Am Educ Res J OnlineFirst, published on April 19, 2010 as doi:10.3102/0002831209361210

American Educational Research Journal
Month XXXX, Vol. XX, No. X, pp. 1–42
DOI: 10.3102/0002831209361210

Selecting and Supporting the Use of Mathematics Curricula at Scale

Mary Kay Stein
Julia H. Kaufman
The Learning Policy Center, University of Pittsburgh

This article begins to unravel the question, "What curricular materials work best under what kinds of conditions?" The authors address this question from the point of view of teachers and their ability to implement mathematics curricula that place varying demands and provide varying levels of support for their learning. Specifically, the authors focus on how teacher capacity (their level of education, experience, and knowledge) and their use of curriculum influence instruction. The study sample is 48 teachers implementing two standards-based mathematics curricula—Everyday Mathematics *and* Investigations—*in two school districts. The data include interviews and surveys with teachers, as well as observations of instruction, over a 2-year period. Findings indicate that teachers' implementation of* Investigations *was considerably better than teachers' implementation of* Everyday Mathematics *in terms of maintaining high levels of cognitive demand, attention to student thinking, and mathematical reasoning. These implementation measures were not correlated to measures of teacher capacity across school districts. However, implementation measures were significantly correlated with teachers' lesson preparation that took into account the big mathematical ideas within curriculum. Further qualitative analysis indicated that the* Investigations *curriculum provided more support to teachers for locating and understanding the big mathematical ideas within lessons compared to* Everyday Mathematics.

Mary Kay Stein holds a joint appointment as a professor in the School of Education and senior scientist at the Learning Research and Development Center; she is also founding director of the Learning Policy Center, University of Pittsburgh, 3939 O'Hara Street, 8th Floor, Pittsburgh, PA 15260; e-mail: *mkstein@pitt.edu*. Her research seeks to understand how institutional, interpersonal, and policy contexts shape teachers' learning and their practice.

Julia H. Kaufman is a research associate at the Learning Research and Development Center and executive director of the Learning Policy Center, University of Pittsburgh; e-mail: *juliak@pitt.edu*. Her research focuses on how innovative practices are implemented in schools and what policies and supports can lead to high-quality teaching and learning around those practices.

KEYWORDS: curriculum, instructional practices, teacher knowledge, mathematics education, educational reform, longitudinal studies

Over the past decade, district policies have become increasingly focused on the improvement of instruction, especially in subjects that are regularly tested under No Child Left Behind (NCLB; Elmore & Burney, 1999; Hightower, Knapp, Marsh, & McLaughlin, 2002; Hubbard, Mehan, & Stein, 2006; Supovitz, 2006). In mathematics, curriculum has traditionally been viewed as the key policy lever for improving instruction and learning on a large scale. Yet, curriculum alone has been shown to have limited influence on teachers' instructional practices (Ball & Cohen, 1996; Coburn, 2001; Fullan, 1991; Fullan & Pomfret, 1977; Wilson, 1990). While it may be relatively easy to get curriculum materials into the hands of large numbers of teachers, it is much more difficult to select appropriate materials for a given school or district context and to design the conditions that will enable teachers to implement them in ways that are intended by the developers.

The past decade has witnessed growing numbers of district-wide adoptions of standards-based mathematics curricula, two of which are studied in this article: *Everyday Mathematics* (*EM*) and *Investigations*. Both of these curricula aim for more ambitious forms of student learning (i.e., conceptual understanding; the capacity to think, reason, and problem solve) than teachers have traditionally been accustomed to and thus represent a significant challenge for teacher learning as well as student learning.

The conventional wisdom is that, even among standards-based curricula, some are more difficult for teachers to implement than are others. When deciding which curriculum to adopt, district leaders often weigh the perceived level of challenge of a particular curriculum against the perceived strengths and weaknesses of their faculty and/or the levels of professional development in which they are willing to invest. For example, in one large urban district, the decision came down to *Everyday Mathematics* versus *Investigations*. According to a key district leader,

> I had worked with both [*EM* and *Investigations*], and I think, for me, . . . that [the district] could have done very well with either series given the right level of support for our schools. . . . However, and this was the union's decision, they felt that *EM* was easier to implement, therefore, there would be less pressure or less stress for teachers to implement *EM*. That was very explicit.

Similarly, the Director of Elementary Mathematics noted,

> I felt that [*EM*] definitely is more scripted than *Investigations* and I felt that for those teachers who needed something more scripted, though that is not necessarily what I promote most, but I felt that it was there and enough to support children.

If true, the above conventional wisdom suggests that districts composed primarily of inexperienced or low-capacity teachers might do better to adopt *EM*, while districts with a higher capacity teaching force might feel "up to the task" of taking on *Investigations*.

Unfortunately, research offers little help to district leaders who find themselves faced with such a decision. Most recent studies (Agodini et al., 2009; Riordan & Noyce, 2001; What Works Clearinghouse, 2007) only focus on the effects of curricula on student achievement without taking into account the major factor mediating the effects of curricula on student learning: teacher instruction. Additionally, such studies do not consider the variable demands on teacher learning required by certain curricula, particularly standards-based mathematics curricula. Moreover, few, if any, studies compare two standards-based curricula to each other; most compare a conventional curriculum (e.g., Saxon) to a standards-based curriculum.

Here we report on research that begins to unravel the question, "What curricular materials work best under which kinds of conditions?" We address this question from the point of view of teachers' capacity to implement curricula that place varying demands on and provide varying levels of support for their learning. A prior analysis of a stratified random sample of lessons from *EM* and *Investigations* reveals that the learning demands on teachers of these two curricula are indeed different (Stein & Kim, 2009). Although both offer tasks that are high-level and cognitively complex for students (as would be expected of standards-based curricula), they differ with respect to the kind of high level tasks that form the majority of their lessons. We use a classification scheme from prior mathematics instructional research (Stein, Grover, & Henningsen, 1996) to distinguish between two kinds of cognitively complex instructional tasks: "procedures with connections to concepts, meaning and understanding" (PWC) tasks and "doing mathematics" (DM) tasks. While DM tasks are less structured and do not contain an immediately obvious pathway toward a solution, PWC tasks tend to be more constrained and to point toward a preferred—and conceptual—pathway to follow toward a solution. Both PWC and DM tasks, however, place high cognitive demands on students, in that these tasks privilege the development of mathematical concepts and ideas over "fool-proof" methods that can be used to get to correct answers. We found that the majority of tasks in *EM* lesson materials used by teachers (79%) were PWC tasks while the majority of tasks in *Investigations* lessons used by teachers (89%) were DM.

These findings suggest that, perhaps, there is some truth to the conventional wisdom that *Investigations* is a more challenging curriculum for teachers. Indeed, past research demonstrates that DM tasks are faithfully implemented less often than are PWC tasks (Stein et al., 1996). Why? Case studies revealed that DM tasks open up the discourse space in sometimes "hard-to-manage" ways for teachers (Henningsen & Stein, 1997). Because a pathway to the solution is not specified, students approach these tasks in unique and sometimes even bizarre ways. Teachers must not only strive to understand how students are making sense of the problem but also begin to align students' disparate ideas and approaches with canonical understandings about the nature of mathematics. Research by other mathematics educators confirms that this is very difficult for most teachers to do (Ball, 2001; Chazen & Ball, 2001; Lampert, 2001; Leinhardt & Steele, 2005; Sherin, 2002).

PWC tasks, on the other hand, channel the route of student thinking along a finite number of pathways (often just one). Because the "learning route" is constrained, the space of the classroom discourse can be expected to be less open. Although research demonstrates that such tasks are susceptible to losing the connection to meaning, we conjecture that the learning demands that PWC tasks place on teachers are more tractable because the variety of student responses that teachers might expect to encounter is bounded and more predictable. Since *EM* comprises primarily these kinds of tasks, whereas *Investigations* comprises primarily DM tasks, it seems reasonable to predict that—all things being equal—teachers will have less difficulty learning to teach with *EM*.

Running counter to the above line of reasoning, however, is a second set of findings regarding opportunities for teacher learning that are embedded in each curriculum (Stein & Kim, 2009). In a nutshell, the *Investigations* curriculum offers more support to teachers than does the *EM* curriculum, where "support" is defined as additional information written specifically *for teachers* in order to help them better understand and teach the lessons. For example, 80% of the *Investigations* lessons sampled by Stein and Kim (2009) provided rationales regarding the mathematics behind the tasks that students were asked to do and 91% helped teachers to anticipate how students might respond to the tasks. The overall difficulty level of the DM tasks coupled with the large amount of supporting materials for teachers led us to identify *Investigations* as a high-demand, high-support curriculum.

Less support for teacher learning was found in the *EM* materials (Stein & Kim, 2009). The rationale for lessons was elaborated in only 21% of the sampled lessons and assistance with anticipating how students might respond to tasks was provided in only 28% of the lessons. Although more in-depth discussions of important mathematical ideas can be found in the *Teacher's Reference Manual*, this manual was read by a small percentage of teachers in our study.[1] The preponderance of PWC tasks coupled with the limited mathematical explanations in the main book (the *Teacher's Lesson Guide*) led us to identify *EM* as a low-demand, low-support curriculum.[2]

This second set of findings complicates the conventional wisdom. Without the findings on teachers' opportunity to learn, one would expect teacher capacity to be more strongly associated with quality of implementation in districts using *Investigations* (i.e., high-capacity teachers would be expected to fare better than low-capacity teachers). In districts using *EM*, on the other hand, one would expect little correlation between teacher capacity and instructional quality (i.e., teachers with low and moderate capacity, as well as teachers with high capacity, should theoretically be able to implement the curriculum well). The educative possibilities offered by the *Investigations* curriculum, however, suggest the need to pay attention to more than just demand on teacher learning exerted by the curriculum and to also attend to the opportunities for teacher learning that are embedded in the curriculum. However, opportunities are just that: opportunities that teachers may or may not take advantage of. Thus, we argue that the real issue dividing high-quality from low-quality implementers of *Investigations* would be *the extent to which teachers take advantage of the opportunities that are presented in the curriculum materials*.

This takes us into a little-researched, but growing, area of curriculum research: the interaction between curriculum features and how teachers use those features (Remillard, Lloyd, & Herbel-Eisenmann, 2009). Brown (2009) argues that understanding how teachers implement a curriculum requires an integrated analysis of curriculum resources, the resources the teacher brings to the task of interpreting the curriculum (teacher capacity), and how the two interact. Building on the concept of "mediated action" (Wertsch, 1998) as applied to teachers' use of the curriculum (Brown, 2009), we assume that teachers and curriculum materials are engaged in a dynamic interrelationship in which each participant (teacher and text) shapes the other; together they shape instruction.[3]

To further understand how teachers' instruction is influenced by both teacher capacity and teachers' use of the curriculum, our study addresses the following research questions:

Research Question 1: How does teachers' quality of implementation differ in comparisons between the two mathematics curricula (*Everyday Mathematics* and *Investigations*)?

Research Question 2: To what extent are teachers' capacity and their use of curricula correlated with the quality of their implementation, and do these correlations vary in comparisons between the two mathematics curricula?

Additionally, because we hypothesize that patterns of curriculum use may vary, depending on the opportunities for teacher learning in the curricula and/or district context, our study also addresses a third and final question to shed some explanatory light on any correlations between use of curricula and quality of implementation:

Research Question 3: How are teachers' patterns of use different in comparisons between the two curricula?

In our work, we use *EM* as an exemplar of a low-demand, low-support curriculum and *Investigations* as an exemplar of a high-demand, high-support curriculum. Our ultimate aim is to draw implications for local decision makers

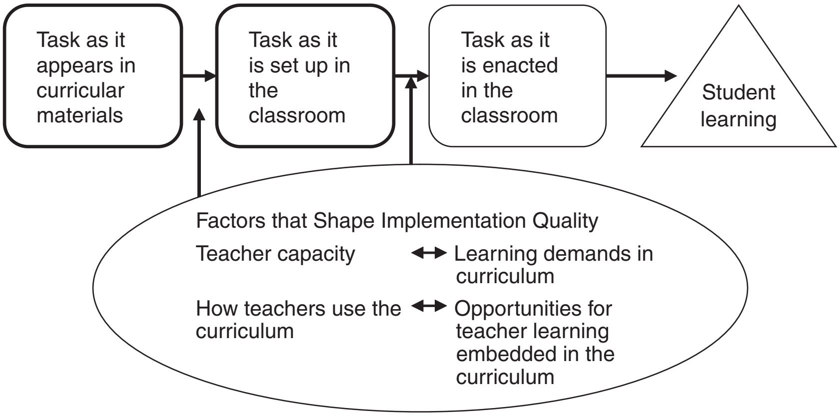


Figure 1. **Phases of curricular task implementation and factors that shape it.**

regarding how to select and how to support the use of mathematics curricula for district-wide improvement efforts.

We should emphasize from the outset that our analysis does not consider how use of *EM* or *Investigations* impacts student learning. Instead, our work focuses on the less-studied link between curricula and instruction. Many attempts have been made to directly link curricula and student learning without considering teacher instruction, despite a large body of research demonstrating that teacher instruction has a major impact on student learning (Cohen & Hill, 2001; Gallagher, 2004; Heneman, Milanowski, Kimball, & Odden, 2006; Rowan, Correnti, & Miller, 2002). But, ironically, those studies that examine teacher instruction and its impact on student learning do not typically focus on the role that the curriculum plays in teacher instruction (Agodini et al., 2009; Riordan & Noyce, 2001; What Works Clearinghouse, 2007). Finally, the studies that have focused on the relationship between standards-based mathematics curricula and instruction (Manouchehri & Goodman, 1998; Tarr, Chavez, Reys, & Reys, 2006) have not emphasized the interaction between features of different standards-based curricula and teachers' capacity to use those features for their instruction.

The remainder of this article unfolds in four sections. First, we provide a framework for identifying high-quality implementation and the factors that shape it. The second section explains our methodology. In the results section, we compare the quality of implementation in two districts—one using *EM* and one using *Investigations*—and then we consider the extent to which teachers' capacity and their use of the curriculum are associated with that implementation quality. Additionally, we delve into qualitative data to more deeply examine how teachers from both districts use the curricula and why that use differs for *EM* and *Investigations*. Our final section concludes with a discussion of implications for policy and practice as well as for research on curricular implementation as a method of scaling up reform.

High-Quality Implementation and the Factors That Shape It

Despite the fact that policymakers often assume a direct relationship between the adopted and enacted curriculum, a variety of studies show that curricula are seldom implemented as intended by their designers (Stein, Remillard, & Smith, 2007). Drawing upon the work of Walter Doyle (1983, 1988) and our own past work (Stein et al., 1996; Stein & Lane, 1996), we frame various sets of possibilities regarding how and where the instructional tasks contained in *EM* and *Investigations* might be altered throughout the course of a lesson.

The framework in Figure 1 depicts the various phases that an instructional task goes through: first as it appears on the pages of a written curriculum, then as the teacher announces or sets up the task inside the classroom, and, finally, as the task is actually enacted in the classroom by students and the teacher.

Based on prior research, we know that the features of an instructional task, especially its cognitive demands, change as the task passes through these phases (Stein et al., 1996; Stigler & Hiebert, 2004). In order to track changes in cognitive demand, tasks can be classified at one of several levels of cognitive demand. These classifications have been developed to measure multiple different types of mathematics curricula across district settings (Stein et al., 1996). As noted earlier, there are two classifications for high-level tasks: DM and PWC tasks. Low-level cognitive demand tasks include procedures-with**out**-connections to underlying meaning or concepts (activities that ask students to perform a set of routinized procedures without knowing why they are doing them or anything about the underlying meaning associated with the operations that they are performing); memorization; unsystematic/nonproductive exploration (a mode of enactment in which students are attempting to work their way through a problem but are making no progress in connecting to the main mathematical idea); and no mathematical activity (students are not attending to the mathematics).

What Constitutes a High-Quality Lesson?

Maintaining cognitive demand. A high-quality lesson is one that begins with a high-level task and that maintains the high level of cognitive demand. Although the cognitive demand of tasks can decline between any of the phases, it is especially important that a high level of demand be maintained through the enactment phase because this phase represents how students actually engage with the task and hence their opportunities to learn what was intended. A commonly observed instructional pattern, however, is for tasks to begin with a high level of cognitive demand (DM or PWC) but then to decline—either in terms of how the teacher sets them up or in terms of the levels of cognitive processing in which students actually engage (Stein et al., 1996). As was indicated earlier in this article, both DM and PWC curricular tasks will—if maintained—constitute high-quality lessons. However, PWC tasks may be less challenging for teachers to maintain because those tasks are more bounded and easier for teachers to grasp and use for instruction.

Attending to student thinking. Both curricula stress the importance of teachers paying close attention to what students do and say as they work on problems so as to be able to uncover and understand their mathematical thinking (e.g., Brendefur & Frykholm, 2000; Hodge & Cobb, 2003; Lampert, 2001; Nelson, 2001; Schoenfeld, 1998; Shifter, 2001). This is commonly done by circulating around the classroom while students work (e.g., Baxter & Williams, in press; Boerst & Sleep, 2007; Hodge & Cobb, 2003; Lampert, 2001). An important goal is to identify the mathematical learning potential of particular strategies or representations used by the students, thereby honing in on which student responses would be important to share with the class as a whole during the discussion phase (Brendefur & Frykholm, 2000; Lampert, 2001; Stein, Engle, Smith, & Hughes, 2008).

Vesting intellectual authority in mathematical reasoning. Both curricula also endorse the view of mathematics classrooms as places where students are "authorized" to solve mathematical problems for themselves, by employing mathematical reasoning rather than relying on the teacher or text (Engle & Conant, 2002; Hamm & Perry, 2002; Lampert, 1990b; Scardamalia, Bereiter, & Lamon, 1994; Wertsch & Toma, 1995). A learning environment embodying the norm of accountability to the discipline regularly encourages students to "account" for how their ideas make contact with those of other mathematical authorities, both inside and outside the classroom (see also Cobb, Gravemeijer, Yackel, McClain, & Whitenack, 1997; Lampert, 1990a; Michaels, O'Connor, Hall, Resnick, & Fellows of the Institute for Learning, 2002).

In summary, a high-quality lesson taught from either of these two curricula would be a lesson in which a high level of cognitive demand is maintained through the enactment phase, in which the teacher attends to student thinking and uses student responses to move the class toward the mathematical goals of the lesson and, finally, in which students are encouraged to solve problems and justify their strategies using mathematical reasoning. Our analysis shows that these three measures of quality implementation are significantly correlated with one another for lessons in either school district ($p < .01$ in both Region Z and Greene), which suggests that these measures work together to demonstrate quality implementation using the curricula in either school district.[4]

What Factors Shape Implementation Quality?

The oval in Figure 1 identifies two teacher-related factors—both of which are interactive with specific curricular features—that potentially shape the quality of mathematics lessons. The first factor is teacher capacity. We use the term *capacity* to refer to teachers' education, experience, and knowledge of mathematics for teaching. An instructional task may be well conceived and well designed in the curriculum (the first box of Figure 1). However, as a teacher prepares for the lesson, a limited understanding of the mathematics involved may lead him or her to fail to recognize the mathematical integrity of the task, thereby altering it in ways that (unintentionally) change (and often reduce) the level of cognitive demand of the task. Similarly, as the task is actually being carried out by students in the classroom, teachers who do not appreciate the mathematical insight to be gained from students' devising various routes through the problem space may inadvertently short circuit students' opportunities to learn by showing students how to find the answer, often in an algorithmic (vs. meaningful) way. While most studies examine the influence of teacher capacity on instruction without regard to the teacher learning demand of the curricular materials, we propose that high levels of teacher capacity would be especially needed to implement high-demand curricula such as *Investigations,* which require teachers to know the mathematical terrain through which students are wandering as they propose ways of solving open-ended, unstructured tasks.

The second teacher-related factor that can potentially shape the quality of mathematics lessons is how teachers actually *use* the materials written for them that appear in the curriculum. As noted earlier, the mathematical rationales for various curricular tasks that appear in the *Investigations*, curriculum invite teachers to prepare for lessons in thoughtful, mathematically rich ways.[5] But they will not produce high-quality lessons in and of themselves. They must be noticed and "used" by the teacher as he or she plans and carries out the lesson. Some teachers actively look for the mathematical point of the lesson and for information regarding how students might respond to the various tasks; others do little more than check for the materials that they need to carry out the lesson (e.g., manipulatives, graph paper) and the activities that students are supposed to do.

We propose that the sophistication levels for how teachers use the curricular materials will differ across the two curricula because of their different affordances (see earlier findings of curricular analysis). Moreover, sophistication of use should be especially related to lesson quality in *Investigations* because teachers who have taken the time to read the support materials will be better prepared to guide student learning through the difficult terrain represented by DM tasks. Teachers who use *EM*, on the other hand, have fewer supports for their learning and (perhaps) less need for such supports because of *EM*'s lower demands on teacher learning. Few studies examine teachers' use of curricular materials as it relates to implementation quality; we know of no studies that compare the relationship between use and implementation quality for two curricula that differ in the kind and number of opportunities for teacher learning as do these two curricula.

Table 1
Number of Coded Classroom Observations per Year, Season, and School

	Number of Observations			
	2004–2005 School Year		2005–2006 School Year	
	Fall	Spring	Fall	Spring
Region Z—School A	15	14	16	15
Region Z—School B	18	17	15	14
Region Z—School C	17	17	16	14
Region Z—School D	18	17	9	9
Greene—School E	18	18	15	15
Greene—School F	18	18	18	18
Greene—School G	17	18	18	18
Greene—School H	18	15	14	14
Total	139	134	121	117
Grand total = 511				

Methods

Setting

The data used in this study were from the two districts selected for a larger NSF-supported research project on how districts scale up elementary mathematics curricula. Greene[6] is an urban K–8 district that includes 16 elementary schools and serves about 20,000 students. About 87% of the students are Hispanic, 86% students are eligible for free or reduced-price lunches, and 50% are English Language Learners (ELLs). The second district was actually a subdivision of New York City: Region Z.[7] Within Region Z, we focused on one school network that included 10 elementary schools, each with a school population ranging from 400 to 800 pupils. Approximately 60% of the students are African American and 35% Hispanic. About 88% of the students receive free or reduced-price lunches, and 10% of the students are ELLs.

Both Region Z and Greene began district-wide implementation of new elementary mathematics curricula in fall 2003, Region Z using *EM* and Greene using *Investigations*. This study reports on data collected between August 2004 and June 2006, hence the information presented herein reflects the second and third years of each district's scale-up effort.

Data Sources

Classroom data. The classrooms that were examined in this study come from eight case-study schools that were selected (four within each district) at the beginning of the study based on recommendations from the district director of mathematics. The schools were selected to represent varying levels

Table 2
Number of Coded Classroom Observations per Grade Level

	Region Z	Greene
Kindergarten	33	54
First grade	30	43
Second grade	51	46
Third grade	47	49
Fourth grade	28	31
Fifth grade	52	47
Total	241	270

of teacher professional community and teacher expertise. All eight had a high percentage of students who qualified for free or reduced-price lunch. Consistent with the demographics of their region, schools in Region Z had a majority of African American students and schools in Greene had a majority of Latino/a students. The schools in Greene also had a higher percentage of ELLs than those in Region Z.

In each case-study school, six teachers (representing the span of elementary grade levels) were selected for classroom observations (three consecutive lessons each fall and three each spring) and interviewing (a pre- and post-interview surrounding each three-lesson set). Altogether, 511 classroom observations were collected between August 2004 and June 2006 and later coded with regard to teachers' implementation of the mathematics curricula (see Tables 1 and 2).

All classroom observations were conducted by trained observers who took detailed field notes and then completed prespecified, qualitative write-ups upon leaving the classroom. The write-ups included a comprehensive lesson summary and answers to a set of questions about cognitive demand, teachers' attention to student thinking, and the location of intellectual authority during the lesson. Answers were required to be backed up by one or more examples from the lesson. These same individuals conducted the pre- and post-lesson interviews, which were audiotaped and transcribed. The interviews contained questions about how teachers prepared for the lesson, including what parts of the curriculum they consulted and what they talked with colleagues about.

Each lesson (along with its associated interview transcripts) was then coded (see coding sheet and decision rules in Appendix B) by one of a group of four trained master's- or Ph.D.-level mathematics educators, all of whom were familiar with the research on cognitive demand. The sources of data that informed the coding for each lesson included the classroom write-up, the artifacts from the lesson, and the transcript of the pre- and post-interview.[8]

In order to prevent coding "drift," the coders met with the authors on a monthly basis to share codes for a randomly selected lesson. These 1- to 2-hour meetings produced 10 "consensus coded" documents plus refinements of the decision rules. In addition, another 9% of the lessons were double coded with an interrater reliability of 75%. For each double-coded lesson, differences were resolved and a consensus code was entered.

The coded variables that are examined for this article included the extent to which the teacher used the lesson plan from the district curriculum; what teachers reviewed in the curriculum and talked about with others in preparation for teaching the lesson; the cognitive demand of the main instructional task of the lesson as it appeared in the curriculum, as it was set up by the teacher, and as it was enacted by students in the classroom; the extent to which teachers worked to uncover and productively use student thinking in the classroom; and the extent to which intellectual authority was vested in mathematical reasoning vs. the teacher or textbook.

Survey data. Surveys were administered to all K–5 teachers in Greene and in the selected network of Region Z in the spring of 2005 and the spring of 2006. Math coaches administered the survey in a group setting when possible; otherwise, teachers completed the surveys on their own time and returned them to their coach. Teachers who completed a survey were given a $10 gift certificate to a bookstore. Here we draw on the survey responses of the observed teachers for whom we have classroom observation data. Of the 48 observed teachers, 46 teachers completed the survey in one or both years. We thus have a total of 86 completed surveys for observed teachers, 44 from Region Z and 42 from Greene.

The survey was developed by the research team to collect information about teachers' knowledge of mathematics for teaching, social capital, curriculum use and implementation, instructional practices, and district strategies (Yuan, Lockwood, Hamilton, Gill, & Stein, 2008). Many of the survey items were drawn from surveys used in other large-scale research on teaching, such as the 2000 National Survey of Science and Mathematics Education (Banilower, Smith, & Weiss, 2002) and the Study of Instructional Improvement (Hill, Schilling, & Ball, 2004). For this study, we drew upon survey items measuring the following variables: teachers' knowledge of mathematics for teaching, teachers' perceived usefulness of the curriculum, the education and experience levels of teachers, and teachers' self-reports of hours of mathematics professional development attended per year. The survey items used in our analysis are included in Appendix C.

Analysis Procedures

Defining quality of implementation. Quality of implementation was defined by three constructs for each lesson: the maintenance of a high level of cognitive demand from the materials phase to the enactment phase of the lesson, the level and kind of attention that the teacher paid to student thinking, and the extent to which the intellectual authority in the classroom was vested in mathematical reasoning (vs. the text and the teacher). More specific information about the scoring for these and all variables used in our analysis can be found in Appendix A.

Defining teacher capacity. Four data sources were used to measure teacher capacity. First, we took into account teachers' responses to a series of survey items that were specifically designed to measure knowledge of mathematics for teaching (KMT; Hill et al., 2004). In particular, we used a set of items that focused on teachers' knowledge of content and students. Items were in the form of instructional scenarios about which teachers were asked a series of forced choice questions. Second, we took into account teachers' survey reports of their education. Third, we included teachers' survey reports of their years of teaching experience. Finally, we included teachers' survey reports of their hours of mathematics professional development overall, as well as their reports of hours of professional development in specific areas associated with mathematics curriculum reform: in-depth study of mathematics content, methods of teaching mathematics, and students' mathematical thinking.

Defining teacher use of curriculum. To investigate teachers' use of curriculum, we first took into account their perceptions about the usefulness of the curriculum, as measured by a composite in our survey. We also took into consideration the percentage of the time that teachers used the curriculum in their lessons, as determined through classroom observations. Finally, we took into account teachers' reports in interviews regarding what they reviewed in the curriculum and what they talked about with others in preparation for their lessons. We specifically noted teachers' reports about what they reviewed and discussed with others in the following three categories: non-mathematical details focused on temporal and structural elements of lessons; materials needed for lessons; and big mathematical ideas in lessons, defined as any teacher talk about lessons that moves beyond basic activities within the lessons and articulates concepts or ideas that are at the heart of the lesson.

Determining relationships between implementation quality, teacher capacity, and use of curriculum. Correlations were calculated in two different ways depending on the source of data:

- *Correlations between lesson quality and information gained from annual surveys*. Surveys were conducted in the spring of 2005 and 2006. Thus, the correlation between information from the surveys and lesson quality was computed using the assigned score from the survey and the average lesson quality score across the entire preceding year (i.e., the average of the 6 lessons observed for a teacher in 2004–2005 was correlated with the same teacher's responses to the 2005 survey and the average of the 6 lessons observed in 2005–2006 was correlated with the same teacher's responses to the 2006 survey). This lowered the size of our sample from 511 (number of observed lessons) to 90 (number of teachers observed each of 2 years).
- *Correlations among lesson quality variables*. For any correlations drawn between variables within a lesson, we were able to maintain our 511 sample size.

Table 3
Means and Standard Deviations for Quality Implementation Variables

	Region Z (*Everyday Mathematics*)		Greene (*Investigations*)	
	N	*M* (*SD*)	*N*	*M* (*SD*)
Quality implementation				
Cognitive demand, materials to setup	210	2.8 (1.2)	251	3.5*** (0.9)
Cognitive demand, setup to enactment	220	2.1 (1.3)	250	3.2*** (1.1)
Total cognitive demand	210	4.9 (2.3)	250	6.7*** (1.8)
Student thinking	241	0.5 (0.6)	268	1.1*** (0.7)
Intellectual authority	241	0.4 (0.6)	269	1.2*** (0.7)

***p < .001.

Additionally, we assessed the relationship between implementation quality and variables derived from our interviews on teachers' use of curriculum, specifically in regard to what in the curriculum lessons they reviewed and what they talked about with others. These variables derived from our interviews were binary (e.g., teacher did or did not review the big mathematical ideas in the curriculum prior to instruction) rather than continuous. Because Pearson correlations are best computed between two continuous variables, we performed independent *t* tests comparing the implementation quality for those teachers in each of the two categories for binary variables (e.g., teachers who did review the big mathematical ideas and teachers who did not). Because interview codes were assigned for a set of three lessons, *t* tests were performed using the average lesson quality score across the three lessons accompanying that interview. This lowered the size of the sample by a third.

In addition, because our initial lesson coding revealed that teachers' review of the big mathematical ideas in the curriculum prior to instruction was connected with higher quality instruction (as described in the following results section), we engaged in a second round of qualitative analysis of the teacher interviews to understand better teachers' responses regarding how they reviewed the big mathematical ideas and the constraints and affordances offered by *Investigations* versus *EM* for their ability to do so.

Results

Quality of Implementation

Greene School District teachers' implementation of *Investigations* was considerably better than Region Z teachers' implementation of *Everyday Mathematics*. Table 3 provides means and standard deviations for the

Table 4
Means and Standard Deviations for Teacher Capacity and Use of Curricula

	Region Z (*Everyday Mathematics*)		Greene (*Investigations*)	
	N	*M* (*SD*)	*N*	*M* (*SD*)
Teacher capacity				
Knowledge of mathematics for teaching (KMT)	39	5.7 (2.6)	41	6.4 (2.7)
Education	25	4.3 (0.6)	25	4.5 (1.2)
Years of experience	25	7.8 (2.8)	25	6.6 (3.1)
Mathematics PD hours/yr	34	10.4 (15.5)	27	26.1** (23.6)
Hours of PD on in-depth study of math content	34	3.9 (4.0)	38	3.5 (4.7)
Hours of PD on methods of teaching mathematics	32	5.1 (4.6)	37	3.7 (3.8)
Hours of PD on students' mathematical thinking	32	3.9 (3.0)	37	4.2 (4.4)
Teacher use of curriculum lessons				
Use of curriculum in daily lessons	241	2.6 (0.8)	270	2.8* (0.7)
Review of non-mathematical details	61	0.2 (0.4)	83	0.3 (0.5)
Review of materials needed	61	0.7 (0.5)	83	0.8 (0.4)
Review of big mathematical ideas	61	0.2 (0.4)	83	0.7*** (0.5)
Talk with others about non-mathematical details	50	0.2 (0.4)	79	0.2 (0.4)
Talk with others about materials needed	50	0.2 (0.4)	79	0.1 (0.3)
Talk with others about big mathematical ideas	50	0.0 (0.1)	79	0.2** (0.4)
Perceptions of curriculum's usefulness	37	3.1 (0.8)	41	3.6** (0.6)

Note. PD = professional development.
*$p < .05$ in independent t-test comparisons between districts. **$p < .01$. ***$p < .001$.

variables that we used to define quality implementation: cognitive demand, student thinking, and intellectual authority.

In the table, cognitive demand data are presented in three ways for lessons: as a 1–4 rating for teachers' maintenance of high cognitive demand from curriculum materials to the setup phase in the classroom; as a 1–4 rating for teachers' maintenance of high cognitive demand from the setup phase to enactment in the classroom; and as a total cognitive demand score that is the sum of the two phases of teachers' instruction (materials to setup + setup to enactment; range of 2–8). Again, more complete definitions for all variables used in this analysis can be found in Appendix A. As can be noted in the table, ratings for cognitive demand were much higher for Greene teachers compared to Region Z teachers ($p < .001$). This finding challenges the conventional wisdom that *Investigations* may be more difficult to implement than *EM* because it contains more DM tasks, which are typically more complex and less bounded than PWC tasks.

As is also indicated in the table, teachers in Greene also received much higher ratings of their work to uncover student thinking in the classroom compared to teachers in Region Z ($p < .001$). In fact, a little over half (51%) of the *EM* lessons in Region Z evidenced no teacher attempt to uncover student thinking, while only 18% of *Investigations* lessons in Greene evidenced no teacher attempt to uncover student thinking. In contrast, in the majority of *Investigations* lessons (61%), the teacher did at least some work to uncover student thinking through questioning strategies and arrangements for public sharing of student responses. Similarly, teachers of *EM* lessons struggled to ensure that intellectual authority was vested in mathematical reasoning compared to teachers of *Investigations* ($p < .001$). While 83% of the teachers in Greene demonstrated some reliance on mathematical reasoning as the source of intellectual authority, only 38% of Region Z teachers did.

Teacher Capacity and Use of Curricula

Table 4 reports the means and standard deviations for two major categories of variables that we hypothesize to influence implementation quality: teacher capacity and teacher use of curriculum. Teacher capacity is represented by four variables measured in our survey: teacher knowledge of mathematics for teaching (KMT), teacher education, years of teaching experience, and mathematics professional development (PD) hours per year. For teacher education and years of teaching experience in the table, we include only the most recent survey response. For KMT, as well as mathematics PD hours per year, we report the cumulative average of teachers' reports over 2 years.

For teachers' use of curriculum, we report on the three major categories for what teachers reported reviewing or talking about with others in their interviews: non-mathematical details of the lesson, materials needed, and big mathematical ideas. If the teacher did not use the curriculum materials for any lessons in his or her lesson cluster, we did not include that teacher's response in our analysis. To understand teachers' use of the curriculum more deeply, we also include means and standard deviations for teachers' use of the curriculum (as rated in lesson observations) and teachers' perceptions of the usefulness of the curriculum (as a composite drawn from the teacher survey).

In comparing teacher capacity in Region Z and Greene, the only variable demonstrating a significant difference between districts is teachers' reports of mathematics professional development hours per year, with *Investigations* teachers reporting much higher hours of mathematics professional development than *EM* teachers ($p < .01$). That said, teachers' responses did not uncover significant differences in regard to hours of professional development for specific topics associated with mathematics reform implementation. In regard to curriculum use, however, we observed several areas of significant differences between districts. First, Greene teachers used *Investigations* materials in their lessons more than Region Z teachers used *EM* materials ($p < .05$). Moreover, when Greene teachers used supplementary materials in their lessons, they were more likely to use materials that were congruent with *Investigations* as compared to the Region Z teachers ($p < .05$). Second, Greene teachers reported higher perceptions of the usefulness of the *Investigations* curriculum compared to teachers' perceptions of *EM* in Region Z ($p < .01$). These higher perceptions of curriculum usefulness in Greene might be seen as an important prerequisite for sophisticated use of curricula.

More important, perhaps, are the different ways in which Greene vs. Region Z teachers used their respective curriculum. While we did not observe significant differences between districts regarding teachers' reports of reviewing or talking about non-mathematical details or needed materials prior to lessons, *Investigations* teachers reported reviewing and talking with others about the big mathematical ideas in the lessons significantly more than *EM* teachers ($p < .001$ for reviewing and $p < .001$ for talking). This suggests a more thoughtful preparation on the part of Greene teachers. However, the reader is reminded that the *Investigations* curriculum provides more support for teachers in this regard than does *EM*. Thus, one might view these findings as representing an interaction between what the curriculum affords and teachers' preparatory routines, which might be further enhanced by mathematics professional development in which Greene teachers participate and the higher perceptions of curriculum usefulness among Greene teachers.

The Relationship Between Implementation Quality and Variables Measuring Teacher Capacity and Use of Curriculum

Table 5 provides correlations between implementation quality (as measured by total cognitive demand, student thinking, and intellectual authority) and variables measuring teacher capacity and use of curriculum.[9] As shown in Table 5, the relationship between teacher capacity and quality of implementation did not play out as expected. In Greene, our expectation that higher capacity teachers would be able to implement the curriculum better was not conclusively demonstrated. While Greene teachers' KMT was positively correlated with cognitive demand, student thinking, and intellectual authority, none of those relationships were significant. Furthermore, Greene teachers' levels of education and years of teaching experience were negatively correlated with intellectual authority. That said, professional development hours reported by Greene teachers were significantly correlated across all quality implementation measures. Additionally, professional development in topic areas associated with reform implementation—in-depth study of mathematics content, methods of teaching mathematics, and students' mathematical thinking—were significantly associated with student thinking and intellectual authority measures for teachers in Greene.

As with Greene teachers, Region Z teachers' education and experience levels were not related to quality of implementation. Region Z teachers' KMT, however, was *inversely* related to student thinking and intellectual authority, meaning that more knowledgeable teachers were less apt to uncover and use student thinking and also less likely to invest intellectual

Table 5
Pearson Correlations Between Implementation Quality and Variables for Teacher Capacity and Use of Curriculum

	Region Z			Greene		
	Total Cognitive Demand	Student Thinking	Intellectual Authority	Total Cognitive Demand	Student Thinking	Intellectual Authority
Knowledge of math for teaching (KMT)	–.10	–.32*	–.33*	.11	.04	.20
Education	–.18	.25	.03	–.06	–.16	–.17
Years experience	.14	.12	.11	–.24	–.18	–.31*
Mathematics PD hours/yr	.09	–.09	.05	.42*	.63**	.57**
PD on in-depth study of math content	–.15	–.16	–.08	.27	.41*	.34*
PD on methods of teaching math	–.08	–.18	–.15	.23	.36*	.19
PD on students' mathematical thinking	–.06	–.09	–.23	.23	.44**	.31
Use of curriculum	.08	.09	.00	.22**	.06	.10
Perceptions of curriculum usefulness	.14	.03	.19	.41**	.34*	.47**

Note. PD = professional development.
$^{*}p < .05$. $^{**}p < .01$.

authority ($p < .05$) in classroom-based mathematical reasoning. Additionally, professional development hours were not related to quality implementation in Region Z. None of our teacher capacity measures were therefore significantly tied to quality implementation when looking across both implementation of *Investigations* curriculum in Greene and implementation of *Everyday Mathematics* curriculum in Region Z.

Additionally, our measures for use of curriculum and perceptions of curriculum usefulness are not significantly correlated with quality implementation of both *Investigations* in Greene and *Everyday Mathematics* in Region Z. Instead, as with professional development, these measures are significantly correlated only with quality implementation in Greene. Taken together, these data suggest that Greene provided professional development that better prepared teachers to implement the *Investigations* curricula and, additionally, that Greene teachers' high perceptions of curriculum usefulness may be actual reflections of the usefulness of the curriculum, at least in terms of the measures we designated for quality implementation.

Only one of our variables—presented in Table 6—was associated with high-quality implementation across *both* curricula: review of big mathematical ideas in the curriculum. While review of and talk about the big mathematical ideas were more prevalent among teachers in Greene School District, lessons for which teacher preparation included a review of the big ideas paid off for *EM* as well as *Investigations*. In both districts, the more teachers prepared by trying to get a handle on the important mathematical ideas at play in the lesson—either by reviewing the big mathematical ideas or talking about them with others—the better they implemented two out of three of our measures of instructional quality: attention to student thinking and intellectual authority based on mathematical reasoning. As can be seen in the table, means for student thinking and intellectual authority were significantly higher among teachers who reviewed and talked about the big mathematical ideas compared to means of teachers who did not review or talk about the big mathematical ideas.

Teachers' reports of reviewing or talking about the non-mathematical details or materials needed for the lesson were not positively associated with our implementation quality measures across districts, although—interestingly—reviewing the non-mathematical details or materials needed was negatively associated with quality implementation in Region Z (not Greene). Unlike teacher capacity, then, *how teachers use the curriculum appears to shape the quality of their lessons*. In particular, when teachers talked about or reviewed big mathematical ideas that students were supposed to be learning in both Greene and Region Z, they tended to have a higher quality lesson.

The Relationship Between Curricular Materials and Teachers' Patterns of Use

While our analysis thus far demonstrates that looking for the big mathematical ideas predicts better quality implementation, much more so than

Table 6

Comparisons of Quality Implementation Means (*SD*s) for Teachers Who Have and Have Not Reviewed and Talked About the Big Mathematical Ideas (BMI) in Curriculum Lessons

	Region Z			Greene		
	Total Cognitive Demand	Student Thinking	Intellectual Authority	Total Cognitive Demand	Student Thinking	Intellectual Authority
BMI reviewed	5.7 (2.0)	0.9* (0.5)	0.7* (0.5)	6.7 (1.5)	1.2** (0.6)	1.3* (0.6)
BMI not reviewed	4.9 (1.8)	0.6 (0.4)	0.4 (0.4)	6.6 (1.2)	0.9 (0.4)	1.0 (0.5)
BMI talked about[a]	8.0 (—)	1.0 (—)	1.3* (—)	6.6 (1.5)	1.4* (0.6)	1.3 (0.4)
BMI not talked about	5.2 (1.8)	0.6 (0.4)	0.5 (0.4)	6.6 (1.5)	1.0 (0.5)	1.2 (0.6)

[a]No standard deviation provided for Region Z talk about big mathematical ideas because only one teacher reported talking about the big mathematical ideas.

*$p < .05$ in independent t tests comparing means of teachers who did and did not talk/review big mathematical ideas. **$p < .01$.

teacher capacity, our data also demonstrate that Region Z teachers indicated looking for the big mathematical ideas far less than Greene teachers. Because the *Investigations* curriculum has a higher percentage of DM tasks compared to the higher percentage of PWC tasks in *EM*, our research might thus suggest that DM tasks lend themselves to exploration of the big mathematical ideas more than the PWC tasks in *EM*. However, when we examined whether a DM or PWC task in either curriculum was associated with teachers' review of the big mathematical ideas, we found no significant correlations. This evidence suggests that there is something beyond the types of task within each curriculum that lends itself to a higher likelihood of *Investigations* teachers reviewing the big mathematical ideas. To probe further what features of curricula led to a higher number of Greene teachers reviewing big mathematical ideas compared to Region Z teachers, we turn now to a more in-depth examination of exactly how teachers combed the curricula for big mathematical ideas and how that process differed across the two curricula.[10]

Number of mathematical foci in a lesson. The most consistent theme differentiating the Region Z teachers and Greene teachers was their perception of the amount of focus provided within the two sets of curricular materials. One Region Z teacher's description of the *EM* curriculum underscores the plethora of activities and ideas found in close proximity to one another:

> . . . [*EM* is] all over the place. "Hello, here's a ruler," the next day "and now we're looking at the nickel, remember the nickel? Great, 'cause here's some dominos." And [the students are] just like, "whoa," but then it comes back again a million times.

In addition to conveying the lack of focus in day-to-day lessons, two other features are noteworthy about this teacher's words. First, as this teacher correctly notes, *EM* lessons are designed to shift rapidly from one topic to the next. As with other spiral curricula, big ideas are presented—at varying levels of difficulty and in various ways—over and over again as the student progresses through the year and through the elementary grades. Rather than spending an entire unit comparing fractions, the child will visit and revisit the idea, first as a "beginning goal," then as a "developing goal," and finally as a "secure goal." Thus, any given set of lessons moves from a snippet of one big idea to a snippet of another, often without full development or closure. According to *EM* curriculum developers, this spiral organization reflects research that children learn best "when new topics are presented at a brisk pace, with multiple exposures over time" (see http://everydaymath.uchicago.edu/educators/faqs.shtml). However, our findings suggest that this spiraling makes extraction of big mathematical ideas difficult for the teacher on a lesson-by-lesson basis. Each lesson provides only a partial window into the developing mathematical idea, thereby making it difficult for the teacher to discern the bigger message toward which the lesson is aimed.[11] Second, the above quotation does not include any talk about ideas, but rather activities. In our judgment, this suggests that when the bigger landscape of the mathematical idea is hidden, the activities become foregrounded.

Further analysis of interviews with Region Z teachers corroborates the above teacher's comment about multiple foci within an *EM* lesson. In more than 50% of our interviews with *EM* teachers who reviewed the big mathematical ideas in lessons (8/14), they named multiple, unconnected topics—always more than two and sometimes more than four topics—that they planned to address in the lesson cluster (three classroom lessons) that we would be observing. Moreover, Region Z teachers, by and large, did not attempt to connect the topics in their description. Take, for example, one Region Z fifth-grade teacher's description of a 3-day set of lessons that she would be teaching:

> We're starting a unit on division. . . . The first lesson is going to be on splitting dividends to obtain a quotient. . . . The next day, on Wednesday, Lesson 4.2 is going to involve us applying an algorithm to the splitting of the dividends . . . and then on Thursday we discussed for Lesson 4.3, having the children use scales and ratios to measure to the nearest quarter of an inch to find distances from one place to the next.

In contrast, only 26% (14/54) of teachers of *Investigations* spoke about addressing more than one topic, and no Greene teachers said that they would be addressing more than two topics. Additionally, when Greene teachers spoke about addressing more than one topic, those topics were often discussed in a connected way, as with this third-grade teacher's description of the consecutive lessons that she would be teaching using *Investigations*:

> The first day, students will work on word problems in small groups or pairs. . . . They are going to decide whether the problem is a multiplication or division problem. . . . On the second day, they will use cubes and act out the problems. Then, they will decide which problem it is. . . . On the third day, they are going to write their own story problem and draw pictures for the problem. They will work on different ways of writing division and multiplication. I'll make a chart, if we have time, showing multiplication in one column and division in the other column and how the inverse property, you know, multiplication is an inverse of division. . . . I want them to use cubes and relate or transfer multiplication to division and vice versa.

This teacher discusses the 3-day sequence as an interconnected whole, with the big idea being that multiplication is the inverse of division.

Other teachers explicitly noted that the *Investigations* curriculum supports a focus on a single topic through multiple lessons, as this teacher who said, "Each *Investigation* has four or five activities that go with it and so the big idea—how I understand it is more those four lessons are with this big idea of math . . . as opposed to each lesson is about this, each lesson is about that." Thus, the focus upon a single concept within (and sometimes across) *Investigations* lessons helps teachers to locate and understand the big mathematical ideas within the curriculum, as is implied by the much larger number of Greene teachers who reported reviewing those big mathematical ideas prior to teaching their lessons.

Support embedded within lessons. Many Greene teachers noted the importance of examining the curriculum materials to prepare prior to teaching a lesson, implying that those materials offered a great deal of support for their teaching. For example, when asked whether she reviews the curriculum materials prior to class, one Greene teacher exclaimed,

> Whew! It makes a huge difference. . . . I can tell if I haven't gone over it very thoroughly because there's not the depth of conversation . . . the more prepared I am, the more I sit and reflect a little bit, the better the flow of the class. . . .

Many Greene teachers specifically talked about how the *Investigations* curriculum provided focus and clear objectives as well as support materials to help them (and their students) grasp the main concepts within a lesson. A kindergarten teacher, for example, noted that the Teacher Notes and Dialogue Box in the *Investigations* curriculum inform her about "what teachers have experienced in the past, how a five or six year old is thinking; it lets you anticipate some questions or problems that might come up." Another teacher noted how clearly the lessons are laid out, particularly in regard to the main mathematical concept within a lesson: "It [an *Investigations* lesson] has at the beginning, like, what is the math, what are you trying to get the kids to see. It has assessments for you, it tells you the mathematical emphasis. . . . It's just laid out very well."

On the other hand, teachers who reviewed the big mathematical ideas in Region Z indicated that they did so despite the lack of clarity in *EM* lessons. Four of these teachers specifically commented that the lessons do not provide enough background or scaffolding for students' learning of newer concepts and assume that all students were exposed to the curriculum in prior years or months. A fifth-grade teacher, for example, noted, "I had the fraction blocks [an *EM* manipulative] and we were comparing fractions . . . but we had never done fractions before. So, it was like, I thought it was much too challenging." No Region Z teachers noted any useful sections or language within *EM* that might help them get across the big mathematical ideas in a lesson.

The *Investigations* curriculum, therefore, appeared to offer more affordances for teachers to extract the big mathematical ideas in the lesson compared to the *EM* curriculum. The following factors pointed out by *Investigations* teachers likely supported their ability to locate and understand those big mathematical ideas: (1) a curricular focus on only a few topics within a set of lessons and some integration among those topics and (2) additional explicit information for teachers within the curriculum to help them grasp the big mathematical ideas in a lesson. In contrast, Region Z teachers were much more apt to talk about multiple topics addressed in *EM* lessons. Furthermore, when speaking about these multiple topics, Region Z teachers were less focused on the mathematical ideas in the lessons as opposed to the many activities within lessons. We hypothesize that this is the case because *EM* lessons often shift rapidly from topic to topic without presenting fully developed mathematical concepts that students investigate over the span of several lessons. Additionally, no Region Z teachers discussed support available within the *EM* curriculum for understanding the big mathematical ideas in the lessons.

Summary and Conclusions

We found that Greene teachers implemented the *Investigations* curriculum at a much higher level than Region Z teachers implemented *EM* in terms of cognitive demand, attention to student thinking, and vesting intellectual authority in mathematical reasoning. This finding was somewhat counterintuitive given research suggesting that the higher percentage of DM tasks within *Investigations* makes that curriculum theoretically more difficult to implement compared to *EM*, which contains a higher percentage of PWC tasks. Our findings on the relationship between traditional measures of teacher capacity and quality of implementation shed little light on Greene teachers' higher implementation of *Investigations*. That is, most of our variables measuring teacher capacity—teacher knowledge, education, and experience—were not strongly related to implementation across curricula. That said, professional development in Greene was significantly correlated with quality implementation, suggesting that Greene's professional development program provided more support to teachers than did Region Z's professional development.

While teacher capacity measures did not clearly relate to quality implementation, curriculum use provided a clearer picture of what might impact teachers' implementation of curricula across settings. Specifically, we found a strong relationship in both districts between teachers who reviewed the big mathematical ideas in the curriculum and teachers who implemented lessons at a high level. Furthermore, a much higher percentage of *Investigations* teachers reviewed the big mathematical ideas that they were preparing to teach compared to *EM* teachers. When we focused upon those teachers in both districts who reviewed the big mathematical ideas, in order to understand the reasons for their ability to focus on those ideas, we saw that Region Z teachers spoke about preparing to teach many more topics within any one *EM* lesson than their counterparts in Greene. Additionally, Greene teachers spoke a great deal about the clarity and support within *Investigations* lessons that helped them locate the big mathematical ideas, whereas Region Z teachers did not speak about useful supports in *EM* to help them locate those ideas. The significant correlations between hours of PD and high-quality implementation in Greene further suggest that PD may have helped teachers focus on the big mathematical ideas in the *Investigations* curriculum, which might in turn have influenced their curriculum use and perceptions about the usefulness of the curriculum.

Our results have a number of implications for districts trying to improve the quality of mathematics instruction and identify curricula to help them do that. First, our findings indicate that when trying to decide which curriculum to use in their district, administrators should consider what affordances curricula offer to help teachers locate and understand the big mathematical ideas within lessons. Our study specifically suggests the following curricular elements that administrators could look for when choosing a mathematics program that will help teachers work with big ideas within lessons: a focus on a single mathematical concept or idea within a curriculum lesson, clarity in presenting that mathematical concept or idea, and ample support and explanation within lessons that will help teachers present the concept to students and skillfully facilitate student thinking and discussion about that concept in the classroom. Second, our results suggest that whether or not the mathematics curriculum currently used by the district includes a focus on big mathematical ideas within a lesson and supports for teachers to teach that idea, the district should implement a professional development program that will help teachers identify the big mathematical ideas in the curriculum and use that curriculum productively to teach high-quality lessons.

More work is needed to tease out the interrelationship between the features of the curriculum and teachers' instruction. Mathematics curricula beyond what was examined in this analysis may offer affordances and constraints for teachers' quality implementation that we have not considered. Additionally, teachers may need different kinds of support for teaching in other subject areas besides mathematics, and those supports might move beyond a clear presentation of the big ideas in a lesson.

Additionally, we did not examine myriad other factors that might influence teacher implementation beyond teachers' capacity and use of curricula. Many aspects of teachers' local environments that we did not examine may be responsible for the differences in teacher behavior described herein. After all, all of the *Investigations* teachers were in Greene while all of the *EM* teachers were in Region Z. Could other aspects of their district environments, including their approach to scaling up the curricula, have been related to the rather strong differences in use patterns and quality of implementation seen across the two sites? Both districts mounted a variety of teacher supports along with their new, mandated curricula including common planning periods, mathematics coaches, and the provision of written district guidelines such as pacing calendars and cross-references to state standards. The differences between the two districts along many of these dimensions are discussed in other project publications (Coburn & Russell, 2008; Stein & Coburn, 2008), the upshot being that, in addition to offering a more teacher-supportive curriculum, the Greene environment was more aligned and consistently focused on teacher learning from those curricular materials than was the Region Z environment. These differences certainly played into the above findings.

Finally, we should add the qualification that our small sample size allows us to report on correlations rather than findings based on causal analysis. Thus, any significant correlational relationships reported here could be influenced by other intervening variables. In spite of this drawback, we believe that our data demonstrate how curriculum materials are used by teachers and how that use relates to quality implementation in ways not previously studied in any education research.

Our objective for this article was not to demonstrate a causal relationship between curricula and student learning or demonstrate that *EM* is a better curriculum than *Investigations* or vice versa. Rather, our work provides evidence that one cannot draw a direct relationship between curriculum and student learning. In our alternative approach to the analysis of effective curriculum materials, we asked what elements of teacher capacity interact with particular curriculum features to influence what teachers do with curriculum. Thus, our focus is on which program leads to better instruction under what conditions. Said another way, if a school leader adopts a particular curriculum, to what features of curriculum—alongside what elements of teacher capacity—must the leader attend in order to implement that curriculum effectively?

Our findings debunk the conventional wisdom that only high-capacity teachers can use *Investigations* in a high-quality way and that teachers with more limited capacity might be able to use *Everyday Mathematics* in a high-quality way. More interestingly, our findings suggest that how a teacher uses a curriculum may be more important than the education, experience, and knowledge that he or she brings to the table. Perhaps another way of conceptualizing teacher capacity—as a teacher who has the capacity to seek out and productively use resources—may be in order (J. Greeno, personal communication, January 19, 2008). This conceptualization could place a stronger emphasis on the interplay between curriculum as tool and teachers' use of curriculum (Brown, 2009) and how that interplay influences instruction. Specifically, our data suggest that curricula may operate as a teaching tool that supports and enhances teacher practice, which might then further influence teachers' skilled use of that tool. Whereas we established a connection between curricula use and instructional quality in our article, future research could delve into whether curricular use changes over time as a result of improved instruction and how different curricular features afford or constrain that relationship.

While mathematics educators have always kept notions of teacher capacity front and center, the ways in which capacity has been studied (teachers' knowledge of mathematics for teaching) have assumed an individual unit of analysis, that is, the teacher by herself or himself as opposed to the teacher-in-interaction with the environment. Here we focus on one key aspect of that environment: the curriculum materials that teachers have been assigned to use. Teachers' use of materials—in interaction with those materials—appears to hold greater explanatory power, in this study at least, than does the more traditional way of defining teacher capacity.

Appendix A
Definitions for Variables Used in Analysis

Implementation Quality	
Cognitive demand	Each phase of the lesson is assigned a cognitive demand code. The phases of a lesson are: the lesson as it appears in the materials used for instruction, the lesson as it is setup/introduced in the classroom, and the lesson as it is enacted following the setup. The high-level cognitive demand codes assigned for each phase of a lesson are: Doing Mathematics and Procedures With Connections The low-level cognitive demand codes assigned for each phase of a lesson are: Procedures Without Connections, Memorization, Unsystematic or Nonproductive Exploration, and No Mathematical Activity. The latter two codes are only assigned to a lesson in setup or enactment.
Maintenance of cognitive demand, materials to setup	Based on coding of each observed lesson using the following scale: 1 point—The teacher maintained a low level of cognitive demand from one phase to the next. 2 points—The teacher transformed a task from a high level of cognitive demand to a low level of cognitive demand. 3 points—The teacher maintained a high level of cognitive demand between two phases but transformed the task from DM to PWC or from PWC to DM. Although the teacher still maintained a high level of cognitive demand, the nature of that cognitive demand essentially shifted in a way that was not consistent with the materials or the teachers' setup. Thus, a teacher received fewer points than if he or she had maintained the same type of high-level cognitive demand from one phase to another. 4 points—The teacher maintained the same high level of cognitive demand from one phase to another without transforming the task into another type of high-level demand or to a lower level of cognitive demand.
Maintenance of cognitive demand, setup to enactment	Based on coding of each observed lesson; coded with the same point system as cognitive demand, materials to setup (above)

(continued)

Appendix A **(continued)**

Total cognitive demand	Cognitive demand score for materials to setup + cognitive demand score for setup to enactment (possible scores from 2 to 8)
Student thinking	Based on coding of each observed lesson using the following scale: 1 point—The teacher did no work to uncover student thinking. 2 points—The teacher did some work to uncover student thinking, including asking students to publicly share their work. 3 points—In addition to #2, the teacher purposefully selected some students to share their work. 4 points—In addition to #2 and #3, the teacher connected or sequenced students' responses in a meaningful way.
Intellectual authority	Based on coding of each observed lesson using the following scale: 1 point—Judgments of correctness derived from teacher or text. 2 points—Judgments of correctness sometimes derived from teacher or text, but also some appeals to mathematical reasoning. 3 points—Judgments of correctness derived from mathematical reasoning.
Teacher Capacity	
Knowledge of Mathematics for Teaching (KMT)	Based on teachers' responses to 12 survey items (see sample items 9–14 in Appendix C). Possible range of 0–12. Cronbach alpha (based on our 2005 survey data for 798 teachers in Greene and Region Z) is .68.
Education	Sum composite based on teachers' survey response (items 2–7 in Appendix C) that includes one point for each of the following: Degree the teacher received (e.g., bachelor's, master's, Ph.D.) Additional credits acquired beyond final degree K–12, ESL, bilingual, math, and/or special education certificates (1 point for each)
Years of experience	Based on teachers' survey response to the question: "Including this school year, how many years have you been teaching?" (item 1 in Appendix C). Categorical responses changed to numerical responses: less than 1 year = .5; 6–10 years = 8; 11 years or more = 11.

Appendix A **(continued)**

Mathematics professional development (PD) hours per year	Sum of teachers' survey responses to questions about the hours of math-related PD received over the past summer and math-related PD received over the current school year (items 15–18 in Appendix C). Teachers' survey responses on hours of PD in specific topics related to the mathematics reform were reported by teachers categorically; categorical responses changed to numerical responses: none = 0; less than 4 = 2; 4–8 = 6; 9–16 = 12.5; more than 16 = 17 (item 19 in Appendix C).
Teacher Use of Curriculum Lessons	
Use of curriculum in daily lessons	Based on coding of each observed lesson using the following scale: 0 points—Teacher used the lesson plan provided by the math curriculum (MC) for 0% of the lesson. 1 point—Teacher used the lesson plan provided by the MC for 1–25% of the lesson. 2 points—Teacher used the lesson plan provided by the MC for 26–75% of the lesson. 3 points—Teacher used the lesson plan provided by the MC for 76–100% of the lesson.
Review of non-mathematical details regarding temporal and structural elements of lessons	Based on coding of teacher interviews prior to and following observed lessons with 0 points for no review and 1 point for review.
Review of materials needed for lessons	Based on coding of teacher interviews prior to and following observed lessons with 0 points for no review and 1 point for review.
Review of big mathematical ideas in lessons	Based on coding of teacher interviews prior to and following observed lessons with 0 points for no review and 1 point for review.
Talk with others about non-mathematical details regarding temporal and structural elements of the lessons	Based on coding of teacher interviews prior to and following observed lessons with 0 points for no talk and 1 point for talk.
Talk with others about materials needed for the lessons	Based on coding of teacher interviews prior to and following observed lessons with 0 points for no talk and 1 point for talk.

(continued)

Appendix A **(continued)**

Talk with others about big mathematical ideas in lessons	Based on coding of teacher interviews prior to and following observed lessons with 0 points for no talk and 1 point for talk.
Perceptions of curriculum's usefulness	Based on teachers' survey responses regarding their agreement with a series of statements (item 8 in Appendix C). Subscale scores were calculated by averaging over all items in the scale. Cronbach's alpha for this scale (based on our 2005 survey data, n = 798 teachers) is .85.

Appendix B
Classroom Observation Coding Instrument

Instructions to Coders:

Before you begin coding, read through the entire lesson observation and transcripts of the pre- and post-interviews (ignore questions in the post-interview that do not directly deal with the observed lessons you are coding). Then, closely read any sections of the lesson write-up or interviews specified below for a code. In assigning your code, rely upon your reading of the whole lesson, interviews, and the specified sections for each code.

All decision rules are listed as footnotes within this document.

MC = Math Curriculum (*Investigations* in Greene or *Everyday Mathematics* in Region Z)

1. Grade Level:
2. Use of MC and materials outside the MC[1] lesson write-up sections for closer reading:
   - Identify all print resources used. (Important: Copies of print materials should be obtained by the observer and appended to this report.)
     - Other curricular materials:
     - Teacher-made materials:
     - Other (e.g., children's literature, article from newspaper, etc.)
   - pdfs for print materials used in the lesson

2a. Use of MC (circle one):

0 Teacher used the lesson plan provided by the MC for 0% of the lesson
1 Teacher used the lesson plan provided by the MC for 1–25% of the lesson
2 Teacher used the lesson plan provided by the MC for 26–75% of the lesson
3 Teacher used the lesson plan provided by the MC for 76–100% of the lesson

2b. Use of materials other than the MC (circle one)[2]:

0 Teacher used materials and/or ideas from sources other than the MC for 0% of the lesson
1 Teacher used materials and/or ideas from sources other than MC for 1–25% of the lesson
2 Teacher used materials and/or ideas from sources other than MC for 26–75% of the lesson

Appendix B **(continued)**

3 Teacher used materials and/or ideas from sources other than MC for 76–100% of the lesson

2c. If materials other than the MC used, indicate what those materials are (circle all that apply):

0 Not applicable (teacher did not use materials other than the MC)

1 District-sanctioned materials (Kathy Richardson materials in Greene and Math Steps in Region Z)

2 Materials congruent with curriculum[3]

3 Materials incongruent with the curriculum[4]

4 Can't tell[5]

3. If the teacher reviewed the MC prior to the lesson, what did the teacher look for in the curricular materials?[6] (circle all that apply)

1 Non-mathematical details (e.g., how to set up activities, how long students should work in groups)[7]

2 Materials needed for the lesson (e.g., manipulatives, graph paper, calculators)

3 Big mathematical ideas that the lesson is meant to get across[8]

4 Other (describe): ____

5 Not asked or not answered

6 Not applicable (teacher did not use the MC)

4. Teacher talk about the lesson[9]

4a. Did the teacher talk to anyone else about the lesson? (circle one)

1 Yes

2 No

3 Not asked or not answered

4b. If the teacher talked to someone else about the lesson, to whom did the teacher talk? (circle all that apply)

1 Grade-level colleague(s)

2 Coach

3 Other (describe): _______________________

4 Not asked or not answered

4c. If the teacher talked to someone else about the lesson, what did the teacher talk about with someone else? (circle all that apply)

1 Non-mathematical details (e.g., how to set up activities, how long students should work in groups)[10]

2 Materials needed for the lesson (e.g., manipulatives, graph paper, calculators)

3 Procedures that the teacher should follow in the lesson (e.g., mathematical representation, vocabulary, order of mathematical procedures)

4 Big mathematical ideas that the lesson is meant to get across

5 Other (describe)[11]: ____

6 Not asked or not answered

5. Cognitive Demand[12]

5a Primary instructional task in the lesson[13]: ______________________

Length: ____________

(continued)

Appendix B **(continued)**

5b. Cognitive demand of the task as it appeared in *EM* or *Investigations* (circle one)[14]:
- 1 No mathematical activity
- 2 Memorization
- 3 Use of Procedures Without Connections
- 4 Use of Procedures With Connections[15]
- 5 Doing Mathematics
- 6 Task did not appear in *EM* or *Investigations*

5c. Cognitive demand of the task as it was set up by the teacher (circle one)[16]:
- 1 No mathematical activity
- 2 Memorization
- 3 Use of Procedures Without Connections
- 4 Use of Procedures With Connections
- 5 Doing Mathematics
- 6 Task did not appear in *EM* or *Investigations*

5d. If the task changes from a 5 (Doing Mathematics) in the curriculum to a 4 (Use of Procedures With Connections) as set up by the teacher, explain how and—if possible—why the teacher made the alteration.

5e. Cognitive demand of the task as it was enacted by students and teacher (circle one)[17]:
- 1 No mathematical activity
- 2 Memorization
- 3 Use of Procedures Without Connections
- 4 Use of Procedures With Connections
- 5 Doing Mathematics
- 6 Unsystematic and/or non-productive exploration
- 7 Task did not appear in *EM* or *Investigations*

6. Teacher Work to Uncover Student Thinking[18] (circle one):
- 0 The teacher did no work to uncover student thinking; he or she did most of the talking in the lesson and/or asked questions with short or one-word answers.
- 1 The teacher did some work to uncover student thinking by asking some open-ended questions, by asking for some explanations, by arranging for public sharing of student responses, and/or by listening respectfully.[19]
- 2 In addition to #1 above, the teacher purposefully selected certain students to share their work during whole-class discussion because she wanted the whole class to hear about the mathematical approach the student took. However, the teacher *did not* sequence or connect students' responses in a mathematically meaningful way (i.e., to move the class toward the mathematical goal of the lesson).
- 3 In addition to #1 and #2 above, the teacher sequenced and/or connected students' responses in a mathematically meaningful way to make student thinking productive for the class as a whole (i.e., to move the class toward the mathematical goal of the lesson).

7. Intellectual Authority[20] (circle one):
- 0 The teacher fostered little or no student construction of mathematical ideas, thinking, and/or reasoning. Judgments about correctness were derived from the text or the teacher, with no appeal to mathematical reasoning.

Appendix B **(continued)**

1 The teacher fostered some student construction of mathematical ideas, thinking, and/or reasoning. However, judgments about correctness were mostly derived from the text or the teacher. Nevertheless, some appeals to mathematical reasoning were made.

2 The teacher fostered student construction of mathematical ideas, thinking, and/or reasoning. Additionally, judgments about correctness were primarily (most of the time) derived from mathematical reasoning and discussion during the class.

Notes to Appendix B

[1]Question 2 involves use of materials only. It does not address anything deeper, like alterations the teacher makes to the math lesson beyond the introduction of additional materials that are not part of the MC. To determine % of time, use the time markers in the lesson observation write-up.

[2]If you cannot tell from the lesson materials you received for this lesson whether additional materials used in the lesson are from the MC, assume that they are from outside of the MC.

[3]For Question 2c, "congruent materials" are at a 4 (Procedures With Connections) or 5 (Doing Mathematics) level of cognitive demand. If the materials include procedures that have the *potential* to connect with meaning, but those connections have not been featured in the task, the materials should be designated as "incongruent," particularly if the materials emphasize fluency and memorization.

[4]For Question 2c, "incongruent materials" are at a 1 (no mathematical activity), 2 (memorization), or 3 (Procedures Without Connections) level of cognitive demand and/or involve test prep.

[5]For Question 2c, select the "can't tell" option if there are no materials available for inspection.

[6]For Question 4, only take into account what the teacher looked for in the MC materials themselves, not what the teacher might have wanted to keep in mind generally prior to the lesson. Also only take into account teachers' review of curriculum materials in regard to the three lessons that will be observed after the pre-interview and before the post-interview. Do not take into account teachers' talk about their review of curriculum materials for other lessons they teach.

[7]For Question 4, "non-mathematical details" refers to the temporal and structural elements of the lesson.

[8]For Question 4, "big mathematical ideas" refers to teacher talk about what they reviewed that moves beyond the activity level and begins to articulate what concepts or ideas are at the heart of the lesson. Teachers' talk about attending to big mathematical ideas might be captured at various points in the pre-interview, including their answers to questions about the goals of the lesson. Keep in mind also that grade level might play a role in teachers' descriptions of big mathematical ideas in the lesson. Specifically, teachers at lower grade levels may not necessarily be articulating as sophisticated and/or complex big mathematical ideas as those at higher grade levels. Thus, the stakes are a little higher for what counts as a big mathematical idea for a teacher at a higher grade level compared to one at a lower grade level.

[9]For Question 5, only take into account teacher talk about the three lessons that will be observed after the pre-interview and before the post-interview. Do not take into account talk that does not refer to those lessons (e.g., who the teacher generally talks to about teaching math or using *Investigations/EM*). If the observer question or the teacher answer is vague or could be referencing general teacher talk rather than teacher talk about these specific lessons, you can choose the "not asked or not answered" option for 5a.

(continued)

Appendix B **(continued)**

[10]For Question 5c, "non-mathematical details" refers to the temporal and structural elements of the lesson.

[11]"Post-lesson talk" can be included as an "other" kind of talk here.

[12]Questions 6b, 6c, 6d, and 6e pertain only to the primary instructional task of the lesson as defined in Question 6a. If you are unsure about assigning a cognitive demand code, refer to the Task Analysis Guide, with its descriptions of the four levels of cognitive demand.

[13]For Question 6a, go with what the lesson observer has identified as the main mathematical task of the lesson.

[14]For Question 6b, code the text materials that formed the entire source of the task, not just a slice that the teacher may have selected to set up.

[15]If the primary instructional task includes procedures that have the potential to be connected with meaning, but those connections are not a featured part of the task, the task should be characterized as "Use of Procedures Without Connections," particularly if the task emphasizes fluency and memorization.

[16]For Questions 6d and 6e, consider whether the teacher did anything to take away from or add to the lesson plan in the curriculum materials. If the teacher did not alter the lesson plan in any significant way, the cognitive demand codes for setup and implementation should carry over from the cognitive demand materials code.

[17]For Question 6e, code should be based upon how the majority of the students responded during the majority of the time they were engaged with the task.

[18]For Question 7, the code should take into account the entire lesson and pertain to teacher instruction during the majority of the lesson for the majority of the students.

[19]The code for Question 7, Choice 1, indicates that the teacher is making at least some effort to engage the students, however successful that effort is.

[20]The code for Question 8 takes into account the entire lesson and pertains to teacher instruction during the majority of the lesson for the majority of the students. Keep in mind also that any code beyond 0 indicates that the teacher is making some effort to foster student construction of ideas, thinking, and/or reasoning. It is not enough that the students are doing the constructing if the teacher has not done anything to foster that student work.

Appendix C
Survey Items Used in Analysis

I. Teacher background

1. Including this school year, how many years have you been teaching?
- ☐ Less than 1 year
- ☐ 1 year
- ☐ 2 years
- ☐ 3 years
- ☐ 4 years
- ☐ 5 years
- ☐ 6-10 years
- ☐ 11 or more years

2. Do you have a bachelor's degree?
- ☐ Yes
- ☐ No

3. Do you have a degree or any advanced credits beyond your Bachelor's degree?
- ☐ Yes
- ☐ No

4. Do you have a master's degree?
- ☐ Yes
- ☐ No

5. Other than a doctorate, do you have any advanced credit beyond your master's degree?
- ☐ Yes
- ☐ No

6. Do you have a doctorate?
- ☐ Yes
- ☐ No

7. Do you have additional endorsements or certifications? Check all that apply.
- ☐ ELL/ESL
- ☐ Gifted
- ☐ Reading
- ☐ Mathematics
- ☐ Middle School
- ☐ Bilingual Education
- ☐ Special Education
- ☐ Principal/Supervisor/Superintendent
- ☐ Other (Please specify):

(continued)

Appendix C **(continued)**

II. Teacher perceptions of curriculum usefulness

8. To what extent do you agree or disagree with the following statements about [*Everyday Mathematics* or *Investigations* (*TERC*)]?

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
a. It contains useful information for me about underlying mathematical ideas	☐	☐	☐	☐	☐
b. It provides me with useful information about how to teach particular mathematical ideas and procedures	☐	☐	☐	☐	☐
c. It provides me with useful information about what students typically know, can do, or have difficulty with	☐	☐	☐	☐	☐
d. It does not emphasize the things that are important for students to learn in mathematics	☐	☐	☐	☐	☐
e. It is difficult to use	☐	☐	☐	☐	☐
f. It (the English version) is not effective with students who speak English as a second language	☐	☐	☐	☐	☐

III. Teacher Knowledge of Mathematics for Teaching

Note*: To measure Teacher Knowledge of Mathematics for Teaching (KMT), we used items developed by the Study of Instructional Improvement (Hill, Schilling, & Ball 2004). The items reprinted below are a **subset of the items that were used. For more information about using these items, go to the following Web site: http://sitemaker.umich.edu/lmt/files/LMT_sample_items.pdf

9. Mr. Garrett's students were working on strategies for finding the answers to multiplication problems. Which of the following strategies would you expect to see some elementary school students using to find the answer to 8 x 8? Mark ***yes***, ***no***, or ***I'm not sure*** for each strategy.

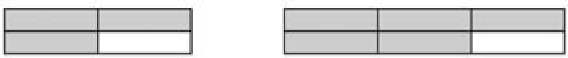
Appendix C **(continued)**

a. They might multiply 8 x 4 = 32 and then double that by doing 32 x 2 = 64.	☐ Yes ☐ No ☐ I'm not sure
b. They might multiply 10 x 10 = 100 and then subtract 36 to get 64	☐ Yes ☐ No ☐ I'm not sure
c. They might multiply 8 x 10 = 80 and then subtract 8 x 2 from 80: 80 – 16 = 64.	☐ Yes ☐ No ☐ I'm not sure
d. They might multiply 8 x 5 = 40 and then count up by 8s: 48, 56, 64.	☐ Yes ☐ No ☐ I'm not sure

10. You are working individually with Bonny, and you ask her to count out 23 checkers, which she does successfully. You then ask her to show you how many checkers are represented by the 3 in 23, and she counts out 3 checkers. Then you ask her to show you how many checkers are represented by the 2 in 23, and she counts out 2 checkers. What problem is Bonny having here?

☐ Bonny doesn't know how large 23 is.
☐ Bonny thinks that 2 and 20 are the same.
☐ Bonny doesn't understand the meaning of the places in the numeral 23.
☐ All of the above.

11. Takeem's teacher asks him to make a drawing to compare 3/4 and 5/6. He draws the following:



and claims that 3/4 and 5/6 are the same amount. What is the <u>most likely</u> explanation for Takeem's answer?

☐ Takeem is noticing that each figure leaves one square unshaded.
☐ Takeem has not yet learned the procedure for finding common denominators
☐ Takeem is adding 2 to both the numerator and denominator of 3/4, and he sees that that equals 5/6.
☐ All of the above are equally likely.
☐ I'm not sure.

(continued)

Appendix C **(continued)**

12. A number is called "abundant" if the sum of its proper factors exceeds the number. For example, 12 is abundant because 1 + 2 + 3 + 4 + 6 > 12. On a homework assignment, a student incorrectly recorded that the numbers 9 and 25 were abundant. What are the most likely reason(s) for this student's confusion.

a. The student may be adding incorrectly.	☐ Yes ☐ No ☐ I'm not sure
b. The student may be reversing the definition, thinking that a number is "abundant" if the number exceeds the sum of its proper factors.	☐ Yes ☐ No ☐ I'm not sure
c. The student may be including the number itself in the list of factors, confusing proper factors with factors.	☐ Yes ☐ No ☐ I'm not sure
d. The student may think that "abundant" is another name for square numbers.	☐ Yes ☐ No ☐ I'm not sure

IV. Teacher Professional Development

13. Not including the in-service days just before the start of classes, did you receive any math-related Professional Development over this past summer?

☐ Yes
☐ No → If no, skip #14

14. About how many hours of math-related Professional Development did you receive over this past summer?

_____ Write number of PD hours

15. During this current school year, did you receive any math-related Professional Development? Please include PD during in-service days.

☐ Yes
☐ No → If no, skip #16

16. About how many hours of math-related Professional Development have you received during this current school year? Please include PD during in-service days.

_____ Write number of PD hours

Notes

Work on this article was supported by a grant from the Interagency Educational Research Initiative (award #0228343). All opinions and conclusions in this article are those of the authors and do not necessarily reflect the views of the funding agency.

[1]Based on a survey of all elementary teachers in our study who used *EM*, almost 30% (n = 82) said that they "never" or "rarely" use the *Teacher's Reference Manual* as they planned lessons. Furthermore, in our interviews with a smaller number of teachers (n = 27), no teachers mentioned the *Teacher's Reference Manual* in response to a question about what they review in the curriculum prior to teaching lessons.

[2]The reader is reminded that these levels of demand and support are stated with respect to teachers and their learning (not students).

[3]This line of inquiry is akin to Kazemi and Hubbard's (2008) recent investigation of the interplay between professional development and instruction. In their work, they emphasize that the typical one-way analysis of how professional development impacts instruction does not unpack the complex relationship between teacher learning and their classroom practice.

[4]Additionally, another stream of our current research (Kaufman & Stein, 2009) indicates that the nature of the high-level task in materials should not make a difference for teachers' work to uncover student thinking and locate intellectual authority in mathematical reasoning. That is, both Doing Mathematics and Procedures With Connections tasks that are maintained from materials to enactment across Region Z and Greene are highly correlated ($p < .01$) with our student thinking and intellectual authority measures. Thus, that the *Everyday Mathematics* curriculum in Region Z contains more Procedures With Connections tasks and that *Investigations* contains more Doing Mathematics tasks should not matter for which district's teachers have higher quality implementation.

[5]Viewing the curriculum as a resource for teachers as well as for students builds on the work of researchers who have investigated the role of curriculum in teacher learning and instructional reform (Davis & Krajcik, 2005). Using the term *educative* to refer to K–12 curriculum materials that are intended to promote teacher learning, Davis and Krajcik (2005) have elaborated on ways in which curriculum materials can be designed to be educative for teachers.

[6]Greene is a pseudonym.

[7]Region Z (a pseudonym) is one of 10 subdistricts into which the New York City schools were subdivided when taken over by Joel Klein.

[8]Because one pre- and one post-interview were conducted per set of three contiguous lessons, the coded data based on those interviews are the same across all lessons in one set.

[9]Because variables measuring teachers' review of curricula and talk with others are binary variables, they are not included in this correlation matrix. We will discuss the relationship between those binary variables and implementation quality variables following the presentation of this correlation analysis.

[10]For this analysis, we focus only on those teachers from each district who reported reviewing or talking about the big mathematical ideas. Thus, our analysis takes into account only 14 pre-interviews with Region Z teachers (11 teachers total) and 56 pre-interviews with Greene teachers (24 teachers total).

[11]To do so, a teacher would have to step back and put into view the entire strand, which is possible because the materials chart the strands and the lessons contributing to them. However, other than in one professional development, we never witnessed teachers doing the work of looking across a strand.

References

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (2009). *Achievement effects of four early elementary math curricula: Findings from first graders in 39 schools* (SCEE 2009-4052). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education.

Ball, D. L. (2001). Teaching with respect to mathematics and students. In T. Wood, B. S. Nelson, & J. Warfield (Eds.), *Beyond classical pedagogy: Teaching elementary school mathematics* (pp. 11–22). Mahwah, NJ: Lawrence Erlbaum.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6–8.

Banilower, E. R., Smith, P. S., & Weiss, I. R. (2002). *Examining the influence of national standards: Data from the 2000 National Survey of Science and Mathematics Education*. Chapel Hill, NC: Horizon Research.

Baxter, J. A., & Williams, S. (in press). Social and analytic scaffolding in middle school mathematics: Managing the dilemma of telling. *Journal of Mathematics Teacher Education*.

Boerst, T., & Sleep, L. (2007). Learning to do the work of mathematics teaching. In M. Lampert (Chair), *Conceptualizing and using routines of practice in mathematics teaching to advance professional education*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Brendefur, J., & Frykholm, J. (2000). Prompting mathematical communication in the classroom: Two preservice teachers' conceptions and practices. *Journal of Mathematics Teacher Education*, *3*(2), 125–153.

Brown, M. W. (2009). The teacher-tool relationship: Theorizing the design and use of curriculum materials. In J. Remillard, G. Lloyd, & B. Herbel-Eisenmann (Eds.), *Teachers' use of mathematics curriculum materials: Research perspectives on relationships between teachers and curriculum* (pp. 17–36). New York: Routledge.

Chazen, D., & Ball, D. L. (2001). Beyond being told not to tell. *For the Learning of Mathematics*, *19*(2), 2–10.

Cobb, P., Gravemeijer, K., Yackel, E., McClain, K., & Whitenack, J. (1997). Mathematizing and symbolizing: The emergence of chains of signification in one first-grade classroom. In D. Kirschner & J. A. Whitson (Eds.), *Situated cognition: Social, semiotic and psychological perspectives* (pp. 151–233). Mahwah, NJ: Lawrence Erlbaum.

Coburn, C. E. (2001). *Making sense of reading: Logics of reading in the institutional environment and the classroom*. Ann Arbor, MI: University Microfilms.

Coburn, C. E., & Russell, J. L. (2008). District policy and teacher social networks. *Educational Evaluation and Policy Analysis, 30*(3), 203–235.

Cohen, D., & Hill, H. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, *34*(3), 3–14.

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

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

Elmore, R., & Burney, D. (1999). Investing in teacher learning: Staff development and instructional improvement. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 263–291). San Francisco: Jossey-Bass.

Engle, R., & Conant, F. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, *20*(4), 399–483.

Fullan, M. (1991). *The new meaning of educational change* (2nd ed.). New York: Teachers College Press.

Fullan, M., & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research*, *47*(2), 335–397.

Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, *79*(4), 79–107.

Hamm, J., & Perry, M. (2002). Learning mathematics in first-grade classrooms: On whose authority? *Journal of Educational Psychology, 94*(1), 126–137.

Heneman, H. G., Milanowski, A., Kimball, S., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay* (CPRE Policy Brief No. RB-45). Philadelphia: Consortium for Policy Research in Education.

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*(5), 524–549.

Hightower, A., Knapp, M., Marsh, J., & McLaughlin, M. (Eds.). (2002). *School Districts and Instructional Renewal*. New York: Teachers College Press.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, *105*(1), 11–30.

Hodge, L. L., & Cobb, P. (2003, April). *Classrooms as design spaces for supporting students' mathematical learning and engagement*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Hubbard, L., Mehan, H., & Stein, M. K. (2006). *Reform as learning: School reform, organizational culture, and community politics in San Diego*. New York: Routledge.

Kaufman, J., & Stein, M. K. (2009). *Quality implementation of standards-based mathematics curricula: A longitudinal comparison of teacher instruction in two urban school districts*. Paper accepted to the annual meeting of the American Educational Research Association, Denver, CO.

Kazemi, E., & Hubbard, A. (2008). New directions for the design and study of professional development: Attending to the coevolution of teachers' participation across contexts. *Journal of Teacher Education*, *59*(5), 428–441.

Lampert, M. (1990a). Connecting inventions with conventions. In L. P. Steffe & T. Wood (Eds.), *Transforming children's mathematics education* (pp. 253–265). Hillsdale, NJ: Lawrence Erlbaum.

Lampert, M. (1990b). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, *27*(1), 29–63.

Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.

Leinhardt, G., & Steele, M. D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction*, *23*(1), 87–163.

Manouchehri, A., & Goodman, T. (1998). Mathematics curriculum reform and teachers: Understanding the connections. *Journal of Educational Research*, *92*(1), 27–41.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. B., & Fellows of the Institute for Learning. (2002). *Accountable Talk$_{SM}$: Classroom conversation that works* [CD-ROM]. Pittsburgh, PA: Institute for Learning, Learning Research and Development Center, University of Pittsburgh.

Nelson, B. S. (2001). Constructing facilitative teaching. In T. Wood, B. S. Nelson, & J. Warfield (Eds.), *Beyond classical pedagogy: Teaching elementary school mathematics* (pp. 251–273). Mahwah, NJ: Lawrence Erlbaum.

Remillard, J., Lloyd, G., & Herbel-Eisenmann, B. (Eds.). (2009). *Teachers' use of mathematics curriculum materials: Research perspectives on relationships between teachers and curriculum*. New York: Routledge.

Riordan, J., & Noyce, P. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education*, *32*(4), 368–398.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of elementary schools. *Teachers College Record*, *104*(8), 1525–1567.

Scardamalia, M., Bereiter, C., & Lamon, M. (1994). The CSILE project: Trying to bring the classroom into World 3. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and educational practice* (pp. 201–228). Cambridge, MA: MIT Press.

Schoenfeld, A. S. (1998). Toward a theory of teaching-in-context. *Issues in Education*, *4*(1), 1–95.

Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction*, *20*(2), 119–150.

Shifter, D. (2001). Learning to see the invisible: What skills and knowledge are needed to engage with students' mathematical ideas? In T. Wood, B. S. Nelson, & J. Warfield (Eds.), *Beyond classical pedagogy: Teaching elementary school mathematics* (pp. 109–134). Mahwah, NJ: Lawrence Erlbaum.

Stein, M. K., & Coburn, C. E. (2008). Architectures for learning: A comparative analysis of two urban districts. *The American Journal of Education*, *114*(4), 583–626.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, *10*(4), 313–340.

Stein, M. K., Grover, B. W., & 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*(2), 455–488.

Stein, M. K., & Kim, G. (2009). The role of mathematics curriculum materials in large-scale urban reform: An analysis of demands and opportunities for teacher learning. In J. Remillard, G. Lloyd, & B. Herbel-Eisenmann (Eds.), *Teachers' use of mathematics curriculum materials: Research perspectives on relationships between teachers and curriculum*, (pp. 37–55). New York: Routledge.

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*(1), 50–80.

Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319–370). Greenwich, CT: Information Age.

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

Supovitz, J. A. (2006). *The case for district-based reform: Leading, building and sustaining school improvement*. Cambridge, MA: Harvard Education Press.

Tarr, J. E., Chavez, O., Reys, R. E., & Reys, B. J. (2006). From the written to enacted curricula: The intermediary role of middle school mathematics teachers in shaping students' opportunities to learn. *School Science and Mathematics*, *106*(4), 191–201.

Wertsch, J. (1998). *Mind as action*. New York: Oxford University Press.

Wertsch, J. V., & Toma, C. (1995). Discourse and learning in the classroom: A sociocultural approach. In L. P. Steffe & J. Gale (Eds.), *Constructivism in education* (pp. 159–174). Hillsdale, NJ: Lawrence Erlbaum.

What Works Clearinghouse. (2007). *WWC topic report: Elementary school math*. Washington, DC: Institute for Education Sciences, U.S. Department of Education. Retrieved October 16, 2007, from http://ies.ed.gov/ncee/wwc/pdf/ESM_TR_07_16_07.pdf

Wilson, S. (1990). A conflict of interests: The case of Mark Black. *Educational Evaluation and Policy Analysis*, *12*(3), 309–326.

Yuan, K., Lockwood, J. R., Hamilton, L., Gill, B., & Stein, M. K. (2008). *Student achievement in districts scaling up mathematics: As related to district policy, human and social capital, curriculum implementation and instructional style*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Manuscript received April 9, 2009
Final revision received November 23, 2009
Accepted December 3, 2009