

Mathematics Teacher Development in the Context of District Managed Curriculum

Mary Kay Stein, Julia Kaufman, and Miray Tekkumru Kisa

Abstract The purpose of this study was to develop and test the viability of a conceptual framework for analyzing mathematics instruction and mathematics teacher development within the context of policies regarding district-wide adoption of curriculum. The framework takes three dimensions of curriculum-based instruction into account independently: use, congruence (the extent to which instruction aligns with district and curricular guidelines), and quality (the extent to which instruction maintains the cognitive demand of appropriately challenging tasks, takes account of and builds on student thinking, and situates intellectual authority in mathematical reasoning). Based on analyses of multiple observations of 36 teachers across two districts, teachers were classified into one of four implementation profiles (flounderer, mechanical, canonical, maverick) that were created by crossing the three dimensions; in addition, their trajectory through those profiles was traced over a two-year period. Results suggest teachers were more likely to use the district-adopted curricula as the source of their lessons than to align their practice with curricular and district guidelines. Teachers' demonstration of high-quality lessons was less frequent. Differences across the two districts in the percentages of teachers falling into each of the implementation profiles suggests that district actions may have shaped teachers' uptake of the curriculum. Finally, results suggest a more uneven pathway toward high-quality instruction than had been initially conjectured.

Keywords Mathematics teacher learning · Mathematics teacher development · Curriculum · Instructional practice · Mathematics teaching · District policy

An earlier version of this paper was presented at the 2011 annual meeting of the American Educational Research Association, New Orleans. This work was supported by a grant from the National Science Foundation (IERI Grant REC-0228343), as well as support by the Institution of Education Sciences and the U.S. Department of Education (Award R305B1000012). The content or opinions expressed herein do not necessarily reflect the view of the National Science Foundation or any other agency of the U.S. Government.

M.K. Stein (✉) · M.T. Kisa
LRDC, University of Pittsburgh, Pittsburgh, USA
e-mail: mkstein@pitt.edu

J. Kaufman
University of Pittsburgh, Pittsburgh, USA

Over the past decade, district policies in the United States have become increasingly focused on the improvement of instruction, especially in subjects that are regularly tested under NCLB (Elmore and Burney 1999; Hightower et al. 2002; Hubbard et al. 2006; Supovitz 2006). In mathematics, curriculum¹ 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 and Cohen 1996; Coburn 2001; Fullan 1991; Fullan and Pomfret 1977; Wilson 1990). While it may be relatively easy to get curriculum materials into the hands of large numbers of teachers, it can be difficult for district leaders to ensure that teachers actually *use* the new materials and more difficult yet to ensure that they use them in a manner that is *congruent* with the pedagogical² features of the curriculum (e.g., group work, manipulative use) and with district guidelines for the sequencing and pacing of lessons/units.

To complicate matters further, even the use of curricula in a congruent manner (as described above) still does not guarantee high-quality instruction, especially for standards-based mathematics curricula that are comprised of cognitively challenging instructional tasks.³ Teachers can set up an instructional task exactly as specified in the curricular materials, yet fail to support students' high-level thinking and reasoning as they actually work on the task (Stein et al. 1996). This is significant because it is not whether students are sitting in groups or using manipulatives or on the right lesson on the right day that matters, rather it is what students are actually thinking about that determines their opportunities to learn.

The purpose of this study was to develop and test the viability of a conceptual framework for analyzing mathematics instruction and mathematics teacher development *within the context of local policies regarding district-wide curriculum adoption and implementation*. Our framework will take *use*, *congruence* and *quality* into account independently as we develop teacher implementation profiles and conjecture pathways of teacher development.

We view the study's contribution as two-fold. First, we believe that our provision of a new framework that takes use, congruence, and quality into account separately represents an advance for the field of research on curriculum implementation and that it can serve as a unifying framework for future studies of large-scale teacher improvement within the context of district managed curricula. The study results suggest that our framework is "up to the task" in that it was able to detect meaningful variation among teachers—variation that appears to be related to the context of the school or district in which they worked. Second, situating the study of teacher development within district reform efforts provides an illustration of how combined

¹In this manuscript, we use the term, "curriculum" to mean a textbook series.

²Non-US readers may prefer the term "didactical." The features to which we refer are those that relate to how to teach the mathematics content.

³Because the two curricula used in this study were standards-based and at least partially funded by the National Science Foundation, the presumption (supported by some prior analyses [see Stein and Kim 2009]) is that the tasks—as they appeared in the curriculum—were high-quality.

attention to district policies and implementation can make progress on understanding and supporting the improvement of teaching on a large scale.

Theoretical Framework

Most models of mathematics teacher development describe teacher learning without reference to the materials with which they interact on a daily basis or the work environment in which their learning occurs (Fennema and Nelson 1997). The contribution of our framework is that it examines teacher learning in a specific, well-defined context: large-scale, district-mandated improvement efforts that rely heavily on the adoption and implementation of standards-based curricula. These kinds of district-wide improvement efforts have become increasingly prevalent over the past decade in the United States with many large urban districts adopting and supporting one carefully selected curriculum (Hightower et al. 2002; Supovitz 2006).

We propose that teacher learning occurs along one or more pathways or trajectories that can be specified. Similar to current efforts to identify *student* learning trajectories that one would expect to emerge within the context of well-conducted programs of instruction (Clements and Sarama 2004), our long-term goal is to identify teacher learning trajectories that could be expected to emerge within the context of well-conducted district improvement efforts.

District Improvement Initiatives as Context for Teacher Learning

Two key features of district improvement efforts that can impact how teacher learning unfolds are (a) the selected *curriculum*; and (b) the *professional support* provided to teachers and other professionals as they are learning to implement the new curriculum.

Selected Curriculum

Past research suggests that standards-based mathematics curricula can offer both challenges and supports for teacher learning (Davis and Krajcik 2005). They offer *challenges* to teacher learning because they aim for more ambitious, cognitively complex forms of student learning (i.e., conceptual understanding; the capacity to think, reason, and problem solve) than teachers have traditionally been accustomed to. Not only did teachers themselves likely not learn mathematics in this less traditional way, but many have also not learned to teach mathematics in ways that foster students' capacity to think, reason, and problem solve (Borko and Putnam 1995).

When designed well, standards-based curricula can offer *support* for teacher learning (Davis and Krajcik 2005; Stein and Kim 2009). Instead of treating the

teacher as an instrument for delivering the curriculum to students, some standards-based curricula invest in the education of the teacher as a critical contributor to the teaching and learning environment. Designers of these so-called *educative* curricula believe that student learning cannot be entirely scripted in advance, but rather unfolds in moment-to-moment, contingent interactions between teachers and students during a lesson; interactions in which materials are a resource for, not the determinant of, learning. In this view of teaching and learning, the teacher must have sufficient knowledge of the mathematical purpose and learning goals of the instructional tasks in the curriculum and insight into how students might respond to those tasks. This kind of information thus becomes integrated into the curricular materials. Despite the increasing popularity of the idea of educative curricula, recent research suggests that standards-based curricula differ widely in the extent to which they are educative for teachers (Stein and Kim 2009).

Professional Support

In addition to the curriculum materials that they select, districts also vary in the nature and extent of *professional support* offered to teachers in the context of district-wide curricular reform initiatives. Most districts now recognize that teachers need more support than that offered by the typical publisher-provided one-day training session. Common support structures include the provision of coaches (Duessen et al. 2007) and common planning periods for teachers on the same grade level. Because of the system-wide nature of these initiatives, professional support is often arranged at contiguous levels of the system. For example, the district mathematics leadership team might provide some kind of ongoing support for principals, as well as hold monthly meetings with coaches; the coaches, in turn, might meet weekly with their building leadership team as well as hold weekly meetings with teachers. Sometimes district math leaders deliver professional development directly to teachers.⁴ It should be noted that, although the above kinds of support structures can be found across many districts, past research suggests that districts vary with respect to *how* these support systems are carried out, with some focusing more on operational features such as how to use materials and pacing guidelines and others focusing more on the big mathematical ideas and the underlying intent of the lessons (Stein and Coburn 2008).

Framework for Analyzing Instruction and Teacher Development

Our conceptual framework for analyzing teaching and teacher learning within the context of district-based improvement efforts is based on (a) the extent to which

⁴This is possible because, in the context of district-paced implementation, teachers on the same grade level should be implementing the same lessons at roughly the same time, thereby allowing the district to “preview” an upcoming unit to teachers across the district on the same date.

teachers actually *use* the selected curriculum as the source of their lessons; (b) the degree of *congruence* of teachers' instruction to curricular and district guidelines; and (c) the *quality* of teachers' instruction (the extent to which it maintains the cognitive demand of appropriately challenging tasks, takes account of and builds on student thinking, and situates intellectual authority in mathematical reasoning). Each of these is described below.

Use of Curriculum We conceptualize curriculum use as the extent to which the teacher draws on the selected curriculum as the source of activities in her lessons. It is important to note that this measure says nothing about *how well* the teachers use the curriculum or even the extent to which they follow the curriculum's and district's guidelines for how to run the lesson. Nevertheless, assessments of use are important because curriculum use constitutes a necessary foundation for large-scale teacher learning within a district-led improvement effort. If the curriculum materials remain swathed in shrink wrap in the closet, teachers and students will not be able to avail themselves of the activities and opportunities for learning contained in them. This aspect of curriculum based reform is often assumed in studies of teacher change, but experience suggests that it should not be taken for granted.

Congruence with Curricular and District Guidelines We conceptualize congruence as the extent to which teachers' instruction aligns with the pedagogical features of the curriculum (e.g., group work, manipulative use) and with district guidelines for the proper sequencing and pacing of lessons/units.⁵ Determining congruence can be accomplished with reference to relatively superficial aspects of instruction, for example, items that might appear on a checklist that a principal uses to evaluate teacher adherence to district mandates. Items that would be relevant for determining congruence include features such as directions for how to set up a lesson (including the manipulatives that will be needed), how to group students for various parts of the lesson, and guidelines for pacing. Items not relevant for determining congruence include an examination of the mathematical ideas at play in the lesson or the extent to which students have the opportunity to learn those ideas.

Assessments of congruence are important because they signal a level of teacher effort that goes beyond using the curriculum materials as a source of activities. Congruent use implies that teachers are actually trying to follow the curriculum in a manner that is aligned with the curriculum developers' and the district's expectations.

Instructional Quality Instructional quality is conceptualized in terms of the affordances for student learning of important mathematical ideas that the instruction provides. Although our criteria for instructional quality adhere to a particular approach to teaching and learning (variously referred to as standards-based, student-centered, or inquiry based), they have not been designed to align specifically with

⁵Judgments about congruence are necessarily district and curriculum specific.

any one particular curriculum. However, standards-based curricula, in general (including the two curricula studied herein) are philosophically compatible with this view of teaching and learning.

We've defined instructional quality in terms of three constructs: the maintenance of high levels of cognitive demand, the level and kind of attention that the teacher pays to student thinking, and the extent to which the intellectual authority in the classroom is vested in mathematical reasoning (vs. the text or the teacher). Each of these is described in more detail below.

1. *Maintenance of high-level cognitive demand.* Cognitive demand refers to the level of thinking and reasoning that is required in order to successfully complete a mathematical instructional task (Doyle 1983; Stein et al. 1996).⁶ High-level tasks often consist of open-ended problems with limited guidance regarding how to solve them, thus requiring students to engage in complex, non-routine thinking and reasoning such as making and testing conjectures, framing problems, representing relationships and looking for patterns. High-level tasks can also be more constrained by orienting students toward the use of general procedures or multiple representations to solve complex problems, but doing so in such a way that concepts, meaning or understanding are illuminated. Low-level tasks focus students' attention on algorithms and routine procedures without attempts to foster conceptual understanding or on memorizing basic facts or definitions.

The cognitive demands of tasks often change as they pass through different phases (Stein et al. 1996). First, tasks exist in print on the pages of curricular **materials**. Next, as the teacher **sets up** the task in the classroom, she may (knowingly or unwittingly) change the cognitive demand of the task (e.g., by inserting easier numbers into the problem; by providing "hints" regarding what to look for). Finally, the students (sometimes with the teacher's help) go about actually working on or **enacting** the task. It is not unusual for the cognitive demand of the tasks to change at this final phase as well, usually as a result of the teacher "taking over" and doing the thinking for the students instead of allowing them to struggle. Past research has shown that students in classrooms in which teachers are able to maintain the high level of cognitive demand of tasks that appear in standards-based materials perform better on tests of higher level thinking and reasoning (Stein and Lane 1996). Thus, we consider one hallmark of a high-quality lesson to be the teacher's ability to maintain the high cognitive demand of instructional tasks.

2. *The level and kind of attention that teachers paid to student thinking.*⁷ Proponents of standards-based instruction 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., Brendehur and Frykholm

⁶A mathematical task is defined as a classroom activity, the purpose of which is to focus students' attention on a particular mathematical idea. An activity is not classified as a new or different task unless the underlying mathematical activity toward which the activity is oriented changes. Standards-based lessons typically consist of one or two tasks.

⁷Judgments about teachers attending to student thinking and about intellectual authority were made based on the entire lesson.

2000; Hodge and Cobb 2003; Lampert 2001; Nelson 2001; Schoenfeld 1998; Shifter 2001). This is commonly done by circulating around the classroom while students work (e.g., Boerst and Sleep 2007; Hodge and 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 (Brendehur and Frykholm 2000; Lampert 2001; Stein et al. 2008). Thus, we consider another feature of a high-quality lesson to be the extent to which the teacher attends to and builds on student thinking.

3. *Intellectual Authority*. Proponents of standards-based instruction 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 and Conant 2002; Hamm and Perry 2002; Lampert 1990; Scardamalia et al. 1994; Wertsch and 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 et al. 1997). Thus, our final feature of high-quality instruction is the extent to which the teacher fosters students’ intellectual authority.

Combining Above Features The unique feature of our framework is that it combines judgments about use, congruence and quality to arrive at a set of instructional profiles. By crossing *use* and *congruence* with *quality*, we have identified the following “implementation profiles”:⁸

- *Canonical Implementer*: High quality, with high use and high congruence. This teacher not only *uses* the district’s selected curriculum and aligns her instruction to be *congruent* with curricular and district guidelines, but she also provides students with *high-quality* opportunities to think, reason and problem solve.

- *Maverick*: High quality, with low use or low congruence. This teacher also provides her students with *high-quality* opportunities to learn to think, reason and problem solve; however, she does so without the curriculum. Either she does *not use* the curriculum at all; or she uses it in a manner that is *incongruent* with curricular and district guidelines.

- *Mechanical Implementer*: Low quality, with high use and high congruence. This teacher *does not provide high-quality* opportunities for student learning but she *uses* the curriculum in a manner that is *congruent* with curricular and district guidelines.

⁸We have identified 4 profiles instead of 8 possible profiles because only 4 profiles were conceptually meaningful. Two profiles (either high or low quality crossed with low use and high congruence) were unlikely because it is difficult to imagine a teacher implementing materials with pedagogical fidelity without actually using the materials. The other two profiles are actually represented in the flounderer and maverick categories which stipulate that (along with either low [flounderer] or high [maverick] quality) the teacher implements with low congruence and either high or low use. Conceptually the dimension that carries the weight of both the flounderer and maverick categories is having low congruence with the pedagogical guidance of the curriculum and district.

- *Flounderer*: Low quality, with low use or low congruence. This teacher is *not providing high-quality* opportunities for student learning and is disregarding the curriculum. Either she does *not use* the curriculum at all; or she uses it in a manner that is *incongruent* with curricular and district guidelines.

As shown by the profiles, this framework separates instructional quality judgments from “following the curriculum” judgments. As such, we are able to differentiate teachers who follow the curriculum in a superficial manner (mechanical implementers) from teachers who follow the curriculum with fidelity to the underlying intent of the curriculum (canonical implementers). In addition, we recognize two different ways of exhibiting high-quality instruction: the canonical implementer and a teacher who sets up and maintains the cognitive demand of appropriately challenging tasks, listens to and challenges student thinking, and encourages students to take mathematical authority, but who does not follow (and may not use) the district-supported curriculum (maverick). In this way, we allow for innovative, high-quality teaching that is not bound to a particular curriculum. Finally, there are also different ways of exhibiting poor-quality teaching: the mechanical implementer who is trying to follow the curriculum, albeit in a superficial manner and the flounderer who is not following (and perhaps not using) the district-supported curriculum but is also not exhibiting high-quality instruction.

Within a well-conducted district-supported implementation, we would conjecture the following pathway for teacher development. The teacher begins by using curriculum materials in *mechanical* ways. That is, she diligently bases her lessons on a set of well-designed curriculum materials and makes a good faith effort to follow the curricular guidelines set forth by her district including what lessons to teach, how quickly to go, what grouping formats to use and so forth. However, the teacher has difficulty delivering on the deeper structure of the curriculum. Over time—if she is well supported by educative materials and by her district—she begins to implement the curriculum in ways that conform to not just the surface features but also the deeper cognitive features that influence how students think and reason (becoming a canonical implementer). Finally, having “learned” a more cognitively challenging, student-centered manner of teaching, she may depart from the standard curriculum and become a maverick, meaning that her teaching is still high quality, but she no longer uses the district-mandated curriculum or she stops adhering closely to the operational guidelines of the curriculum and/or the district.

The purpose of this study was to develop and test the viability of this framework for analyzing mathematics instruction and mathematics teacher development within the context of local policies regarding district-wide curriculum adoption and implementation. The following questions guided this study:

1. How do teachers participating in district-wide curricular-based initiatives vary with respect to use of the mandated curriculum, congruence with curricular and district guidelines regarding how to use the curriculum, and quality of instruction?
2. How do teachers participating in district-wide curricular-based initiatives vary with respect to the framework’s four profiles?

3. What within-teacher patterns, if any, emerge with respect to the four profiles as teachers participate in the district-wide initiative over multiple years?
4. In what ways, if any, might the above identified patterns be related to the nature of the curricular materials and/or the nature of the professional support provided by the district?

Methods

Data Sources

Data for the present study come from a large NSF-supported multi-year study of the initial years of district-wide implementation of *Investigations* and *Everyday Mathematics* in two urban districts. In Fall 2003, Greene School District⁹ mandated implementation of *Investigations*, whereas Region-Z mandated implementation of *Everyday Mathematics*; both are standards-based elementary (grades K-5) curricula. Six focal teachers in each of 4 case-study schools in each district were selected for observation. Schools were selected to represent the range of schools in each district with respect to teacher capacity and extent of teacher professional communities; teachers were selected to represent the range of talent and grade levels in the building. For this study we used all the teachers for whom we had data for the 2004-05 and 2005-06 school years, which includes 19 Greene teachers and 17 Region-Z teachers.

Most teachers were observed six times per year (for 3 consecutive lessons in the fall and 3 consecutive lessons in the spring). All classroom observations were conducted by trained observers who took detailed field-notes and then completed pre-specified, qualitative write-ups upon leaving the classroom.¹⁰ 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.

Each lesson was coded by one of a group of four trained Masters- or PhD-level mathematics educators, all of whom were familiar with the first author's prior 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.¹¹ In order to prevent coding "drift," the coders

⁹Pseudonyms.

¹⁰The individuals who were selected to conduct the observations and create the write ups had expertise in either mathematics education or a social science field that relied heavily on observation (e.g., anthropology). They participated in a 2-day, in-person, group training at the start of the project. This training involved watching videos of mathematics lessons and creating write ups that were critiqued by project leaders and their peers. During the course of the project, the observers were provided feedback on their write-ups and participated in at least one follow up group session.

¹¹Because one pre- and one post-interview were conducted per set of 3 contiguous lessons, the coded data based on those interviews is the same across all lessons in one set.

met with the authors on a monthly basis to share codes for a randomly selected lesson. These 1–2 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 inter-rater reliability of 81 %, 67 %, and 75 % for use, congruence, and quality respectively.¹² For each double-coded lesson, differences were resolved and a consensus code was entered.

In addition to teacher observations, we have copies of all the curricular materials adopted by the two districts, transcripts of teacher pre- and post-lesson set interviews, observations of professional development at different levels of the system, and transcripts of interviews with principals, mathematics coaches and district leaders. We did not analyze these data sources firsthand, but instead drew on previous project analyses that examined (a) the nature of demand and support in the curriculum materials (Stein and Kim 2009; Stein and Kaufman 2010); the nature of district-wide support (Stein and Coburn 2008; Coburn and Russell 2008); differences across schools (Sutherland et al. 2007); and the evolution of reform mandates and supports over time (Kaufman and Stein 2010).

Procedures of Analysis

Our initial analysis focused on characterizing each of the 36 teachers according to use, congruence, and quality across the three observed lessons that they delivered in each of four semesters over the course of two years: Fall 2004, Spring 2005, Fall 2005, and Spring 2006. Each lesson write-up was coded by a mathematics educator according to *use* (on a scale of 0–4 according to the portion of the lesson that used the curriculum as the source of activities in the lesson); *congruence* (an aligned/non-aligned judgment based on the math educators’ assessment of the lesson’s congruence with the curriculum’s and district’s guidelines [specifically constructed for each curriculum]); and *quality* (a score of 1 to 8 based on judgments of the levels of cognitive demand at the set-up and enactment phases of the lesson coupled with mathematics educators’ judgment of where intellectual authority resided and the extent to which the lesson built on student thinking; this coding system builds on earlier work and is explained in Stein and Kaufman 2010).

Next, the scores for each of the teacher’s three lessons were averaged across the three observations to represent a season-year score on each dimension. Finally, teachers’ practice was identified as high or low use, congruent or non-congruent implementation, and high- or low-quality based on cut scores that were conceptually determined. Each of these analytic phases for use, congruence and quality is described below.

¹²Inter-rater reliability was computed as the number of agreements divided by the total number of possible agreements/disagreements.

Curriculum Use

Curriculum use was measured on the following scale:

- 0 = 0 % of the lesson drew on *Investigations or Everyday Mathematics*
- 1 = 1–25 % of the lesson drew on *Investigation or Everyday Mathematics*
- 2 = 26–75 % of the lesson drew on *Investigations or Everyday Mathematics*
- 3 = 76–99 % of the lesson drew on *Investigations or Everyday Mathematics*
- 4 = 100 % of the lesson drew on *Investigations or Everyday Mathematics*

High use was defined as a teacher with an average score of 3.0 or higher across the three lessons she taught during each semester, meaning that over 75 % of the time—on average—the teacher would have drawn on curricular materials for the three lessons she taught. Thus, in addition to those teachers who used the selected curriculum the entire time of all of their observed lessons, we also included teachers who used *Investigations or Everyday Mathematics* as the source of their classroom activities between 76 and 99 % of the time.¹³ Anyone who used *Investigations or Everyday Mathematics* 75 % of the time or less, on average, was characterized as a “low” user.

Congruence

We developed two separate checklists—one with indicators of congruent instruction and one with indicators of incongruent instruction—for each curriculum based on an in-depth analysis of the curriculum and the district’s expectation of how that curriculum should be implemented. For example, *Everyday Mathematics* relies on a spiral structure where lessons that happen later in the sequence depend upon material that was covered earlier. Because of this design, skipping particular lessons would be considered to be incongruent; whether a teacher skipped a lesson in *Everyday Mathematics* is one of the *Everyday Mathematics* indicators for whether a teacher is incongruent. In contrast, *Investigations* has a modular design. The curriculum does not require that teachers use all units and there is flexibility in the order that units are employed. Because of this different design, skipping a unit would not be considered incongruent and is not part of the set of indicators determining incongruence for *Investigations*.

After a coder completed the checklists for congruent and incongruent indicators, that coder would determine the overall lesson to be “congruent” through a holistic

¹³This seemed reasonable because it is not unusual for teachers to do non-textbook activities for a small portion of a class period. For example, they might review a skill such as “telling time” because an early dismissal has been announced for the day. On the other hand—because *Investigations and Everyday Mathematics* are comprehensive curricula with daily lessons—a teacher who failed to use them at all for one or more lessons (of the six observed lessons) would be considered to be an inconsistent user.

judgment of the lesson, using checkmark counts for congruence versus incongruence as a source of evidence for making that holistic judgment, as well as taking into account whether the teacher engaged in congruent instruction for the majority of the lesson.

A congruent set of three lessons within a semester is defined as a set of lessons where only one lesson out of three is incongruent. That is, the majority of the lessons within a semester had to be congruent.

Quality

The quality score is comprised of three measures: maintenance of cognitive demand, attention to student thinking, and intellectual authority. Our scale for maintenance of cognitive demand is based upon (1) the extent to which the teacher maintained the same cognitive demand for the primary instructional task from the materials phase to the set up phase; and (2) the extent to which the teacher maintained the same cognitive demand from the set up phase to the enactment phase. For each of these two transitions, we allocated 1–4 points to each teacher’s lesson in the following way:

- 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 one kind of high-level task into another type.¹⁴ 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 intent of the instructional task. Thus, a teacher received fewer points than if s/he 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 a task into another type of high-level demand or to a lower level of cognitive demand.

Through this point system, the maintenance of cognitive demand score could be from 2–8 points.

For scoring for *attention to student thinking*, teachers were assigned a score of 1 to 4 depending on the extent to which they uncovered student thinking and made it available to other students in a way that would help the class’s learning as a whole. The guidelines for score assignments were:

¹⁴Either from a “doing mathematics” task to a “procedures-with-connections” task or from a “procedures-with-connections” task to a “doing mathematics” task (Stein et al. 1996).

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 point 2, the teacher purposefully selected some students to share their work.

4 points—In addition to points 2 and 3, the teacher connected or sequenced students' responses in a meaningful way.

Finally, for scoring *mathematical authority*, teachers were assigned a score of 1 to 3 depending on the extent to which students had such opportunities in the lesson. The guidelines for score assignments were:

- 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.

Teachers with high quality instruction are differentiated from teachers with low quality instruction by establishing a “high quality” cut score for each of the three constructs: maintenance of cognitive demand (CD), teachers' work to uncover and productively use student thinking (ST), and the extent to which intellectual authority was vested in mathematical reasoning (IA). For CD, high quality was defined as an average score of 7.0 for teachers' lessons in one semester. For ST, high quality was defined as an average score of higher than 1.0 for all a teachers' lessons in a semester. For IA, we also set the cut score as higher than 1.0 for all a teachers' lessons in a semester. We set these cut scores based on our knowledge of each construct and our own expectations regarding what constitutes a high-quality lesson for that construct. Finally, we judged a teacher as having an overall high-quality set of lessons across the year if s/he scored as “high quality” for CD and *either* ST or IA. We did not require teachers to have a “high quality” score for both ST and IA with the rationale that both constructs equally reflect high-quality instruction and receiving a high score on one of the two constructs alongside a score above the cut for cognitive demand would reflect ample opportunity for student learning.

Assigning Instructional Profiles

For each semester, teachers were classified as flounders, mechanical implementers, canonical implementers, or mavericks according to their use, congruence, and quality ratings as described on pages 355–356.

Identifying Features of District Improvement Strategies

If patterns of cross-site differences and/or within-teacher development of instructional profiles over time were identified, we consulted findings associated with previously analyzed data to build conjectures regarding why the patterns emerged.

Results

We present the results according to the research questions, beginning with an assessment of the variation across teachers and sites in their levels of use, congruence and quality. The fourth question (contextual features associated with observed patterns) is addressed throughout the results section as patterns are identified.

Teachers' use, congruence and quality

As shown in Table 1, there was variation across the three dimensions of use, congruence and quality; quite noticeable variation between Region Z and Greene teachers; and some variation over time. We discuss each of these in turn.

The data in Table 1 suggest that teachers were more likely to use their respective curricula as the source of their classroom activities than to align their instructional practice with curricular and district guidelines. Approximately 80 %–90 % of the teachers used their curricula to a high degree (i.e., more than 75 % of the time) whereas as few as 53 % of the teachers (and never more than 71 %) exhibited instructional practice that was judged to be highly congruent with curricular and district guidelines. Each of these dimensions, however, exceeded teachers' capacities to demonstrate high-quality lessons. The percentage of teachers with high quality lessons hovered around 25 %, much lower than the first two dimensions.

Perhaps more interesting are the differences between Region Z and Greene in terms of use, congruence and quality. With respect to all three dimensions, Greene teachers exhibited higher levels at all time points except one.¹⁵ The differences are most marked with respect to quality and least marked with respect to use.

Variations over time are more difficult to detect. There do not appear to be strong differences over time in Region Z, but Greene teachers exhibited fairly substantial declines in congruence (from 95 % to 68 %) and in quality (from 53 % to 37 %) between the Spring of 2005 and the Fall of 2005.

What does all of this suggest? Early in these two district-wide initiatives, it appears to have been easier to obtain relatively high levels of use—and to maintain that high level of use over time—than to command greater teacher investments in terms of congruence or quality. In both districts, messages from central office were very clear: teachers were expected to use the new curriculum and principals would be checking to make sure that they were. Thus, mandates appear to work in terms of the lowest levels of compliance, that is, they drive teachers to take books out of their shrink wrap, distribute them to students, and teach out of them. Just one step beyond that, however, mandates are less effective. Many fewer teachers used the materials according to even the most superficial guidelines for their use (i.e., the

¹⁵Fall 2005 where 82 % of Region Z teachers were high users and 79 % of Greene teachers were high users.

Table 1 Teachers' Use, Congruence, and Quality

	Fall 2004 (n = 35)	Spring 2005 (n = 35)	Fall 2005 (n = 36)	Spring 2006 (n = 36)
High Use (use of mandated curriculum more than 75 % of the time in each lesson, on average)	89 % (31/35)	83 % (29/35)	81 % 29/36	78 % 28/36
... Region Z	81 % (13/16)	75 % (12/16)	82 14/17	71 12/17
... Greene	95 % (18/19)	89 % (17/19)	79 % 15/19	84 % 16/19
High Congruence (practices aligned with features at least 2 out of 3 lessons/semester)	71 % (25/35)	66 % (23/35)	53 % 19/36	58 % 21/36
... Region Z	44 % (7/16)	31 % (5/16)	35 % 6/17	41 % (7/17)
... Greene	95 % (18/19)	95 % (18/19)	68 % (13/19)	74 % (14/19)
High Quality (high quality on cognitive demand and either student thinking or mathematical authority)	29 % (10/35)	29 % 10/35	22 % 8/36	28 % 10/36
... Region Z	6 % (1/16)	0 % 0/16	6 % 1/17	6 % 1/17
... Greene	47 % (9/19)	53 % 10/19	37 % 7/19	47 % 9/19

kinds of markers that principals would be looking for in their classrooms to indicate that teachers are being faithful to the curriculum). Finally, quality was, by far, the most difficult thing to achieve, suggesting that mandates alone cannot dictate transformations of practice. Given that such transformations require teacher learning, additional investments in the professional development of teachers appears to be required.

The differences across Region Z and Greene with respect to quality beg the question of possible differences in how Region Z teachers versus Greene teachers were supported. In earlier analyses of how these two districts created organizational environments to support their respective reforms (Stein and Coburn 2008), we found that Greene was able to create significant opportunities for teacher learning that

aligned with reform goals while efforts in Region Z coordinated teachers' actions but failed to spur meaningful opportunities for teacher learning. For example, while coaches played a role in both districts' reform efforts, the selection process used in Greene yielded better coaches. Not surprisingly, the substance of what coaches talked about with teachers and with principals was very different across the two districts. In Greene, coaches' interactions were more substantive and more focused on mathematics teaching and learning; in Region Z interactions primarily focused on how to manage the *Everyday Mathematics* materials, gathering manipulatives and other tools for teachers, and providing general pointers regarding how to plan for and teach a lesson with little or no discussion of mathematical content or student thinking. Similarly interactions in teacher communities in Greene were more likely to move beyond pacing and managing materials to also include more substantive conversations about instructional strategies, student learning, and at times, the mathematics itself. Also, the principals in Greene were more likely to receive training on the mathematics reform and to work closely with their mathematics coaches in assessing and improving instruction in teachers' classrooms. The principals in Region Z, on the other hand, either turned over the mathematics program completely to their coaches or used their coaches for non-mathematics tasks.

Finally, it appears as though use alone does not buy district leaders much if their ultimate goal is high-quality instruction. Despite use levels that were not much lower than Greene's, the vast majority of Region Z teachers' instructional practices were judged to be low quality. On the other hand, the data in Table 1 suggest that congruence may play a more influential role in creating high-quality instruction if for no other reason than substantially greater percentages of Greene teachers exhibit congruent instruction and also exhibit high-quality instruction (although at lower rates).

The decline that occur between the Spring of 2005 and the Fall of 2005 in Greene co-occurred with a policy shift. Specifically, new state-level requirements for teachers' professional development hours related to English as a Second Language instruction necessitated a much larger emphasis on ESL professional development at the district and school level in Greene, which led to many fewer opportunities for teachers to engage in mathematics professional development (Kaufman and Stein 2010). In addition, a newly hired superintendent made it clear that teachers were free to use whatever materials they wished to address learning goals and, especially, ESL concerns. In other words, *Investigations* was no longer a mandated curriculum. Interestingly, from the Spring of 2005 to the Fall of 2005 teachers showed less decline in their use of the curriculum as the source of their daily activities (from 89 % to 79 %) than they did in congruence which dropped quite precipitously (from 95 % to 68 %). Perhaps this reflects the fact that teachers had been forced to relinquish their old curriculum materials and thus had no other materials on hand. Quality declined less, suggesting that teachers had developed some internal capacity to teach mathematics at a high level without necessarily following a specific curriculum.

Table 2 Teachers' instructional profiles

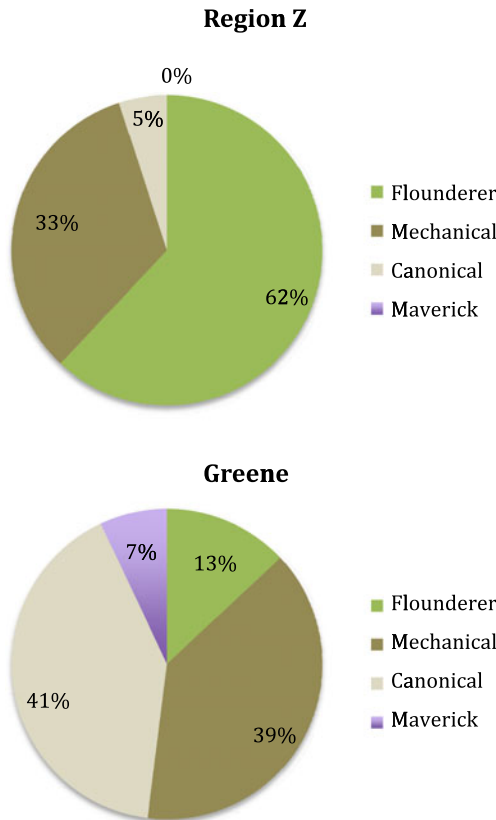
	Fall 2004 (n = 35)	Spring 2005 (n = 35)	Fall 2005 (n = 36)	Spring 2006 (n = 36)	Overall
Flounderer	29 %	34 %	42 %	39 %	
	10/35	12/35	15/36	14/36	
... Region Z	56 %	69 %	65 %	59 %	62 %
	9/16	11/16	11/17	10/17	41/66
... Greene	5 %	5 %	21 %	21 %	13 %
	1/19	1/19	4/19	4/19	10/76
Mechanical	43 %	37 %	36 %	31 %	
	15/35	13/35	13/36	11/36	
... Region Z	38 %	31 %	29 %	35 %	33 %
	6/16	5/16	5/17	6/17	22/66
... Greene	47 %	42 %	42 %	26 %	39 %
	9/19	8/19	8/19	5/19	30/76
Canonical	26 %	26 %	17 %	28 %	
	9/35	9/35	6/36	10/36	
... Region Z	6 %	0 %	6 %	6 %	5 %
	1/16	0/16	1/17	1/17	3/66
... Greene	42 %	47 %	26 %	47 %	41 %
	8/19	9/19	5/19	9/19	31/76
Maverick	3 %	3 %	6 %	3 %	
	1/35	1/35	2/36	1/36	
... Region Z	0 %	0 %	0 %	0 %	0 %
	0/16	0/16	0/17	0/17	0/66
... Greene	5 %	5 %	11 %	5 %	7 %
	1/19	1/19	2/19	1/19	5/76

Teacher Profiles

As shown in Table 2, teachers were unevenly distributed across the four profiles. There are—once again—noticeable differences in Region Z teachers versus Greene teachers; and there is some change over time. Each of these is discussed in turn.

Across all four time periods teachers were most likely to be classified as flounders or mechanical implementers. As noted earlier, these profiles reflect low-quality implementations with the difference being that the mechanical implementers are using the curriculum as the source of activities for the majority of their lesson activities and are attempting to follow curricular and district guidelines regarding how to use the curriculum while the flounders are not attempting to follow guidelines and, in some cases, were making limited or no use of the materials. There were many fewer canonical implementers, and fewer still, mavericks.

Again, there were differences between the two districts, but also one important similarity. Similar percentages of Greene and Region Z teachers were classified as mechanical implementers (39 % and 33 % respectively). However, despite these similarities, the balance of the teachers in Region Z tended to be flounders while the balance of the Greene teachers were canonical implementers. Thus, the overall distribution of teachers in each of these profiles looks very different across the two districts.



Over time, the largest change appears between the Spring 2005 and Fall 2005 time periods in Greene when the percentages of flounders increased from 5 % to 21 % and the percentages of canonical implementers decreased from 47 % to 26 %. There were no noticeable changes over time in the Region Z data.

What does all of this suggest? Across the early years of district-wide improvement efforts, the two districts’ approaches to mandated, curriculum-based reform appeared to have yielded a lot of flounders and mechanical implementers, neither of which, according to our definitions, was providing high quality opportunities for student learning. As noted earlier, mandates alone do not appear to work in producing high-quality instruction.

Table 3 Region Z within-teacher instructional profiles

		Fall 2004	Spring 2005	Fall 2005	Spring 2006
Predominantly Flounderer (59%)	EN	Flounderer	Flounderer	Flounderer	Flounderer
	EB	Flounderer	Flounderer	Flounderer	Flounderer
	BT	Flounderer	Flounderer	Flounderer	Flounderer
	SD		Flounderer	Flounderer	Flounderer
	BE	Mechanical	Flounderer	Flounderer	Flounderer
	DS	Mechanical	Flounderer	Flounderer	Flounderer
	BH	Flounderer		Mechanical	Flounderer
	PQ	Flounderer	Mechanical	Flounderer	Flounderer
	QJ	Flounderer	Flounderer	Flounderer	Mechanical
Flounderer / Mechanical Mix (18%)	MD	Flounderer	Flounderer	Flounderer	Mechanical
	UW	Mechanical	Flounderer	Mechanical	Flounderer
	KT	Flounderer	Mechanical	Mechanical	Flounderer
Predominantly Mechanical (18%)	HQ	Mechanical	Flounderer	Flounderer	Mechanical
	OG	Flounderer	Mechanical	Mechanical	Mechanical
Other (6%)	UF	Mechanical	Mechanical	Mechanical	Mechanical
	TT	Canonical	Flounderer	Flounderer	Mechanical
	NC	Mechanical	Mechanical	Canonical	Canonical

Despite both districts having similar numbers of mechanical implementers, however, Greene appears to have been able to foster a non-trivial amount of canonical implementation, meaning that teachers were using the district curriculum to create worthwhile learning opportunities for students. Thus, it appears as though mandates accompanied by support for teacher learning can yield positive outcomes related to quality. Finally, as noted earlier, our proposed pathway of teacher learning suggests that well-supported teachers develop from mechanical implementers to canonical implementers. The data in Table 2 suggest that this *might have* happened in Greene, but not in Region Z. We now turn to a within teacher analysis over time to examine this claim.

Patterns Over Time

Because the patterns over time are so different in Region Z versus Greene we will discuss the teachers in the two districts separately.

As shown in Table 3, across the two-year period, 10 Region Z teachers (59 %) displayed a predominately flounderer profile, never growing out of that profile for more than one time period. However, many teachers who stayed with the curriculum over time—and even tried to follow their guidelines (the mechanical implementers)—also never improved to a canonical profile. As shown in Table 3, a group of three teachers toggled back and forth between the mechanical and flounderer patterns and two teachers by and large remained mechanical implementers throughout the two-year period. Finally, in the “other” pattern, we find one teacher who appeared to actually progress nicely from a mechanical to canonical implementer and another teacher who is hard to classify.

Table 4 Greene within-teacher instructional profiles

	Teacher	Fall 2004	Spring 2005	Fall 2005	Spring 2006
Predominantly Canonical and Maverick (37%)	LH-6	Canonical	Mechanical	Maverick	Canonical
	NQ	Canonical	Mechanical	Canonical	Maverick
	DN	Canonical		Maverick	Canonical
	CD	Canonical	Canonical	Mechanical	Canonical
	BX	Canonical	Mechanical	Canonical	Canonical
	KE	Canonical	Mechanical	Canonical	Canonical
	XN	Canonical	Mechanical	Canonical	Canonical
Canonical Mechanical/ Mix (26%)	FT	Mechanical	Canonical	Flounderer	Canonical
	TS	Mechanical	Canonical	Flounderer	Canonical
	LI	Mechanical	Canonical	Mechanical	Canonical
	NN	Canonical	Canonical	Mechanical	Mechanical
	UN	Mechanical	Canonical	Canonical	Mechanical
Predominantly Flounderer /Mechanical (37%)	KN	Flounderer	Flounderer	Flounderer	Flounderer
	QS	Mechanical	Mechanical	Flounderer	Mechanical
	SH	Mechanical	Mechanical	Mechanical	Mechanical
	DT	Maverick	Mechanical	Mechanical	Flounderer
	SN	Mechanical	Canonical	Mechanical	Flounderer
	LH-7	Mechanical	Canonical	Mechanical	Flounderer
	WH	Mechanical	Canonical	Mechanical	Mechanical

What do these patterns suggest? A closer look at the preponderance of flounderers, who never improved (the first group), reveals that one school contributed 5 out of the 10 teachers. Unlike the rest of our focal schools in Region Z, this particular school was not “on board” with the mandated nature of the mathematics reform. From the start, it was clear that the principal sanctioned a wide variety of materials in addition to—or in place of—*Everyday Mathematics*, often claiming the rationale that there were “rumors” that the district was going to switch to a different curriculum (Sutherland et al. 2007). Moreover, the coach in this school took on a range of duties beyond that of mathematics coach.

The only Region Z teacher who improved over time (NC), came from a school in which there was some degree of conscientiousness about following the curriculum, including the help of a coach who proclaimed to be a new convert to the *Everyday Mathematics* curriculum. However, three of NC’s colleagues in the school (HQ, OG, and UF) did not progress as she did, but rather remained trapped in a mechanical profile (OG, UF) or a flounder/mechanical mix (HQ).

As shown in Table 4, the Greene teacher patterns (with the exception perhaps of the final group) are very different than the Region Z teacher patterns. First, there are 7 teachers (37 %) who, for the most part, stay within the two high-quality profiles, either canonicals or mavericks. Interestingly, these 7 teachers appeared to be “strong out of the gate,” that is, they displayed a canonical profile at the first data collection point (the reader is reminded however, that the first data collection point was the beginning of the second year of the reform in both districts).

The second group of teachers displayed a mixture of canonical and mechanical profiles. The first three teachers are especially interesting because they began with

a mechanical implementation but ended as canonical implementers. The fact that these same three teachers “slipped” into a flounderer or mechanical profile in the Fall of 2005 is interesting because that is when the new superintendent lifted the mandate to use the *Investigations* curriculum. Finally, the third group of Greene teachers appear similar to the Region Z teachers in that (except for KN) they all tried to use the curriculum at some point (there is a preponderance of mechanical implementations), but were rarely able to break into a sustained canonical profile.

What can we make of the Greene patterns? First, the teachers who were predominantly canonical or maverick were never flounderers (the first group). This suggests that perhaps floundering should be a red flag to observers or evaluations. It is sometimes argued that teachers should be permitted to go with their own decisions regarding curriculum; this study suggests that this will not lead to high quality—whether canonical or maverick.¹⁶

Closer examination of the first consistently good profile (the canonical/maverick group), reveals that 4 out of the 7 teachers came from one school, a school that had a principal who was a former mathematics coach and a consistent supporter of the *Investigations* curriculum. Even when the district pulled back its support of the reform, this school continued to support mathematics teachers with coaching and professional development (Kaufman and Stein 2010).

An optimistic interpretation of the next group—the five teachers labeled as the canonical/mechanical pattern—could be that they are “on the way” to sustained canonical implementations. Four of these five teachers came from the same school. The principal was an advocate of the reform and the *Investigations* curriculum during the early years, but then embraced the freedom to supplement in year 3 when the new superintendent lifted the mandate. Interestingly, at that point in time, four of the teachers slipped into a lower-quality profile; the fact that all but one re-emerged as a canonical implementer suggests that they perhaps had really learned from the earlier years implementing the curriculum and were thus able to reconfigure their practice at that higher level after “flirting” with the freedom from the mandate.

Finally, Greene was not immune from the flounderer/mechanical pattern that only rarely develops beyond low-quality instruction. This group of teachers came from all four of our schools, suggesting that no one school was immune from it as well.

Closer examination of the differences in the patterns *across* the two districts suggests that, perhaps, one leg of our proposed pathway of teacher development—from mechanical to canonical implementer—did indeed occur and it occurred more in Greene than in Region Z. We have already discussed how these two districts organized very different opportunities for teacher learning associated with their respective reforms. We have not, however, examined the curricula that each district selected to anchor their reform. As noted earlier, both were standards-based reforms. Both provided access to high-level, cognitively demanding tasks. However, one (*Investigations*) was found to have substantially more educative features than the other (Stein and Kim 2009). The *Investigations* materials more often identified the big

¹⁶There were no flounderers that led to mavericks.

mathematical idea at play in the lesson (and often provided a brief tutorial on it), thereby allowing teachers to apprehend the purpose of the activities in which they were about to engage their students. These materials also helped teachers to anticipate how students might respond to the many open-ended activities, thus helping them prepare ahead of time for how they might handle divergent and otherwise unexpected student responses. *Everyday Mathematics*, on the other hand, tended to have less open-ended tasks and to channel students and teachers toward a particular route through the problems. Teachers are provided with few in-depth details regarding how students might be expected to respond to the problems.

In short, the two curricula can be viewed as taking different stances toward teacher learning. *Investigations* does not script the teaching and learning that should occur in the classroom believing that student learning is always an emergent phenomenon, one that teachers must be attuned to through their attention to student thinking. As such, it helps teachers to (a) develop a nuanced understanding of the mathematical content to be learned; and (b) ways in which students might address this content. By doing so, it is investing in the teacher as an important element in the teaching and learning equation. *Everyday Mathematics*, on the other hand, appears to place the bulk of the expected learning between the student and the materials, with the teacher acting as a deliverer of those materials. Much less investment in the teacher is provided. Thus, another contributing factor to the greater preponderance of mechanical implementers developing into canonical implementers in Greene may be that the curricular materials provided greater transparency about their intent and potential student responses and, as such, helped teachers to move beyond a superficial, follow-the-directions style of implementation.

Conclusions

This work has implications for research on characterizing mathematics instruction within the context of district improvement strategies that rely on curricula and for research on teacher learning pathways. In addition, local policy makers could use findings generated here to help inform their designs for large-scale, curriculum based reforms.

Characterizing Instruction

The utilization of three dimensions (use, congruence, and quality) to characterize instruction offers a multi-dimensional view of instructional practice within district-wide, curriculum-based reform efforts. The fact that these three dimensions varied independently from one another suggests that each is offering a unique contribution to characterizing the nature of instruction. Yet, most often, only congruence *or* quality is measured. We believe that our framework represents an advance for

the field of research on curriculum implementation and that it can serve as a unifying framework for future studies of large-scale teacher improvement within the context of district managed curricula. This includes our method of delineating four profiles of instruction, which our results suggest are viable as well. These profiles (flounderer, mechanical, canonical and maverick) captured variation across the teachers and appeared to be responsive to difference in contexts across the two districts.

Characterizing Teacher Development

The findings do not suggest that we have identified a clear, uniform pathway for teacher development within the context of district-wide, curriculum based reforms. Instead of straightforward development from mechanical to canonical to maverick, some of our data toggled back and forth between two or more different profiles. This raises questions about the instructional profiles as reliable platforms on the road to teacher improvement.

The results suggest a more uneven pathway toward high-quality instruction than we had proposed. First, teachers who achieved a canonical implementation did not always stay at that level of implementation. In most cases, they exhibited mechanical (or even flounderer) profiles after they had achieved a canonical implementation. We conjectured that these “declines” may sometimes have been related to changes in district-level enforcement of the curricular mandate. Another potential contributor could be the topic. Perhaps a mechanical profile was exhibited because the teacher was on a challenging topic (for her) and therefore more comfortable with a procedural, follow-the-book style of implementation. Or perhaps the teacher changed grade level and therefore did not have her earlier command of the conceptual field.

The results do, however, support our notion that teacher pathways are relevant with respect to a particular context. Both districts were in the midst of large-scale, curriculum-based reforms. However, past analyses suggested that the amount and type of support each district provided for teacher learning varied significantly. The present analysis suggests that, under a supportive context, more than half of the teachers may be able to achieve canonical implementations; in a less-supportive context, however, the vast majority of implementations will most probably consist of a mixture of flounders and mechanical implementations.

Finally, the fact that there were few mavericks suggests that a common concern raised about district-based curricular reforms may not be warranted. Often, critics complain that excellent teachers are muffled by heavy-handed, top-down district reforms that force them to use a particular curriculum. The low incidence of mavericks in our data set (even after the mandate was lifted in Greene in the second year) suggests that this worry may be unfounded. A much larger worry, on the other hand, is the large number of flounders and mechanical implementers that such reforms may foster.

Implications for Large-Scale District Reform

The findings reported herein suggest that expecting all teachers to implement standards-based curricula places a huge responsibility on the district to—not only monitor where teachers are on any given date—but also to support them as they try out new and often unfamiliar materials. By positioning teacher development against the backdrop of various ways in which teachers implement the district curriculum, this study's findings provide important foundational knowledge for the development of efficient and effective large-scale teacher support systems in environments characterized by district-wide managed curriculum.

Our findings suggest that district policies must go beyond mandates. Alone, mandates delivered only the lowest level of implementation: use. They were relatively ineffective for assuring that teachers implement the curriculum in a way that is aligned with the pedagogical guidelines in the curriculum and with district guidelines. They were not effective in delivering quality. The canonical implementers were almost exclusively in Greene, the district with effective support systems accompanying their roll-out of the new curriculum.

Our study also suggests ways in which our framework might be useful to the design of district support systems. Knowing the profiles of teachers in one's school or district would be useful in planning professional development. Not only do teachers with different profiles require different kinds of professional development (the maverick could be challenged by innovative offerings outside the district while the flounderer needs basic support), but teachers can be paired with one another in ways that take advantage of their differences. For example, a mechanical implementer could learn from a canonical implementer, but leaders would not want to send a flounderer into a maverick's classroom because—although it would be high-quality—without a curricular roadmap, it would be unclear to the flounderer how the teacher accomplishes what she does. Overall, considering various implementation profiles in the context of district-wide, curriculum-based improvement efforts is a promising approach to both diagnosing teachers' needs and identifying and using the strengths already present in the district (the canonical implementers) to address those needs.

References

- 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.
- 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.
- Borko, H., & Putnam, R. (1995). Expanding a teacher's knowledge base: a cognitive psychological perspective on professional development. In T. Guskey & M. Huberman (Eds.), *Professional development in education* (pp. 35–65). New York: Teachers College Press.

- Brendehur, J., & Frykholm, J. (2000). Prompting mathematical communication in the classroom: two preservice teachers' conceptions and practices. *Journal of Mathematics Teacher Education*, 3, 125–153.
- Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning*, 6(2), 191–221.
- 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: Erlbaum.
- Coburn, C. E. (2001). *Making sense of reading: logics of reading in the institutional environment and the classroom*. Ann Arbor: University Microfilms.
- Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, 30(3), 203–235.
- 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(2), 159.
- Duessen, T., Coskie, T., Robinson, L., & Autio, E. (2007). "Coach" can mean many things: five categories of literacy coaches in reading first. REL Northwest (REL 2007-No. 005). IES: US Department of Education.
- 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. A., & Conant, F. C. (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.
- Fennema, E., & Nelson, B. (Eds.) (1997). *Mathematics teachers in transition*. Hillsdale: Erlbaum.
- 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.
- Hamm, J. V., & Perry, M. (2002). Learning mathematics in first-grade classrooms: on whose authority? *Journal of Educational Psychology*, 94(1), 126–137.
- Hightower, A. M., Knapp, M. S., Marsh, J. A., & McLaughlin, M. W. (Eds.) (2002). *School districts and instructional renewal*. New York: Teachers College Press.
- 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. H., & Stein, M. K. (2010). Teacher learning opportunities in a shifting policy environment for instruction. *Educational Policy*, 24(4), 563–601.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: mathematical knowing and teaching. *American Educational Research Journal*, 27(1), 29–63.
- Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven: Yale University Press.
- 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: Erlbaum.
- 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: MIT Press.
- Schoenfeld, A. S. (1998). Toward a theory of teaching-in-context. *Issues in Education*, 4(1), 1–95.
- 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 clas-*

- sical pedagogy: teaching elementary school mathematics* (pp. 109–134). Mahwah: Erlbaum.
- Stein, M. K., & Coburn, C. E. (2008). Architectures for learning: a comparative analysis of two urban school districts. *American Journal of Education*, *114*, 583–626.
- Stein, M. K., & Kaufman, J. (2010). Selecting and supporting the use of mathematics curricula at scale. *American Educational Research Journal*, *47*(3), 663–693.
- 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.), *Mathematics teachers at work: connecting curriculum materials and classroom instruction* (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., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: an analysis of mathematical tasks used in reform classroom. *American Educational Research Journal*, *33*(2), 455–488.
- Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: helping teachers learn to better incorporate student thinking. *Mathematical Thinking and Learning*, *10*(4), 313–340.
- Supovitz, J. A. (2006). *The case for district-based reform: leading, building and sustaining school improvement*. Cambridge: Harvard Education Press.
- Sutherland, S., Smith, J., & Wallace, C. (2007). *The missing link? The coaches' role in policy translation*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- 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: Erlbaum.
- Wilson, S. (1990). A conflict of interests: the case of mark black. *Educational Evaluation and Policy Analysis*, *12*(3), 309–326.