior knowledge and beliefs, and (c) the role of social interaction in stimulating and maintaining this type of conflict.

Subjects

The 19 secondary mathematics teachers participating in the ESP Project during the 2004–2005 school year were invited to be the subjects in this study. Eighteen teachers agreed to provide classroom artifacts (instructional tasks and student work), and 11 of these teachers also elected to participate in classroom observations. The teachers were provided with a stipend for participating in the ESP professional development workshop (during the 2004–2005 school year) and a leadership and mentoring workshop (during June 2005); receipt of the stipend was not contingent upon teachers' participation in the study.

A contrast group was selected to provide a means of assessing whether the knowledge and instructional practices of ESP teachers differed from the knowledge and instructional practices of teachers who did not participate in the ESP professional development workshop. Two school districts in the region (but not participating in ESP) agreed to recruit teachers to participate in the contrast group, and all mathematics teachers in the two schools were given the option of participating. District administrators (i.e., the superintendent, curriculum director, and building principals) were provided with an overview of the study, the goals of the ESP professional development workshop, and a listing of the data to be collected from contrast group teachers. The purpose of the teachers' participation in the study was described to district administrators and to teachers as helping to determine, at the end of the school year, if teachers who had participated in a yearlong professional development workshop with a very specific focus had different knowledge or instructional practices (related to the specific focus) than teachers who had not participated in the professional development workshop. A total of 10 teachers (5 from each school) agreed to serve as contrast subjects.

The data in Table 1 enable comparisons between the ESP teachers and the contrast group with respect to years of teaching experience, level of their teaching assignment (middle school or high school), and possession of a secondary mathematics certification. The one ESP teacher identified as not having a secondary mathematics certification was originally certified as an elementary teacher and attained middle school mathematics certification by successfully completing the Praxis exam. Demographics of the schools in which ESP teachers worked ranged from a large, urban, economically disadvantaged school district to a midsized affluent suburban district to several small middle-class suburban districts. The teachers' professional development opportunities (outside of the ESP project) varied greatly, as did exposure to and use of *Standards*-based mathematics curricula and ways of teaching mathematics. The two school districts that employed the contrast group

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

teachers were located in suburban areas consisting of middle-class to upper-middle-class communities. One school was implementing *Standards*-based curricula in both middle school and high school, and teachers were participating in a large-scale professional development project in the region. The other school did not use *Standards*-based curricula and did not offer quality professional development opportunities for mathematics teachers. Hence, the schools and teachers in the contrast group reflected the diversity of the ESP group along several dimensions: years of teaching experience, teaching level (middle school and high school), school demographics, professional development opportunities, and mathematics curricula (*Standards*-based and conventional).

Table 1
Characteristics of ESP Teachers and Contrast Group Teachers

	Years of experience					Grade level		Secondary mathematics certification	
Teachers	Range	Average	< 10	10–15	> 16	Middle school	High school	Yes	No
ESP	3–30	8.5	12	4	2	10	8	17	1
Contrast group	2–31	11.8	4	3	3	6	4	10	0

Data Sources

As part of the effort to document ESP activities and their potential impact on teacher participants, data were collected on teachers' experiences in the ESP workshop, teachers' knowledge, and teachers' instructional practices. The analyses reported herein utilize only the data on teachers' instructional practices (see Boston, 2006, for an analysis of teachers' experiences in the ESP workshop and of changes in teachers' knowledge). The data on ESP teachers' instructional practices (i.e., instructional tasks, student work, and classroom observations) were collected at three points during the 2004–2005 school year: an early data collection in the Fall (October, 2004), a middle data collection in the Winter (January, 2005), and a final data collection in the Spring (May, 2005). Data from the contrast group were collected in the Spring (April–May, 2005). The paragraphs that follow describe the sources of data on teachers' instructional practices used in this investigation: instructional tasks, student work, and lesson observations.

Instructional Tasks

For the collection of instructional tasks, ESP teachers submitted all instructional tasks used over 5 consecutive days of instruction for one of their courses. Teachers

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

completed a daily log sheet that indicated how each task was used in the lesson (i.e., warm-up, review, main instructional task, practice, homework, etc.). Only the main instructional task in each of the five lessons was scored for analysis. For the majority of lessons, the main instructional task could be clearly identified by the following criteria: (a) It was identified by the teacher as the main instructional task, (b) it fit coherently into the sequence of learning tasks and mathematical ideas being developed over the 5 days of instruction, and (c) it was reported to have taken up the largest amount of instructional time in the lesson. If two tasks equally satisfied all three criteria, the scores for the two tasks were averaged into one task score for the lesson. The ESP project team did not advise teachers regarding the selection or the source of their instructional tasks; teachers simply submitted the entire set of tasks they had used during 5 consecutive days of instruction, independently selected in their own manner and from their own sources.

Student Work

Teachers submitted class sets of students' written work for three of the instructional tasks used within each of the 5-day data collection periods. Research conducted by Matsumura and colleagues at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) (Clare, 2000; Clare & Aschbacher, 2001; Matsumura, 2003; Matsumura et al., 2002) determined that collections of student work can provide stable indicators of classroom practice. Although the CRESST studies focused on reading comprehension, recent research by the Instructional Quality Assessment team (Boston & Wolf, 2006; Matsumura et al., 2006) suggests that students' work on mathematics assignments tends to be even more stable within teachers and highly correlated with observed instruction. Teachers completed a cover sheet for each class set of student work, in which they described the directions given to students to complete the task, how students worked on the task (i.e., individual work, group work, homework), and how students' work was assessed.

Lesson Observations

ESP teachers were observed once during each 5-day data collection period. Lesson observations were conducted by the first author and by another graduate student on the ESP team. Lesson observations lasted an entire class period, with the researcher observing, taking field notes, and collecting relevant classroom artifacts (i.e., instructional tasks or handouts).

Data From the Contrast Group

In the spring of 2005, teachers in the contrast group completed an assessment of their knowledge of the cognitive demands of instructional tasks, participated in one lesson observation, and provided copies of the instructional tasks used during the lesson. Data from teachers in the contrast group were collected and coded by the

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

first author. To ensure reliability, the tasks used by contrast group teachers during the lesson observations were mixed with the tasks of the ESP teachers for blind coding by a second rater. Teachers in the contrast group did not participate in professional development activities related to the ESP project. Data from the contrast group were intended to provide a snapshot of the knowledge[1] and instructional practices of secondary mathematics teachers who were similar demographically to the ESP group but had not experienced the ESP professional development workshop.

Instrumentation and Coding

Data on teachers' classroom practices were scored using the Instructional Quality Assessment (IQA) Academic Rigor (AR) in Mathematics rubrics for *Potential of the Task* (Appendix A) and *Overall Implementation* (Appendix B). These rubrics have undergone reliability and validity testing by the IQA team (Boston & Wolf, 2006; Matsumura, Slater, Garnier, & Boston, 2008). The Potential of the Task rubric assesses the level of cognitive demand necessary for students to produce the best possible response to the task, and the Overall Implementation rubric assesses the level of cognitive demand of students' actual engagement with or responses to the task. Stein and colleagues' work on cognitively challenging tasks and task implementation (e.g., Stein, Grover, & Henningsen, 1996; Henningsen & Stein, 1997; Stein et al., 2000) formed the basis of the score levels for each rubric. Score levels 1 and 2 represent low-level cognitive demands as described in the Task Analysis Guide (Figure 1). Score levels 3 and 4 represent high-level cognitive demands in which the connections to meaning and understanding are implicitly (score level 3) or explicitly (score level 4) required by the task. A score of 0 indicates that the lesson tasks or activities were nonmathematical in nature (i.e., typing a report, looking for information on the Internet).

The five main instructional tasks in an ESP teacher's data collection were scored on a scale of 0–4 using the Potential of the Task (Appendix A) dimension of the IQA AR-Math rubrics. A teacher's five task scores were then averaged to provide the teacher's mean task score for that data collection period. Similarly, each class set of student work received a score of 0–4 for the Potential of the Task, and these three scores were averaged to provide a mean student-work Potential of the Task score for that data collection period. Each class set of student work also received a score for the cognitive processes evident in the students' written work for solving the task, using the Overall Implementation rubric (Appendix B). Class sets of student work were scored for Overall Implementation based on the highest level of cognitive demand exhibited by the majority of student-work samples within the set. A teacher's three student-work Implementation scores were then averaged to provide a mean student-work Implementation score for that data collection period. Comparing a teacher's mean scores for the student-work Potential of the Task and Implementation

[1]See Boston (2006) for comparisons of ESP and contrast-group teachers' knowledge of the cognitive demands of mathematical tasks.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

of the Task provides an indication of whether the high-level cognitive demands were maintained during implementation.

The lesson observations were scored for Potential of the Task and for Overall Implementation. The Implementation score was based on the highest level of cognitive demand exhibited as students engaged with the main instructional task throughout the lesson. The IQA Lesson Checklist (Appendix C), derived from the classroom-based factors presented in Figure 3 (Stein, Grover, & Henningsen, 1996), was utilized to guide the scoring of the lesson observations. The Lesson Checklist also served as a tool to identify the factors that characterized instruction in the observed lesson and to compare the nature of instruction in teachers' classrooms across data collection periods.

All data on teachers' instructional practices were collected and scored by the first author. A blind, stratified random sample of the instructional tasks and student work was scored by a second rater (not associated with the ESP project or teachers) to determine reliability. A subset of classroom observations was conducted and coded by a second rater as well. Interrater agreement for instructional tasks was 83.3%; student-work tasks, 86.7%; student-work implementation, 93.3%; and lesson observations (tasks and implementation), 100%.

The officially adopted mathematics curricula for the courses from which teachers were submitting tasks and student work were coded as "*Standards*-based" or "conventional." Curricula coded as "*Standards*-based" were identified as "exemplary or promising curricula" by the U.S. Department of Education (USDE, 1999), were rated highly in the review of mathematics curricula conducted by the American Association for the Advancement of Science (AAAS, 2000), or were developed under the auspices of an NSF grant. This coding enabled a comparison of the level of tasks and student work submitted by teachers with officially adopted *Standards*-based or conventional curricula. In other words, this comparison assessed whether the availability of a certain type of curricular resource influenced the level of tasks used to engage students in learning mathematics.

In summary, each ESP teacher received (a) a mean score for the Potential of the set of five instructional tasks, (b) a mean score for the Potential and for the Implementation of their three student-work tasks, (c) a score for the Potential and for the Implementation of the task used during the lesson observation, and (d) a Lesson Checklist identifying the factors influencing the task Implementation score for the lesson observation. Contrast group teachers received a score for the Potential and a score for the Implementation of the task used during the lesson observation.

Analysis

Descriptive statistics and nonparametric, one-tailed statistical tests with a significance level of $p < .05$ were used to identify increases in teachers' selection and implementation of high-level tasks. Changes over time were compared using one-tailed Wilcoxon signed-rank tests for nonparametric, paired data. Two-way ANOVAs with "curriculum type" and "time" as the grouping variables were

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

conducted to determine whether the availability of a *Standards*-based or conventional curriculum influenced teachers' selection and implementation of mathematical tasks before and after their participation in ESP. To supplement the results of the ANOVA, Mann-Whitney tests were used to compare differences in means between data collections, and chi-squared tests were used to compare differences in the number of tasks and implementations at a high level or low level between data collections. Finally, teachers from ESP and contrast groups were compared using Mann-Whitney tests (one-tailed). When statistically significant differences were found, the nature of the differences were described qualitatively using the rubric descriptors for task Potential (Appendix A) and Implementation (Appendix B) and the factors identified by the lesson observer on the Lesson Checklist (Appendix C). The analyses in the current study coordinated multiple sources of evidence to identify, substantiate, and describe teachers' instructional practices with respect to the selection and implementation of cognitively challenging tasks throughout their participation in the ESP professional development sessions. In the following section, we present the results of the analyses.

RESULTS

The results of the analyses are reported in two sections corresponding to teachers' selection and implementation of cognitively challenging tasks, respectively.

Teachers' Selection of Instructional Tasks

Collections of tasks were analyzed by comparing (a) the mean task scores between data collections, (b) the mean task scores between teachers using *Standards*-based and conventional curricula, and (c) the number of high-level and low-level instructional tasks between data collections.

Differences in Mean Scores Between Task Collections

The mean of a teacher's five task scores serves as an indicator of the level of cognitive demand of the tasks used in the teacher's classroom over the 5-day data collection. Table 2 reports the descriptive statistics for the task collections. The Fall task mean of 2.54 lays almost exactly on the demarcation line between high-level and low-level cognitive demands (i.e., between score levels 2 and 3). By Winter, the task mean had increased to 2.93, and according to the Mann-Whitney test, the change of 0.39 represented a significant increase in the level of cognitive demand of teachers' main instructional tasks ($z = 1.79$; $p = .04$ [one-tailed]). The Spring task mean of 3.01 represents an overall significant increase of 0.47 from Fall ($z = 2.34$; $p < .01$ [one-tailed]).

Differences in Mean Scores Between Curriculum Type

Table 3 provides descriptive statistics for task scores grouped by curriculum type.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Table 2
Descriptive Statistics on the Potential of the Task Scores for the Task Collection

Data collection	Number of teachers	Mean (*SD*)	95% confidence interval
Fall	18	2.54 (0.48)	2.30–2.77
Winter	16	2.93 (0.55)*	2.64–3.22
Spring	14	3.01 (0.53)*	2.70–3.31

*Significant increase from Fall at $p < .05$.

A two-way ANOVA with *time* (Fall, Winter, Spring) and *curriculum type* (*Standards*-based or conventional) as the grouping variables was conducted on the subset of 12 teachers for whom data were available for all three data collections to identify whether teachers' task scores (and the increases over time) were influenced by the type of curriculum available for the courses for which the teachers were submitting data. Consistent with the Mann-Whitney test, the ANOVA indicated that ESP teachers significantly increased the level of tasks in their data collections from Fall to Spring ($F = 7.35$; $p < .01$). Interestingly, these gains were not influenced by curriculum type ($F = 3.61$; $p = .09$), and an insignificant interaction between time and curriculum ($F = 1.12$; $p = .35$) indicated that the increase in task levels over time was not significantly greater in one group than in the other; teachers with each type of curricula experienced similar gains in the levels of the tasks in their data.

Table 3
Descriptive Statistics on Potential of the Task Scores Grouped by Curriculum Type

Data collection	Mean (*SD*)				Difference in means
	Standards-based ($n = 6$)		Conventional ($n = 6$)		
Fall	2.83	(0.43)	2.35	(0.53)	0.48
Winter	3.18	(0.27)	2.62	(0.48)	0.56
Spring	3.18	(0.60)	2.99	(0.37)	0.19

Differences in the Number of High-Level Tasks Between Data Collections

The next set of tests determined whether the significant increases in the mean task scores over time identified an increase in the number of high-level tasks in teachers' data collections. Myriad increases in task scores (i.e., increases between score levels 1 and 2 or between score levels 3 and 4) could affect the mean but not affect the percent of tasks at a high level vs. low level of cognitive demand. The number and percent of tasks at each score level is portrayed in Table 4, for all available data in Fall, Winter,

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

and Spring. In Fall, low-level tasks were selected more often than high-level tasks (i.e., 56% vs. 44%) as the main instructional tasks in ESP teachers' classrooms. By Spring, this ratio was approximately 3:1 in favor of high-level tasks (73% vs. 27%). A chi-squared test indicated that the number of high-level instructional tasks in teachers' data collections increased significantly over time ($\chi^2(2) = 16.18; p < .01$).

Table 4
Number (and Percent) of Tasks at Each Score Level for Potential of the Task

Data collection (number of teachers)	Number of tasks ($n \times 5$)	Low-level (L-L) cognitive demands			High-level (H-L) cognitive demands		
		Score = 1	Score = 2	Total L-L	Score = 3	Score = 4	Total H–L
Fall ($n = 18$)	90	3 (3%)	47 (52%)	50 (56%)	31 (34%)	9 (10%)	40 (44%)
Winter ($n = 16$)	80	3 (4%)	23 (29%)	26 (33%)	33 (41%)	21 (26%)	54 (67%)
Spring ($n = 14$)	70	0 (0%)	19 (27%)	19 (27%)	33 (47%)	18 (26%)	51 (73%)

Due to the nature of the IQA AR-Math rubric, changes in score levels reflect qualitative differences in teachers' instructional tasks. In the Fall data collection, over half (52%) of the main instructional tasks were categorized as "procedures without connections" (i.e., Potential score = 2), such as performing computations or following a set of rote, prescribed steps in an algorithm. Only 10% of Fall tasks provided opportunities for students to explicitly demonstrate mathematical reasoning and understanding (i.e., Potential score = 4), such as having to demonstrate connections between mathematical representations or describe the reasoning behind a mathematical procedure. In the Winter and Spring data collections, teachers incorporated a greater percentage of tasks with high-level cognitive demands. Approximately two thirds of the main instructional tasks in the Winter and three fourths of the main instructional tasks in the Spring (67% and 73%, respectively) provided students with opportunities to engage in high-level thinking and reasoning (i.e., a Potential score of 3 or 4).

Teachers' Implementation of Tasks

Teachers' implementation of mathematical tasks was assessed through collections of students' work and through lesson observations. Comparisons between the lesson observations of the ESP group and the lesson observations of the contrast group are also presented.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Collections of Students' Work

Table 5 provides descriptive statistics for the student-work Potential and Implementation scores. Due to the overlap between teachers' main instructional tasks and the student-work tasks (i.e., the tasks for which teachers submitted students' work were often a subset of the instructional tasks analyzed previously), only the analyses of student-work Implementation scores will be discussed in this section.

Table 5
Descriptive Statistics on Student-Work Scores for Potential and Implementation

Data collection (number of teachers)	Sets of student work ($n \times 3$)	Student-work mean (*SD*)	
		Potential	Implementation
Fall ($n = 16$)	48	2.63 (0.79)	2.27 (0.57)
Winter ($n = 15$)	45	3.02 (0.81)	2.59 (0.72)
Spring ($n = 13$)	39	3.03 (0.80)*	2.86 (0.81)*

*Significant difference from Fall at $p < .05$, one-tailed.

Analyzing student-work Implementation scores. Mann-Whitney tests were used to identify significant increases in student-work Implementation scores between data collections, as an indicator of whether instructional tasks were implemented by teachers and enacted by students at a high level of cognitive demand. Between Fall and Winter, the increase in Implementation scores of 0.32 was just below the significance level ($z = 1.64$; $p = .05$; [one-tailed]). Student-work Implementation means continued to increase between Winter and Spring, which contributed to an overall significant increase of 0.59 between Fall and Spring ($z = 2.94$; $p = .002$ [one-tailed]).

Table 6 identifies the number of student-work tasks that were rated as high level for Potential and for Implementation. Similar to the argument stated previously regarding increases in the task collection means, significant increases in Implementation means do not guarantee that teachers implemented more tasks at a high level (i.e., at a score of 3 or 4) if the increases occurred between score levels 1 and 2 (implementations that improved but remained low level) or between score levels 3 and 4 (improvements in implementations that were already high level). The data in Table 6 were analyzed to determine whether increases in the mean Implementation scores for student work reflect a true increase in the number of high-level implementations. A chi-squared test comparing the number of high-level vs. low-level implementations among Fall, Winter, and Spring indicated that the number of high-level implementations evident in students' work increased significantly over time ($\chi^2(2) = 16.11$; $p < .001$).

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Table 6
Comparison of Implementation Scores for Student-Work (SW) Sets Rated as High Level (H-L) for Potential

Data collection	Sets of SW	SW sets rated as high level for Potential		SW sets rated as high level for Implementation		
		Number	Percentage of total SW sets	Number	Percentage of total SW sets	Percentage of SW sets rated H-L for Potential
Fall	48	23	48%	12	25%	52%
Winter	45	32	71%	24	53%	75%
Spring	39	30	77%	26	67%	87%

Student-work Implementation scores improved significantly from Fall to Spring, but earlier results indicated that teachers were also using high-level instructional tasks more frequently in Spring. Arguably, using a greater number of high-level tasks may have enabled teachers to enact a greater number of high-level implementations. This raises the question of whether the increase in Implementation scores is simply a natural consequence of the increase in Potential scores rather than a true reflection of improvements in teachers' ability to maintain high-level cognitive demand during instruction. Another way to analyze improvements in teachers' implementation is to examine the relationship between student-work Potential and Implementation *within* each data collection. Comparisons between the Potential and Implementation data listed in Tables 5 and 6 indicate that in all three data collections: (a) Implementation means are lower than Potential means and (b) the number of high-level implementations is lower than the number of high-level tasks. These differences suggest that even when teachers selected cognitively challenging tasks, the tasks were sometimes enacted in ways that did not provide evidence of students' engagement in high-level thinking and reasoning. However, the data in Table 6 also indicate that the number of student-work tasks maintained at a high-level during implementation increased from 25% (12 out of 48) in Fall to 67% (26 out of 39) in Spring. These statistics suggest that fewer high-level tasks in the student-work collection were declining during implementation in the Spring data collection than in the Fall.

A chi-squared test was used to assess whether teachers' ability to maintain the cognitive demands of high-level student-work tasks improved over time. For student-work tasks that were coded as high-level for Potential, Table 6 reports the number and percent of high-level vs. low-level Implementations in Fall, Winter, and Spring. The chi-squared test indicated that significant changes did occur between data collections ($\chi^2(2) = 7.96$; $p = .02$) in the number of high-level student-work tasks that were maintained at a high level during implementation (i.e., student-work tasks at a score of 3 or 4 for Potential that were maintained at a score level of 3 or 4 for Implementation).

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Descriptors for the Potential and Implementation rubrics confirm the results of the statistical tests, indicating that fewer high-level student-work tasks declined into "procedures without connections" (i.e., score level 2) over time; subsequently, engagement with high-level cognitive demands was evident in a greater number of sets of students' work over time. Only three student-work tasks increased from a score of 3 for Potential to a score of 4 for Implementation. No student-work tasks coded as low-level (i.e., a score of 1 or 2) for Potential were subsequently coded as high-level for Implementation.

Analyzing the influence of curriculum on teachers' student-work Implementation scores. To investigate whether the changes in student-work Implementation scores were influenced by the presence of a *Standards*-based or conventional curriculum, a two-way ANOVA was conducted using the subset of 10 teachers who submitted students' work in all three data collections. Table 7 provides descriptive data for the student-work Implementation scores grouped by curriculum type. Results of the ANOVA test indicated that the main effect for curriculum type ($F = 2.71$; $p = .152$) and the interaction between time and curriculum ($F = 0.72$; $p = .50$) were both nonsignificant. Similar to findings on teachers' selection of tasks, teachers with a *Standards*-based curriculum did not implement student-work tasks at a significantly higher level than teachers with a conventional curriculum, and the improvement in student-work Implementation scores was not significantly greater in one group than in the other. Only the main effect for time was significant ($F = 7.95$; $p = .004$).

Lesson Observations

Lesson observations were conducted for 11 ESP teachers and 10 contrast group teachers. ESP teachers were observed once in each data collection, and contrast group teachers were observed once in the Spring. Table 8 provides descriptive statistics on the Potential and Implementation of the lesson observations and identifies the significant differences between data collections and between ESP teachers and

Table 7
Descriptive Statistics on Student-Work Implementation Scores Grouped by Curriculum Type

Data collection	Curriculum type	*n*	Mean	*SD*
Fall	All	10	2.40	0.38
	Standards-based	5	2.67	0.33
	Conventional	5	2.13	0.18
Winter	All	10	2.60	0.52
	Standards-based	5	2.67	0.53
	Conventional	5	2.53	0.56
Spring	All	10	3.05	0.47
	Standards-based	5	3.20	0.38
	Conventional	5	2.90	0.55

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Table 8
Descriptive Statistics on Lesson Observations

Data collection	Potential mean (*SD*)	Implementation mean (*SD*)
ESP Teachers ($n = 11$)		
Fall	2.68 (0.72)	2.45 (0.69)
Winter	3.23 (0.88)*	2.86 (0.90)
Spring	3.18 (0.75)*	2.91 (0.83)
Contrast group ($n = 10$)	2.40 (0.52)**	2.20 (0.42)**

*Significantly higher than Fall at $p < .05$.
**Significantly lower than ESP in Spring at $p < .05$.

the contrast group. Implementation means from the ESP teachers' lesson observations exhibited an overall increase of 0.46 between Fall and Spring. Results of the Mann-Whitney tests indicate that this increase was not statistically significant ($z = 1.21$; $p = .11$ [one-tailed]).

No significant differences were found between the contrast group and the ESP teachers' Fall scores for Potential ($z = 0.81$; $p = .21$ [one-tailed]) or for Implementation ($z = 0.67$; $p = .25$ [one-tailed]). This indicates that in the Fall data collection, prior to participation in the ESP workshop, ESP teachers used similar levels of tasks and implemented tasks in similar ways as teachers in the contrast group. Conversely, ESP teachers' Spring lesson observation scores were significantly higher than the contrast group in Potential ($z = 2.15$; $p = .02$ [one-tailed]) and in Implementation ($z = 1.87$; $p = .03$ [one-tailed]).

Interestingly, no lesson observations in any data collection increased from a low-level score for Potential to a high-level score for Implementation. Qualitative changes in lesson implementation are reflected in the factors identified by lesson observers on the IQA Lesson Checklist (Appendix C). Between Fall and Winter, a comparison of the factors selected on the Checklist indicated an increase in (a) teachers holding students accountable for high-level products and processes, (b) teachers providing consistent press for explanation and meaning, and (c) teachers providing encouragement for students to make conceptual connections. Winter Lesson Checklists also indicate a corresponding decrease in the occurrence of (a) teachers providing a set procedure for solving the task; (b) the focus shifting to procedural aspects of the task or focus on correctness of the answer rather than on meaning and understanding; (c) feedback, modeling, or examples being too directive or not leaving any complex thinking for the student; and (d) students not being pressed or held accountable for high-level products and processes or for explanations and meaning. The changes identified between Fall and Winter were sustained between Winter and Spring, although no new changes or patterns emerged. The contrast group more closely reflected the patterns and factors of implementation exhibited by the ESP teachers in the Fall. Hence, following their participation in ESP, the ESP teachers were more likely to exhibit classroom factors that main-

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

tained high-level cognitive demands and less likely to exhibit classroom factors that reduce high-level cognitive demands (Stein et al., 1996) than (a) prior to their participation in the ESP workshops and (b) a group of teachers who did not participate in the ESP workshops.

DISCUSSION

The results of this study provide evidence that, following their participation in the ESP workshop, ESP teachers significantly increased (a) the level of cognitive demand of the main instructional tasks in their data collections and (b) their ability to maintain high-level cognitive demands in students' work. Furthermore, as compared to a contrast group of secondary mathematics teachers who did not participate in the ESP professional development workshop, the ESP teachers had significantly higher task and implementation mean scores for classroom observations. Changes in ESP teachers' selection and implementation of cognitively challenging tasks are discussed in the sections that follow.

ESP Teachers Increased Their Selection of High-Level Instructional Tasks

The level of cognitive demand of the main instructional tasks in ESP teachers' data collections increased significantly from the Fall to the Spring data collection, which mark the beginning and end, respectively, of teachers' participation in the ESP year-1 workshop. The increase of 0.47 between the Fall task collection mean of 2.54 and the Spring task collection mean of 3.01 is important in terms students' opportunities for learning mathematics with understanding because the increase occurred exactly when it was crucial for moving teachers' main instructional tasks from low level (i.e., a score of 1 or 2) to high level (i.e., a score of 3 or 4) cognitive demands. Note that an increase between score levels 1 and 2 would have indicated that teachers were still using low-level tasks, just different types of low-level tasks. Similarly, an increase between score levels 3 and 4 would have indicated that teachers enhanced the high-level tasks that they were already using. Instead, the increase between score levels 2 and 3 illustrates a change from instructional tasks characterized by low-level cognitive demands in the Fall data collection to instructional tasks characterized by high-level cognitive demands in the Spring.

The ANOVA results in this study indicated that the presence of a *Standards*-based or conventional curriculum in teachers' classrooms was not a significant influence on teachers' selection or implementation of high-level tasks. This finding is consistent with current studies conducted by Tarr and colleagues (Tarr et al., 2008) in which a reform-oriented learning environment was shown to support higher student achievement regardless of the type of curriculum (conventional or Standards-based) in place in the teacher's classroom. The lack of an effect of curriculum type on task selection, however, is particularly surprising; Standards-based curricula are specifically designed to contain a greater percentage of cognitively challenging tasks (AAAS, 2000; USDE, 1999). The nonsignificant results do not indicate an absence

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

of high-level tasks in Standards-based curricula. Rather, they reflect a prevalent research finding that mathematics teachers do not always use *Standards*-based curricular materials as intended by the curriculum developer (Lloyd, 1999; Lloyd & Wilson, 1998; Remillard, 1999; Remillard & Bryans, 2004). Teachers' conceptions "act as critical filters" (Lloyd & Wilson, 1998, p. 250) that govern their use of curricula in ways that can be supportive of or antithetical to reform-oriented mathematics pedagogy. In the current investigation, this statement was true of teachers using both *Standards*-based and conventional curricula. Close examination of the main instructional tasks revealed that teachers who had access to *Standards*-based curricula often did not use the high-level tasks offered by the curricula as their main instructional tasks, and teachers with access to conventional curricula often used supplementary materials or created extensions that increased the cognitive demands of the tasks in their curricular materials. Both groups selected the main instructional tasks for their data collections in all of the ways identified by Remillard and Bryans (2004, p. 363): *guided by* the curricula (teachers used their *Standards*-based or conventional curricula as the source of the task and of the lesson activities); *drawn from* the curricula (teachers used the tasks from the curricula, but implemented these tasks in their own way), *adapted from* the curricula (teachers altered the tasks in the curricula), or replaced the curricula with *other resources* or tasks of their *own design*. Remillard and Bryans refer to teachers' selection and adaptation of instructional materials as teachers' "orientation toward curricula," defined as "a set of perspectives and dispositions about mathematics, teaching, learning, and curriculum that together influence how a teacher engages and interacts with a particular set of curricular materials" (p. 364).

Hence, our findings contribute to a growing body of research that suggests that changing teachers' orientations toward their curricula in ways that support the selection of cognitively challenging instructional tasks is a necessary and vital step in promoting reform-oriented instructional practices and enhancing students' opportunities to learn mathematics with understanding. Too often, school resources are directed primarily toward the purchase and adoption of new curricula while overlooking the importance of providing professional learning opportunities that will enable teachers to enact the curricula as intended. This study, with all caveats due to small sample size, suggests that professional development specifically focused on the selection and implementation of cognitively challenging tasks can support teachers to develop an orientation toward using such tasks as the basis of instruction. Furthermore, teachers in this study began to select a greater number of high-level tasks as the main instructional tasks in their Winter and Spring data collections regardless of the curriculum in place in their classrooms. We speculate that an awareness of the importance of cognitively challenging mathematical tasks, and access to tools to assess the level of cognitive demand (i.e., the Task Analysis Guide in Figure 1), can support teachers to use, adapt, or supplement their curricular materials in ways that capitalize on the cognitively challenging tasks that are available or alleviate the curriculum's weaknesses in providing such tasks. Greater resources should be extended toward providing mathematics teachers with professional

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

development experiences that will maximize the benefits of a *Standards*-based curriculum or enhance the opportunities for high-level thinking and reasoning offered by conventional curricula.

One question raised by this analysis is the extent to which the main instructional tasks submitted in teachers' data collections generalize to the main instructional tasks used in teachers' classrooms on a regular basis. One possibility is that the results *do not* generalize to teachers' everyday instructional practices; rather, teachers inferred from the professional development activities that we were "looking for" high-level tasks in their data collections and used such tasks as the main instructional task during the week of data collection *only*. The initial, statistically significant jump in task means between Fall and Winter, however, indicates that ESP teachers selected high-level tasks for instruction during the week of the Winter data collection, following their participation in three of six professional development sessions. One indication that many teachers' Winter task collections did *not* generalize to their typical classroom practices is the absence of a parallel increase in student-work *Implementation* means from Fall to Winter. This could indicate that several teachers and their students were still unaccustomed to working with high-level instructional tasks and had not firmly developed the habits of practice that sustain high-level cognitive engagement. The Winter data collection occurred immediately following the third 1-day professional development session. At this point, teachers had engaged in solving and reflecting on cognitively challenging tasks and analyzing the level of cognitive demand of mathematical tasks. They had also been introduced to the Task Analysis Guide. Hence, the emphasis of the professional development prior to the Winter data collection was the selection of high-level tasks. It would not be surprising, then, if some teachers provided high-level tasks in their Winter data collections simply to please the researchers.

However, the continued, gradual improvements in student-work Implementation from Fall to Winter and from Winter to Spring indicate that teachers and students were becoming more skilled at maintaining high-level cognitive engagement. Implementing high-level tasks in ways that maintain the cognitive demands is not a trivial endeavor, as documented by large-scale studies such as QUASAR (Henningsen & Stein, 1997), the TIMSS 1999 Video Study (NCES, 2003), and Horizon Research (Weiss & Pasley, 2004; Weiss et al., 2003). We assert that, if teachers were not *selecting* high-level tasks on a regular basis, improvements would not be evident in the ways in which they were *implementing* high-level tasks.

ESP Teachers Improved Their Implementation of High-Level Tasks

From Fall to Spring, teachers' student-work samples increased in (a) the mean Implementation score, (b) the number of high-level implementations, and (c) the number of high-level tasks maintained at a high level during implementation. Comparisons of the implementation of high-level student-work tasks indicated that high-level demands were less likely to decline in Spring than in Fall. As noted earlier, the difficulty of maintaining high-level task demands is well documented

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

in research (Henningsen & Stein, 1997; NCES, 2003; Weiss & Pasley, 2004), which helps to dissipate the notion that improvements in implementation were limited to the student work submitted for the data collections. If teachers could simply will themselves to implement tasks at a high level, every high-level task in the data collection would have been maintained at a high level during implementation, and this change would have been immediately evident in the Winter data collection. Rather, the gradual increase in the student-work Implementation means from Fall to Spring and the factors selected on the Lesson Checklist indicate that ESP teachers were consistently improving their ability to maintain high-level cognitive demands. Teachers were working toward instructional change, and evidence of improvement is an indication that they were making an effort to implement and maintain high-level tasks in their classrooms at other times beyond the weeks of data collection. *Working toward* instructional change does not imply that all ESP teachers were experts at implementing high-level tasks following their participation in ESP. Rather, 5 ESP teachers enhanced their previous ability to implement high-level tasks and 8 teachers improved their implementation of high-level tasks in ways that pushed student-work Implementation means over the demarcation line between instruction characterized by low-level cognitive demands and instruction characterized by high-level cognitive demands. The increases occurred exactly when they mattered most in term of students' opportunities to engage in high-level thinking and reasoning during mathematics instruction.

Although student-work Implementation increased significantly, Implementation for lesson observations did not. This discrepancy can be attributed to the differences between the characteristics of the teachers providing student work and those participating in the lesson observations. Ten of the 11 teachers participating in the lesson observations also submitted sets of students' work. Hence, the teachers providing lesson observations are a subset of 10 of the 17 teachers providing student work. From this subset, 4 of the teachers who participated in the lesson observations were already implementing high-level tasks on a regular basis. Although their lesson observation Implementation scores improved slightly over time, their improvements and the changes exhibited by 6 of the remaining 7 teachers did not amount to a statistically significant increase in the Implementation scores for lesson observations. The analysis of student-work implementation was based on a (slightly) larger sample size, was based on three samples of student work per data collection (vs. one lesson observation), and consisted of more teachers who experienced significant growth in their ability to implement cognitively challenging tasks. Within the data from individual teachers, it is important to note that student work and lesson observations portray a consistent picture of instruction in the teacher's classroom. Also, the Implementation mean scores for student work and lesson observations are quite similar in the Fall (2.27 vs. 2.45), Winter (2.59 vs. 2.86), and Spring (2.86 vs. 2.91).

Though small sample size, lack of statistical power (due to one lesson observation per data collection), and the subset of four skilled teachers all worked against the statistical significance of the increase in Implementation scores for lesson obser-

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

vations, several qualitative improvements in the lesson implementations are noteworthy. A comparison of field notes across lesson observations in the Fall, Winter, and Spring indicates that even the subset of four skilled teachers improved the quality of the mathematical discussions in their classrooms and the questions they asked to support students' work. Changes in the factors selected on the Lesson Checklists also suggest that they incorporated new ways of holding students accountable for high-level thinking and reasoning. Of the seven remaining teachers who were observed, six teachers exhibited improvements in task implementation during their lesson observations and one exhibited no change. In the Fall data collection, this group of teachers consistently proceduralized the demands of any high-level instructional tasks in their student work or lesson observations. In other words, when a high-level task (i.e., score of 3 or 4) was used as the main instructional task, the tasks consistently fell to a score of 2 for Implementation. By Spring, six of the seven teachers had developed instructional strategies that supported students' thinking and reasoning, as indicated by their ability to maintain high-level demands from Potential to Implementation.

Significant increases in teachers' selection and implementation of cognitively challenging tasks may suggest that ESP teachers utilized the MTF and other "tools" provided by the ESP workshop as a focus for changing their instructional practices (Smith, Boston, & Steele, 2006). According to the National Academy of Education (1999),

> tools—including student assessments, curriculum and professional-development materials . . . , and protocols for observing classrooms or professional meetings—are powerful carriers of theory and knowledge. Carefully designed tools that educators find useful in their practice can, then, become a powerful means of changing educational practice. (p. 35)

Further investigation into whether (and how) the tools enabled the ideas that emerged in the ESP sessions to "travel" into teachers' classrooms, to support teachers' selection and implementation of high-level tasks, and to focus teachers' analysis of their own instructional practice will be an important facet of our future work. We close this discussion by drawing conclusions from the results of this study and suggesting additional directions for future endeavors.

CONCLUSIONS

The results of this study are important to mathematics professional development, teaching, and learning primarily because the changes in teachers' instructional practices, throughout their experiences in the ESP professional development workshop, have the potential to improve students' opportunities for learning mathematics. Although this study did not measure the impact of the changes in teachers' instructional practices on students' learning, evidence from other studies (e.g., Hiebert & Wearne, 1993; NCES, 2003; Stein & Lane, 1996) suggests that increased exposure to cognitively challenging tasks and extended engagement with high-level cognitive demands increases students' learning of mathematics.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

This study contributes to a small body of professional development research that utilized classroom observations and artifacts to provide evidence of enhancements in teachers' instructional practices following their participation in a professional development intervention. The current investigation strengthens the knowledge base on teachers' instructional change by describing changes in teachers' implementation of high-level tasks and by utilizing classroom artifacts and observations, rather than self-report data, as the main data source. Furthermore, this investigation utilized a tool for analyzing classroom observations and collections of student work (i.e., the IQA Academic Rigor in Mathematics rubric) that provided qualitative data *and* served as a statistically sound instrument for collecting quantitative data on teachers' selection and implementation of cognitively challenging tasks. This study also contributes to research on the use of student work as an indicator of classroom practice (Clare, 2000; Clare & Aschbacher, 2001; Matsumura, 2003; Matsumura et al., 2002). Recent research by the IQA team has shown that in mathematics, student work is a stable measure of instructional practice that is highly correlated with observed instruction (Matsumura et al., 2006). This suggests that student work can provide a proxy for lesson observations that is statistically and qualitatively consistent with observed instruction and thus holds implications for the design of future research into teachers' instructional practices. We contend that the IQA rubrics can serve as one of a small number of tools that have the potential to measure instruction at scale, such as the artifacts packages developed at CRESST (Borko, Stecher, Alonzo, Moncure, & McClam, 2005) and the instructional logs developed by the Study of Instructional Improvement (Rowan, Camburn, & Correnti, 2004; Rowan, Harrison, & Hayes, 2004).

In summary, this study has identified improvements in teachers' selection and implementation of cognitively challenging tasks following their participation in a specifically focused professional development workshop. These instructional changes hold promise for improving students' learning of mathematics in ESP teachers' classrooms and suggest that the ESP workshop can serve as one model of the type of professional development capable of supporting improvements in teachers' instructional practices and students' learning more broadly. Though small sample size confines the results of this study, the nature and extent of changes in teachers' instructional practices merits further investigation into the design and content of the ESP workshop and its relationship with teachers' instructional change. Future directions for research on the efficacy of this model include a follow-up assessment of the maintenance and continued growth of ESP teachers' selection and implementation of high-level instructional tasks; replicating the results of the study with larger groups of teachers or with a different set of professional development providers; and establishing the link among professional development, teachers' knowledge and instructional practices, and students' achievement in mathematics.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

REFERENCES

Arbaugh, F. (2000). Time on tasks: Influences of a study group on secondary mathematics teachers' knowledge, thinking, and teaching. (Doctoral dissertation, Indiana University). Retrieved November 10, 2008, from Dissertations & Theses: A&I database. (Publication No. AAT 9993602).

Arbaugh, F. (2003). Study groups as a form of professional development for secondary mathematics teachers. *Journal of Mathematics Teacher Education, 6*, 139–163.

Arbaugh, F., & Brown, C. A. (2002, April). *Influences of the mathematical tasks framework on high school mathematics teachers' knowledge, thinking, and teaching.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Arbaugh, F., Lannin, J., Jones, D. L., & Park-Rogers, M. (2006). Examining instructional practices in Core-plus lessons: Implications for professional development. *Journal of Mathematics Teacher Education, 9,* 517–550.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Towards a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey-Bass.

Barnett, C. (1998). Mathematics teaching cases as a catalyst for informed strategic inquiry. *Teaching and Teacher Education, 14*, 81–93.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers College Record, 110*, 608–645.

Borasi, R., & Fonzi, J. (2002). Engaging in scaffolded instructional innovation. *Foundations: Professional development that supports school reform* (pp. 83–98). Washington, DC: National Science Foundation.

Borko, H. & Putnam, R. T. (1995). Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development. In T. R. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices*. Mahwah, NJ: Lawrence Erlbaum.

Borko, H., Stecher, B. M., Alonzo, A., Moncure, S., & McClam, S. (2005). Artifact packages for characterizing classroom practice: A pilot study. *Educational Assessment, 10*, 73–104.

Boston, M. D. (2006). *Developing secondary mathematics teachers' knowledge of and capacity to implement instructional tasks with high-level cognitive demands* (Doctoral dissertation, University of Pittsburgh). *Dissertation Abstracts International* (Publication No. AAT 3223943).

Boston, M. D., & Wolf, M. K. (2006). *Assessing academic rigor in mathematics instruction: The development of the instructional quality assessment toolkit: CSE technical report 672.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*, 499–531.

Carpenter, T. P. & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema and T. Romberg (Eds.), *Mathematics classrooms that promote understanding* (p. 19–32). Mahwah, NJ: Lawrence Erlbaum.

Clare, L. (2000). *Using teachers' assignments as an indicator of classroom practice: CSE technical report 532.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Clare, L., & Aschbacher, P. R. (2001). Exploring the technical quality of using assignments and student work as indicators of classroom practice. *Educational Assessment, 7*, 39–59.

Clarke, D. M. (1997). The changing role of the mathematics teacher. *Journal for Research in Mathematics Education, 28*, 278–308.

Cobb, P., Yackel, E., & Wood, T. (1991). Curriculum and teacher development: Psychological and anthropological perspectives. In E. Fennema, T. P. Carpenter, & S. J. Lamon (Eds.), *Integrating research on teaching and learning mathematics*. Albany, NY: State University of New York Press.

Denzin, N. K., & Lincoln, Y. S. (2000). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1–28). Thousand Oaks, CA: Sage Publications.

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

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

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

Fennema, E., Carpenter, T., Franke, M., Levi, L., Jacobs, V., & Empson, S. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403–434.

Fernandez, C. (2005). Lesson study: A means for elementary teachers to develop the knowledge of mathematics needed for reform-minded teaching? *Mathematical Thinking & Learning, 7,* 265–289.

Frykholm, J. (2005). Innovative curricula: Catalysts for reform in mathematics teacher education. *Action in Teacher Education, 26*(4), 20–36.

Ginsburg, A., Cooke, G., Leinwand, S., Noell, J., & Pollock, E. (2005). *Reassessing U.S. international mathematics performance: New findings from the 2003 TIMSS and PISA.* Washington, D.C: American Institutes for Research.

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education, 28*, 524–549.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., et al. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

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

Hughes, E. K., Smith, M. S., Hogel, M., & Boston, M. D. (in press). Case stories: Supporting teacher reflection and collaboration on the implementation of cognitively challenging mathematical tasks. In F. Arbaugh & P. M. Taylor (Eds.), *Inquiry into mathematics teacher education.* Association of Mathematics Teacher Educators (AMTE) Monograph Series, Volume 5.

Jacobs, V. R., & Philipp, R. A. (2004). Supporting teacher learning: Mathematical thinking: Helping prospective and practicing teachers focus. *Teaching Children Mathematics, 11*, 194–201.

Janesick, V. J. (2000). Choreography of qualitative research design: Minuets, improvisations, and crystallization. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 379–399). Thousand Oaks, CA: Sage Publications.

Kazemi, E., & Franke, M. L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education, 7*, 203–235.

Lemke, M., Sen, A., Pahike, E., Partelow, L., Miller, D., Williams, T., et al. (2004). *International outcomes of learning in mathematics literacy and problem solving: Pisa 2003 results from the U.S. perspective.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Little, J. W. (2004). Looking at student work in the United States: Countervailing impulses in professional development. In C. Day and J. Sachs (Eds.), *International handbook on the continuing professional development of teachers* (pp. 94–118). Buckingham, England: Open University Press.

Lloyd, G. M. (1999). Two teachers' conceptions of a reform-oriented curriculum: Implications for mathematics teacher development. *Journal of Mathematics Teacher Education, 2*, 227–252.

Lloyd, G. M., & Wilson, M. L. (1998). Supporting innovation: The impact of a teacher's conception of functions on his implementation of a reform curriculum. *Journal for Research in Mathematics Education, 29*, 248–274.

Matsumura, L. C. (2003). *Teachers' assignments and student work: Opening a window on classroom practice: CSE technical report 602.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Matsumura, L. C., Garnier, H., Pascal, J., & Valdes, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment, 8*, 207–229.

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale." *Educational Assessment, 13*(4), 267–300.

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., et al. (2006). *Reading comprehension and mathematics instruction in urban middle schools: A pilot study of the instructional quality assessment toolkit: CSE technical report 681.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). TIMSS 2003 international mathematics report. Boston: TIMSS & PIRLS International Study Center, Boston College.

National Academy of Education. (1999). *Recommendations regarding research priorities: An advisory report to the National Educational Research Policy and Priorities Board.* Retrieved November 10, 2008, from http://www.naeducation.org/Research_Priorities_Publication.pdf

National Center for Education Statistics. (2003). *Teaching mathematics in seven countries: Results from the TIMSS video study.* Washington, DC: U.S. Department of Education.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

RAND Mathematics Study Panel. (2003). *Mathematical proficiency for all students.* Santa Monica, CA: RAND.

Remillard, J. (1999). Curriculum materials in mathematics education reform: A framework for examining teachers' curriculum development. *Curriculum Inquiry, 29*, 315–342.

Remillard, J. T., & Bryans, M. B. (2004). Teachers' orientation toward mathematics curriculum materials: Implications for teacher learning. *Journal for Research in Mathematics Education, 35*, 352–388.

Romagnano, L. (1994). *Wrestling with change: The dilemmas of teaching real mathematics.* Portsmouth, NH: Heinemann.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum in large-scale surveys: A study of literacy teaching in 3rd grade classrooms. *Elementary School Journal, 105*, 75–102.

Rowan, B., Harrison, D., & Hayes, A. (2004). Using instructional logs to study elementary school mathematics: A close look at curriculum and teaching in the early grades. *Elementary School Journal, 105*, 103–127.

Schoen, H. L., Cebulla, K. J., Finn, K. F., & Fi, C. (2003). Teacher variables that relate to student achievement when using a standards-based curriculum. *Journal for Research in Mathematics Education, 34*, 228–259.

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

Smith, M. S. (2000). Balancing the old and new: An experienced middle school teacher's learning in the context of mathematics instructional reform. *The Elementary School Journal, 100*, 351–375.

Smith, M. S. (2001). *Practice-based professional development for teachers of mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Smith, M. S., Bill, V., & Hughes, E. K. (2008). Thinking through a lesson: Successfully implementing high-level tasks. *Mathematics Teaching in the Middle School, 14,* 132–138.

Smith, M. S., Boston, M., & Steele, M. (2006, January). *Focusing on challenging mathematical tasks: A strategy for improving teaching and learning.* Paper presented at the 10th annual meeting of the Association of Mathematics Teacher Educators, Tampa, FL.

Smith, M. S., Stein, M. K., Arbaugh, F., Brown, C., & Mossgrove, J. (2004). Characterizing the cognitive demands of mathematical tasks: A task sorting activity. In G. Bright and R. Rubenstein (Eds.), *Professional development guidebook for perspectives on teaching of mathematics: Companion to the sixty-sixth yearbook* (pp. 45–72). Reston, VA: National Council of Teachers of Mathematics.

Stein, M. K., Baxter, J. & Leinhardt, G. (1990). Subject matter knowledge and elementary instruction: A case from functions and graphing. *American Educational Research Journal, 27*, 639–663.

Stein, M. K., Grover, B., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*, 455–488.

Stein, M. K., & Lane, S. (1996). 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, 2,* 50–80.

Stein, M. K., Silver, E. A., & Smith, M. S. (1998). Mathematics reform and teacher development: A community of practice perspective. In J. Greeno & S. Goldman (Eds.), *Thinking practices in mathematics and science learning* (pp. 17–52). Hillsdale, NJ: Erlbaum.

Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development.* New York: Teachers College Press.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

Stigler, J. W., & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership*, *61*(5), 12–16.

Swafford, J. O., Jones, G. A., & Thornton, C. A. (1997). Increased knowledge in geometry and instructional practice. *Journal for Research in Mathematics Education, 28*, 467–483.

Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. (2008). The impact of middles-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*, 247–280.

Thompson, C. L., & Zeuli, J. S. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 341–375). San Francisco: Jossey-Bass.

United States Department of Education. (1999). *Exemplary and promising mathematics programs.* Retrieved November 10, 2008, from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/15/de/f1.pdf.

Wallen, S. B., & Williams, S. R. (2000). Validating classroom issues: Case methods in support of teacher change. *Journal of Mathematics Teacher Education, 3,* 3–26.

Weiss, I. R., & Pasley, J. P. (2004). What is high quality instruction? *Educational Leadership, 61*(5), 24–28.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K–12 mathematics and science education in the United States (Highlights report).* Chapel Hill, NC: Horizon Research, Inc.

Authors

Melissa D. Boston, School of Education, Duquesne University, 600 Forbes Avenue, Pittsburgh, PA 15282; bostonm@duq.edu

Margaret S. Smith, Department of Instruction and Learning, University of Pittsburgh, 5515 WWPH, 230 S. Bouquet St., Pittsburgh, PA 15260; pegs@pitt.edu

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

APPENDIX A

IQA Academic Rigor: Mathematics Rubric for Potential of the Task
(Adapted from Boston & Wolf, 2006, and Matsumura et al., 2006)

4	The task has the potential to engage students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships, such as: • Doing mathematics: using complex and nonalgorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example). OR • Applying procedures with connections: applying a broad general procedure that remains closely connected to mathematical concepts. The task must explicitly prompt for evidence of students' reasoning and understanding. For example, the task *may* require students to— • solve a genuine, challenging problem for which students' reasoning is evident in their work on the task; • develop an explanation for why formulae or procedures work; • identify patterns and form generalizations based on these patterns; • make conjectures and support conclusions with mathematical evidence; • make explicit connections among representations, strategies, or mathematical concepts and procedures; and • follow a prescribed procedure in order to explain/illustrate a mathematical concept, process, or relationship.
3	The task has the potential to engage students in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships. However, the task does not warrant a "4" because— • it does not explicitly prompt for evidence of students' reasoning and understanding; • students may be asked to engage in doing mathematics or procedures with connections, but the underlying mathematics in the task is not appropriate for the specific group of students (i.e., too easy or too hard to promote engagement with high-level cognitive demands); • students may need to identify patterns but are not pressed for generalizations; • students may be asked to use multiple strategies or representations, but the task does not explicitly prompt students to develop connections between them; and

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

3	• students may be asked to make conjectures but are not asked to provide mathematical evidence or explanations to support conclusions.
2	The potential of the task is limited to engaging students in using a procedure that is either specifically called for or its use is evident based on prior instruction, experience, or placement of the task. There is little ambiguity about what needs to be done and how to do it. The task does not require students to make connections to the concepts or meaning underlying the procedure being used. The focus of the task appears to be on producing correct answers rather than developing mathematical understanding (e.g., applying a specific problem-solving strategy, practicing a computational algorithm). OR The task does not require student to engage in cognitively challenging work; the task is easy to solve.
1	The potential of the task is limited to engaging students in memorizing or reproducing facts, rules, formulae, or definitions. The task does not require students to make connections to the concepts or meaning that underlie the facts, rules, formulae, or definitions being memorized or reproduced.
0	The task requires no mathematical activity.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

APPENDIX B

IQA Academic Rigor: Mathematics Rubric for Implementation of the Task
(Adapted from Boston & Wolf, 2006 and Matsumura et al., 2006)

4	Students engaged in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships, such as: • Doing mathematics: using complex and nonalgorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example). OR • Applying procedures with connections: applying a broad general procedure that remains closely connected to mathematical concepts. There is explicit evidence of students' reasoning and understanding. For example, students may have— • solved a genuine, challenging problem for which students' reasoning is evident in their work on the task; • developed an explanation for why formulae or procedures work; • identified patterns and formed generalizations based on these patterns; • made conjectures and supported conclusions with mathematical evidence; • made explicit connections among representations, strategies, or mathematical concepts and procedures; • followed a prescribed procedure in order to explain/illustrate a mathematical concept, process, or relationship.
3	Students engaged in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships. However, the implementation does not warrant a "4" because— • there is no explicit evidence of students' reasoning and understanding; • students engaged in doing mathematics or procedures with connections, but the underlying mathematics in the task was not appropriate for the specific group of students (i.e., too easy or too hard to sustain engagement with high-level cognitive demands); • students identified patterns but did not make generalizations; • students used multiple strategies or representations, but connections between different strategies/representations were not explicitly evident; • students made conjectures but did not provide mathematical evidence or explanations to support conclusions.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

2	Students engaged in using a procedure that was either specifically called for or its use was evident based on prior instruction, experience, or placement of the task. There was little ambiguity about what needed to be done and how to do it. Students did not make connections to the concepts or meaning underlying the procedure being used. Focus of the implementation appears to be on producing correct answers rather than developing mathematical understanding (e.g., applying a specific problem-solving strategy, practicing a computational algorithm). OR Students did not engage in cognitively challenging work; the task was easy to solve.
1	Students engaged in memorizing or reproducing facts, rules, formulae, or definitions. Students did not make connections to the concepts or meanings that underlie the facts, rules, formulae, or definitions being memorized or reproduced.
0	Students did not engage in mathematical activity.

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions

APPENDIX C

IQA Lesson Checklist
(Adapted from Boston & Wolf, 2006, and Matsumura et al., 2006)

Check each box that applies:

A ↑ The lesson provided opportunities for students to engage with the high-level demands of the task:	B ↓ During the lesson, the high-level demands of the task were removed or reduced:
☐ Students engaged with the task in a way that addressed the teacher's goals for high-level thinking and reasoning.	☐ The task expectations were not clear enough to promote students' engagement with the high-level demands of the task.
☐ Students communicated mathematically with peers.	☐ The task was not complex enough to sustain student engagement in high-level thinking.
☐ Students had appropriate prior knowledge to engage with the task.	☐ The task was too complex to sustain student engagement in high-level thinking (i.e., students did not have the prior knowledge necessary to engage with the task at a high level).
☐ Teacher supported students to engage with the high-level demands of the task while maintaining the challenge of the task.	☐ Classroom management problems interfered with students' opportunities to engage in high-level thinking.
☐ Students had opportunities to serve as the mathematical authority in the classroom.	☐ Teacher provided a set procedure for solving the task.
☐ Teacher provided sufficient time to grapple with the demanding aspects of the task and for expanded thinking and reasoning.	☐ The focus shifted to procedural aspects of the task or on correctness of the answer rather than on meaning and understanding.
☐ Teacher held students accountable for high-level products and processes.	☐ Feedback, modeling, or examples were too directive or did not leave any complex thinking for the student.
☐ Teacher provided consistent presses for explanation and meaning.	☐ Students were not pressed or held accountable for high-level products and processes or for explanations and meaning.
☐ Teacher provided students with sufficient modeling of high-level performance on the task.	☐ Students were not given enough time to deeply engage with the task or to complete the task to the extent that was expected.
☐ Teacher provided encouragement for students to make conceptual connections.	☐ Students did not have access to resources necessary to engage with the task at a high level.
☐ Students had access to resources that supported their engagement with the task.	☐ Other:
☐ Other:	

This content downloaded from 150.131.192.151 on Thu, 02 Apr 2015 14:03:37 UTC
All use subject to JSTOR Terms and Conditions