Learning Lessons from Instruction:

Descriptive Results from an Observational Study of Urban Elementary Classrooms

Heather C. Hill, Erica Litke

Katherine Lynch, Cynthia Pollard and Barbara Gilbert

Harvard Graduate School of Education

Draft:

Do not cite or publish

Author Note

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C090023 to the President and Fellows of Harvard College to support the National Center for Teacher Effectiveness. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

For decades, U.S. policy-makers and professional associations have pushed for more ambitious teaching and learning in the nation's schools, seeking more disciplinary integrity and higher cognitive demand in classrooms. In mathematics, the Common Core State Standards (CCSS) ask teachers to foster student fluency with procedures, make mathematics relevant through practical applications, and foster conceptual understanding, all competencies found in older policy and reform documents (e.g., National Mathematics Advisory Panel, 2008). In addition, the CCSS also asks for more emphasis on disciplinary practices, such as developing students' abilities to model mathematically, make mathematical arguments, and generalize from patterns (NGA, 2010).

The instructional shifts required by these reforms may be challenging, particularly for U.S. urban school districts where, despite an influx of resources and committed staff, schools are often faced with high percentages of struggling students, competing reform initiatives, and variable teacher quality (e.g. Jacob, 2007; Lankford, Loeb, & Wyckoff, 2002). Nevertheless, instructional improvement in urban districts has been an important focus of U.S. federal and state policy for decades, as progress in these districts is seen as key to closing the academic achievement gap and leveling the playing field for low-income, minority students. Understanding where urban districts are, in terms of progress toward state and national standards, may help the field understand how much progress has been made in the last decades, and how far such districts need to go.

To address this issue, we conducted a video-based observational study of mathematics instruction in five urban school districts on the Eastern seaboard of the US. These districts varied in terms of size, student population, and history of efforts to implement prior mathematics reform initiatives. Teachers allowed us to capture up to six videos per person over a two-year time span (2010-2012). Our final video dataset, which included 1735 videos from 329 teachers, was analyzed via descriptive memos and also coded using the Mathematical Quality of Instruction (MQI), a classroom observation protocol that captures both Common Core practices and other critical elements within classroom mathematics instruction (Hill et al., 2008). Using both the descriptive memos and the MQI coding, we can ask:

- What is the nature of mathematics instruction in these urban classrooms?
- Does instruction vary by student population?
- Does instruction vary substantially by district? If so, why?

Findings suggest that there are a significant number of high-quality mathematics lessons in this sample; in many other lessons, teachers deploy specific reform practices but have largely traditional instruction and occasional mathematical imprecisions, as suggested in older studies in this vein (e.g., Cohen, 1990). We found considerable variation in instruction by district, with the district with the most coherent and long-lasting reform effort in mathematics having the strongest overall instruction.

Literature Review

Research prior to the 1990s contained few glimpses into classrooms, yet those that did exist suggested largely teacher-directed, didactic lessons constituted the steady diet for most American students. In Larry Cuban's (1993) outstanding book, *How Teachers Taught*, Cuban traces constancy in U.S. schoolhouses from the 1890s through the 1980s. Despite waves of progressive reforms encouraging student-centered practices during these years, only incremental change appeared to occur. Where students in the 1890s engaged in silent study, drill, and recitation at the direction of the teacher, students in the 1980s still experienced largely teacher-led instruction with little student talk (p. 219). Using data from classroom visits conducted while he was superintendent of the Arlington, VA public schools, Cuban wrote "If [the topic] were math, social studies, science or language arts, the teacher would work from a text with the entire class answering questions from it or either from dittoed sheets or workbooks" (p. 218). However, the modern incarnation of teacher-centered classrooms also contained subtle differences from the 1890s. In many classrooms, seating arrangements had changed, with students sitting in small groups and having some latitude for self-directed movement around the classroom.

Even when more substantial reform took root, it rarely spread widely. Cuban's account notes stories of progressive reforms' implementation in some urban classrooms, as reformers like Edward Sheldon, Francis Parker, and John Dewey sought to transform the largely drab instruction they found therein. In many classrooms – particularly early-grade classrooms – teachers followed and enabled students' interests, encouraged activity-based learning, and allowed student collaborative work and discussion. Work in such classrooms was often documented by reformers, school officials, journalists, and a nascent professional teacher educator corps, all apparently eager to celebrate and publicize success. Yet Cuban soberly evaluates the spread of these ideas, estimating that they reached only 10 to 20 percent of classrooms (p. 135); he also notes that reforms often appeared in hybridized form, as teachers took on some elements, such as increased student expression, small group work, and activities, and forsook others that required more fundamental changes in classroom life.

More modern reform efforts, including those that aimed to change the level of involvement of students in lessons and to foster changes in the way content was presented, have met similar fates. In mathematics, efforts began in the 1980s to increase the proportion of applied problems in the curricula, the cognitive demand with which students engage those problems, and student understanding of fundamental mathematical principles rather than rote memorization (Cohen, McLaughlin & Talbert, 1993; Lampert, 2001). Similar to the primary sources describing reformed classrooms in Cuban's work, scholars during this period focused their energy on transformed classrooms, crafting carefully questioning (Cohen, 1990; Heaton, 1992) and occasionally celebratory portraits of classrooms outside the norm. Yet as Cuban found when he looked at classrooms broadly prior to 1980, more representative surveys post-1990 found little evidence that reform entered classrooms at scale. In a path-breaking study using video to examine a nationally representative sample of eighth grade classrooms, Hiebert and colleagues (2005) found that a small but nationally representative sample of 8th grade U.S. mathematics lessons were characterized by a low level of mathematical challenge. A decade later, the *Measures of Effective Teaching* (MET) study found that mathematics instruction in six urban districts contained few disciplinary practices or sense-making, opportunities for students to engage in mathematical reasoning, or teacher use of student ideas. Only 1% of lessons in this study scored at high levels on these three competencies and for the vast majority (70%) of lessons, none of these strengths were present (Kane & Staiger, 2012). Weiss, Pasley and colleagues (2003) observed a widely dispersed sample of 57 K-5 mathematics teachers and found that while classrooms were generally well-resourced and well-managed, there was also a low level of intellectual rigor (p. C-20) in most. Of class time, 59% was spent in large group settings, with the rest spent in small group or individual work time (p. 24).

Similarly, such studies depicted the content of most mathematics lessons as largely procedural. Hiebert et al. (2005) found that 91% of problems in 8th grade lessons were solved through the use of procedures or by simply stating an answer. Likewise, only 34% of the problems solved in an average lesson involved the application of mathematical ideas. Of Weiss, Pasley and colleagues' observations, only one-quarter received a high rating for the dimension involving mathematical sense-making; in contrast, 40% received a rating that suggests purely procedural instruction (p. C-17), with the remainder in the middle.

Finally, numerous scholars during this period noted less than ideal features of mathematics classrooms. Weiss and colleagues (2003) estimated that 8% of observed class time was spent on non-instructional activities; MET, a decade later, reported that 6.6% of mathematics lessons contained segments in which the classroom work was not focused on mathematics for more than half of the 7.5 minute segment (Kane & Staiger, 2012). Furthermore, the accuracy of mathematics instruction appeared high but not perfect; Weiss et al. (2003) reported that lessons scored a 3.73 out of 5 on this indicator (C-18). Hill, Kapitula and Umland (2011) also suggest that some instruction is mathematically problematic.

In all, a comparison of mathematics instruction across decades suggests only incremental changes have taken root despite tremendous effort by reformers. Why instruction remains constant has been the subject of much speculation. At the most basic level, support from central office administrators and school officials appears a necessary condition for reform to make modest gains into classrooms. Cuban (1993) notes that "Where these [reforms] seemed to appear in strength were in school districts where top administrators gave formal approval for the effort, established organizational machinery to carry it out, and persisted in its implementation" (p. 136). More recently, Stein and Coburn (2008) compared two urban districts' efforts to implement new, ambitious mathematics curricula. Only in the district where principals engaged in meaningful interactions with coaches, principals were significantly involved in discussions and decision-making around instructional improvement in mathematics, and the focus was on core changes to mathematics teaching and learning did teachers have substantial opportunities to learn and implement the districts' intended instructional reforms.

Another, related explanation for limited adoption of reformers' ideas comes from the environments in which schools and teachers operate. In this line of thinking, successful reform requires the confluence of many elements: not only supportive district and school administrators, but also curriculum materials that support reform's instructional goals, professional development that supports teachers learning about curriculum materials and instruction, and assessments that both incent attention to new ways students can learn subject matter and provide opportunities for teachers to learn new ways of teaching (Cohen & Hill, 2001). In this narrative, elements missing from this puzzle can subvert reform as well. Portraits of schools responding to accountability pressure in the late 1990s and 2000s, for instance, detailed a stunning variety of responses (e.g., narrowing of the curriculum, focusing on "bubble kids," explicit instruction in test-taking strategy; see Diamond & Spillane, 2004), but rarely did any reports describe accountability making instruction more cognitively demanding or disciplinarily rich for students. In fact, Diamond (2007) argues that despite a policy environment that suggests reform instruction will help resolve race and class-based inequalities, conventional instruction persists because of the incentives embedded in basic-skills state tests.

To explain within-district variability, and to explain the adaptation of reform practices into more typical formats, scholars often note that teachers filter reform initiatives through pre-existing knowledge and beliefs (Cohen, 1990), making sense of the new in light of the old. Thus reform ideals, such as introducing mathematical ideas through more authentic problem-solving situations, regressed into older formats (e.g., story problems) as reforms spread at scale. As in Cuban's account, this meant that teachers changed reform more than reform changed teachers (Cohen, 2011; Cuban, 2013); incremental advances in classroom instruction may have occurred over this period, but wholesale revision to instruction was seldom seen.

Equity, specifically whether students of different races received access to the reform elements that did exist, has been an important theme in this literature. Cuban (1993) recounts depictions of rural schools serving African-American students during the period 1920-1940; in addition to describing terrifically poor resources (including lack of instructional materials and unsafe school buildings), he argues that few of reformers' ideals made it into the classroom. Numerous studies also demonstrate that low-income, non-white students are less likely to be assigned to quality teachers (Hill & Lubienski, 2007; Jacob, 2007; Lankford, Loeb & Wyckoff, 2002), where quality is defined as teacher preparation, certification status, and knowledge.

Thus both historic and recent research provides little evidence that reform ideals have entered classrooms in their original form, and that when they do, they appear disproportionately in higher-status districts and classrooms before fading out. We seek to update and extend this literature, focusing on the extent to which mathematics reform efforts begun in the 1990s (NCTM, 1989; 2000) have penetrated classrooms in five urban districts. By using an unusually detailed coding system, we can match reform ideals along multiple dimensions to enacted practice; because we have learned about district efforts to enact these reforms, we can also match patterns in implementation to patterns in support.

Methods

This paper features a mix of descriptive and exploratory analysis. It is descriptive in the use of structured observations and analytic lesson memos to characterize classroom instruction across multiple districts. It is exploratory in the linking of observed patterns in instruction to district background (for a similar study, see Stein & Coburn, 2008).

Sample

The study sample was comprised of 329 fourth- and fifth-grade teachers from five urban school districts in the Eastern United States. We recruited schools based on district referrals and size, requiring that schools have at least two teachers in each sampled grade. We included in the sample all teachers from the 2010-2011 and 2011-2012 academic years who completed at least three videotaped lessons. The final dataset includes 1,735 videotaped lessons.

The five participating districts were situated in a mix of smaller cities and large metropolises, all serving large numbers of low-income and minority students. All five districts had taken actions supportive of the NCTM mathematics reforms of the 1990s; however, as we describe in further detail below, their patterns of support differed. Districts 1, 2, 3, and 5 had adopted textbooks supportive of the NCTM mathematics reforms, including *Investigations* and *Everyday Mathematics*, while District 4 had adopted *Harcourt Brace*, a more conventional textbook. All five districts had implemented teacher professional development in mathematics, although as we discuss below, the format and content of the professional development varied considerably. Table 1 shows demographics of students in the sample by district. We discuss characteristics of the sampled districts in further detail below.

Data collection

We use two primary data sources for the current study: video recordings of participating teachers' lessons, and interviews with district personnel. We describe each of these data sources below.

Video recordings. We recorded teachers' mathematics lessons over two academic years. Each teacher was videotaped a minimum of three and a maximum of six times. Lessons were videotaped using a three-camera unmanned videocapture rig, which project staff turned on at the beginning of each lesson and off at the end. Each lesson lasted approximately 45-60 minutes. Teachers were permitted to choose the days on which they were videotaped, with the requirement that they select lessons typical of their regular instruction and exclude days on which students would be taking tests.

Each of the 1,735 lessons in the current dataset was scored by trained, certified raters using the *Mathematical Quality of Instruction* (MQI) classroom observation instrument (Hill, Blunk, Charalambous, Lewis, Phelps, Sleep, & Ball, 2008). The MQI is designed to capture the mathematical quality of elementary and middle school classroom instruction. Prior research indicates that teachers' scores on some dimensions of the MQI instrument are positively associated with their students' mathematics achievement (e.g. Hill, Charalambous, & Kraft, 2012; Hill, Kapitula & Umland, 2011). Prior to rating videos, raters were required to participate in initial training and pass an initial certification exam. Raters participated in ongoing calibration seminars in order to maintain rating consistency and minimize rater drift (see Hill et al., 2012). Raters were blind to the district from which each classroom lesson video was derived. Two trained raters watched each videotaped lesson, and scored each 7.5-minute segment on 17 MQI codes.

The current analysis is focused on three major dimensions of the MQI: (1) the Richness of the Mathematics (*Richness*); (2) Teacher Errors and Imprecision (*Errors*); and (3) Common Core Aligned Student Practices (*CCASP*). Each dimension is comprised of between two and five items, as well as a holistic code that captures raters' summary judgment of the segment's overall quality on that dimension. *Richness* captures the depth of the mathematics provided to students, including the extent to which the mathematics offered in the lesson emphasizes the meaning behind facts and procedures or important mathematical practices. *Errors* captures teacher errors or imprecisions of language or notation, or lack of clarity or precision in the teacher's delivery of mathematical content. *CCASP* captures evidence of students' involvement in cognitively activating mathematical activities, such as providing explanations, engaging in reasoning, or posing mathematically motivated claims or questions. Table 2 displays a list of relevant codes.

District Interviews. After the conclusion of video data collection, interviews were conducted with the individuals who served as district mathematics directors or coordinators from each participating school district in the years during and leading up to our study. Interviews were conducted by the study authors and were semi-structured; interviewers utilized a protocol that included a standard set of questions and follow-up questions as needed. Each 45-75 minute interview was conducted by study authors via Adobe Connect and telephone conference line. Interviews were audio-recorded, and subsequently transcribed for analysis.

Analyses

Quantitative: In order to provide a descriptive picture of instructional quality in these urban districts, we calculated the frequencies with which teachers were observed participating in various instructional practices in their classrooms. Next, we utilized regression analysis to examine the relationships between instructional quality dimensions and student background characteristics. Further details are discussed below.

Illustrations of instruction: In order to provide illustrations of the instructional practices characteristic of teachers with different average levels of observed instructional quality, we conducted a qualitative analysis of lessons from teachers at the 'high,' 'middle,' and 'low' points of the instructional quality distribution. For each teacher in the sample, we calculated average overall MQI scores, as well as average scores shrunken for lesson number to address the fact that some teachers had more lessons than others. We extracted for analysis those teachers who showed fair consistency in their percentile ranking on both sets. The breakdown of reviewed lessons was as follows: (1) 8 teachers were categorized in the 'high' percentile (90th-90st percentile) group with a total of 26 lessons; (2) 16 teachers were categorized in the 'middle' percentile group (49th-51st percentile) with a total of 104 lessons; and (3) 15 teachers were categorized in the 'low' percentile group (8th-12th percentile) with a total of 64 lessons.

Lesson summaries, short documents describing the lesson written and reviewed by two raters during the lesson scoring period, provided insight into what instruction looked like in each of these strata. Two authors independently read through all lesson summaries at each level, noting themes observed in the instruction. The authors recorded and discussed their observations, and identified themes across lessons as indicative of instruction among lessons at that particular level. We then used the lesson summaries from operational scoring to select one teacher from each group whose lessons were illustrative of the overall profile we articulated. Two authors then watched videos of these teachers' lessons, taking notes and focusing on specific instances where we saw the aspects of instruction illustrative of the instruction at that level. Finally, these authors reconciled notes and developed case profiles illustrative of teaching at each level.

District Interviews: After the interviews were transcribed, analysis of the interview data was conducted by two authors. The interview questions were then utilized as a framework for developing in-depth case studies of two districts. Analysis of the interviews and the development of the case studies were conducted via an iterative process. Both authors independently reviewed the interview transcripts, then held meetings to discuss findings. As the case studies were developed, the authors continued to meet frequently in order to discuss findings and emerging themes that were identified from the cases.

Results

In the following three sections, we describe frequency of reform practices within classrooms, and examine whether these reform practices are distributed evenly across student demographic characteristics; we describe typical teachers in the center and tails of the MQI distribution; and we discuss the district effects observed in the data.

Frequency of MQI Features

We first begin by describing the observed frequency of reform practices within classrooms in our data. As each lesson was scored on the MQI by two raters, we selected for each lesson one set of scores from a randomly chosen rater. This procedure left us with 13166 segments in the 1735 lessons to describe frequencies of classroom phenomena.

Table 3 shows the prevalence of four instructional formats within our 1735 lessons. These formats are similar to those studied in prior literature, in that they capture the amount of teacher versus student control, and the proportion of the lesson in which students worked on applied problems relative to other types of mathematical activities. Similar to past accounts, direct instruction dominated; raters reported that in 61% of segments, teachers entirely controlled the delivery of mathematical content, including conducting presentations of material, going over homework or problems, launching tasks, or supervising student work time. Another 32% of segments were rated as partly teacher-controlled. In four-fifths (80%), the teacher was actively presenting new material for either the entire or part of the segment; in roughly 20%, students worked individually or in groups to complete tasks. Whole-class discussion (meaning students talking with one another, rather than back and forth through a teacher) occurred in some or all of only 5% of segments. Working on applied problems, where such problems were defined as any problem involving a contextualized situation (e.g. a short word problem to find "total distance traveled" from two segments of a trip), occurred in 28% of segments.

These statistics paint a picture of instruction in these urban districts as largely aligned with past depictions: teacher-led, with little student talk, and with only a modest number of applied/contextualized problems. But blunt descriptions of lesson format obscure other trends in the data. Although teachers still do most of the communication of mathematical content to students, *what* they convey appears different in this sample of teachers. In contrast to older characterizations of most U.S. instruction consisting of "procedures without connections," teachers in this sample made use of several meaning-making elements (see Table 4). Over a quarter of segments contained either an explanation for why a procedure works, why a solution method makes sense, why an answer is true, or what a solution means in the context of the problem. Another 29% of segments contained connections among different representations of mathematical ideas or procedures (e.g., a linear graph and a table both capturing a linear relationship) or among different mathematical ideas (e.g., fractions and division). Almost 15% of segments contained instruction in which teachers or students applied multiple solution methods to a single problem or a set of similar problems. Though these proportions may seem modest, they represent segment-level prevalence of these activities; aggregated to the lesson level, we see that 65% of lessons featured at least one explanation, 66% featured at least one connection among representations or ideas, and 45% contained multiple solution methods. In all, 85% of lessons had one of these three features.

Thus in contrast to past reports, we find a significant amount of meaning-oriented instruction occurring in classrooms, albeit meaning primarily conveyed via direct instruction. Our coding system makes possible one addition to this depiction: the quality of these elements. For the three practices above, the difference between a "mid" and "high" score is the level of detail, substance and precision in that practice. For example, a middle score for the mathematical explanation code refers to explanations that are partial, lack detail, or specific only to one problem (e.g., why five is the answer to 25 ÷ 5=5). A high score means that the explanation(s) in a segment were general in nature (e.g., in division, we partition the dividend into the number of groups shown by the divisor). A similar logic is true for the connections and multiple methods codes, where a high score signifies that the mathematical event is marked by explicit and/or detailed connections or comparisons between the representations, topics, or methods. These score distinctions were built into the MQI because of research showing the effectiveness of the practices noted under the high score point (e.g., Rittle-Johnson & Star, 2007). However, Table 4 shows that seldom are these instructional features enacted with such characteristics in our classrooms. About five percent of segments contained high-quality connections, and less than three percent of segments contained high-quality explanations. Detailed and explicit links between multiple methods was enacted in just over one percent of segments.

Two other codes within the MQI that focus on mathematical practices - teacher and students' use of mathematical language and developed generalizations - returned more discouraging results. Frequencies for the former, which captures the use of mathematical terms, show that while mathematical language is often used in these segments (42% of segments), it is rarely used densely or with explicit attention to developing fluency and precision in classroom talk (5% of segments). The latter, which captures instances in which teachers and/or students examine multiple instances or examples of a phenomena then make a general statement about that phenomena, was only coded at either mid or high in 4% of segments. Both are central to the Common Core practice standards, and results here suggest high-level use, as the practice standards suggest, is uncommon in these urban districts.

Turning now to the MQI codes that concern student involvement in cognitively challenging work, we see patterns that again appear different from past depictions of U.S. instruction. For instance, in one-fifth of segments, raters reported that students themselves offered at least one explanation for why a procedure works, why a solution method makes sense, why an answer is true, or what a solution means in the context of the problem. In nearly a fifth of segments, raters observed students engaging in mathematical questioning and reasoning of some sort, for instance by requesting an explanation (e.g., "Why does this rule work?"), making conjectures, claims or counter-claims about mathematical phenomena, forming conclusions based on patterns or engaging in reasoning in the general sense (e.g., "Because the angles in any triangle add up to 180 degrees, a triangle should have at least two acute angles.") And in 25% of segments, raters observed students engaged in providing at least some mathematical input into the development of the mathematics – for instance, offering an explanation or working on a cognitively demanding task with little teacher scaffolding. Though the fraction of segments in which such cognitively challenging work dominated the segment was small, at 3%, the at least partial engagement of students in mathematical thinking suggests they are not nearly as passive as past accounts might suggest. Overall, 84% of lessons contained at least one of these three elements, and 21% contained one of these elements at a high level.

The MQI also tracked the frequency with which the teacher presentation contained a mathematical error or lacked mathematical precision or clarity. Outright mathematical errors – e.g., teachers solving problems or defining terms incorrectly – were few, at 6% of segments. By contrast, 16% of lessons contained a language imprecision on the part of the teacher, for instance by saying "division makes numbers smaller" without specifying that this statement is true only for whole numbers, and without specifying which number becomes smaller. Raters judged another 10% of segments to feature unclear mathematical presentations on the part of the teacher. We illustrate the nature of such mathematical errors below, in our case studies.

Relationship with student characteristics

Next, we examined the relationship between these features and student demographics, to determine whether average instructional quality differed by the students served in a given classroom. As an initial step in this analysis, we aggregated scores for specific items to the teacher level, then created dimension-level scores composed of the items under each broad category. For instance, for the Richness dimension of the MQI, we aggregated the five items shown in Table 2 at the segment level to the teacher level, then averaged specific items across segments to arrive at a teacher-level score. We followed a similar procedure for each of the codes of the MQI. We then regressed the teacher-level scores on each dimension within a school year on the following classroom characteristics for that teacher in the same school year: percentage of students eligible for subsidized lunches, percentage of students designated as English Language Learners, percentage of students designated as special education, and average student prior mathematics achievement. We also controlled for differences in scores from year to year using a school year fixed effect; furthermore, our regression used robust standard errors and accounted for the fact that some teachers had multiple years of MQI and classroom-level data.

In Table 5, we present parameter estimates and associated *p*-values for each of the dimensions of the MQI. Examining Table 5, we see that a teacher's Richness and CCASP score in a given year is positively associated with that teacher's classroom percentage of ELL students and that teacher's classroom average prior math achievement. The teacher's classroom percentage of special education students is also negatively associated with the teacher's errors score in that given year (i.e., a higher percentage of special education students is associated with fewer errors) and positively associated with the teacher's CCASP score. Classroom-level eligibility for subsidized lunches did not demonstrate any significant associations with instructional quality in a given year. Thus we find that concern that disadvantaged student populations may receive lower-quality instruction is not, in this dataset, substantiated.

In sum, our study finds more diversity in the quality of instruction than previously found in many urban districts We continue now by making this depiction more concrete via illustrative case examples of three teachers.

Illustrations of instruction

Because the depiction of instruction presented above diverges from others commonly found in the research literature, we chose to illustrate instructional quality by examining teachers near the 10^{th}, 50^{th} and 90^{th} percentiles of the sample MQI score distribution, with the goal of choosing a teacher with typical instruction from each set. To do this, we first read documents created by raters as they scored the lessons from each teacher; for each lesson, raters described the lesson content and activities, and noted mathematical weaknesses and strengths. (Pairs of raters reconciled the lesson summaries.) By reading through these lessons summaries, we generated a list of themes at each level, then re-read the lesson documents described above, settling on a teacher at each percentile group who illustrated those themes. We watched all of the lessons from each of the selected teachers to develop a descriptive profile for that teacher. We present both the general themes and specific illustrations below.

Teachers at the 90^{th} Percentile

In analyzing summaries from teachers at or near the 90^{th} percentile, we noted that most lessons contained some elements of rich mathematics, such as explanations of mathematical concepts and mathematical sense-making. For instance, in a lesson on fractions, one teacher repeatedly emphasized why the bigger the denominator, the smaller the number of pieces when breaking up the whole. Lessons in this category also often included some elements of strong mathematical practices such as linking between representations, multiple solution methods, and strong mathematical language. These teachers frequently took advantage of correspondences between representations (e.g. fraction bars and numeric representations or word problems and equations) to make mathematical meaning in their lessons. In addition, lessons included only occasional small errors such as minor imprecisions or sloppiness in language (e.g. saying vertexes instead of vertices) or the omission of the parts of a definition or explanation (e.g. neglecting to say that the whole must be divided into equal pieces when discussing fractional parts). These imprecisions did not tend to obscure the larger mathematical point of the lesson.

Raters frequently noted that students were afforded the opportunity to actively engage in the mathematics in the lessons in this category. In particular, the questions asked by these teachers and the tasks they posed encouraged students to make meaning of the mathematics presented to them. Many of the lessons in this category included notable student mathematical contributions: students offering alternative solution methods, noticing mathematical patterns, or student explanations and reasoning. For example, in one lesson students were asked to classify shapes according to their characteristics and did so with significant reasoning comparing and contrasting shapes. In another lesson in which students worked on adding fractions by finding the common denominator and multiplied a fraction by 2/2, the teacher insisted that students explain why this process is considered multiplying by a whole, and prompted students to give full explanations of this concept. While lessons in this category seldom included all of the features listed above simultaneously, these features were observed frequently across lessons and teachers.

Thomas

After reviewing the lesson summaries, we selected one 5th grade teacher, Thomas, to illustrate instruction among teachers in the 90^{th} percentile. Thomas taught in a large, diverse urban district (District 14) with approximately half of its students enrolled in the Free and Reduced Price Lunch program and 15% designated as English Language Learners. In contrast to the rest of the district, Thomas' school has over 75% of students receiving Free and Reduced Price Lunch and in the second year of data collection, one third of his students were English Language Learners. Math instruction in this district overall was marked by direct presentation of content to students (as opposed to more inquiry-oriented lessons), few behavioral issues, and only sporadic attention to meaning-making around mathematical concepts (Hill et al., 2013). At the time of the study, Thomas had been teaching for 6 years, having arrived in his classroom via an alternative certification program. Thomas reported using both a standard mathematics textbook published by Harcourt, as well as a reform-oriented NSF-funded Harcourt series entitled *Think Math!* that focuses on a combination of conceptual understanding and computational fluency.

Overall, Thomas' instruction was consistently solid, similar to other teachers in this category. His lessons included frequent opportunities for students to engage with the mathematics in cognitively activating ways and provided opportunities for sense-making around mathematical concepts and ideas. Thomas' lessons were error-free except for occasional imprecise uses of language (e.g. stating that a radius went from the middle of the circle to *an* edge rather than the edge). However, these imprecisions did not hinder the mathematics of the lesson. His lessons were characterized by a high level of engagement from the students – students were not only eager to participate but also routinely willing to engage with open-ended mathematical questions and tasks, present their solutions, and correct their errors.

Thomas frequently focused on definitions and underlying mathematical ideas in his lessons, often asking students to set aside procedures and formulas in favor of spending time understanding mathematical structure. For example, Thomas began a lesson on finding the area of a circle by telling students, "Our essential question is: How do I calculate the area of a circle … Before we can worry about formulas and begin to start doing calculations, we need to make sure we understand and know the parts of a circle." He then spent a significant portion of the lesson developing precise definitions of radius and diameter and told students that they would later use these to derive the formula for the circumference of the circle. In another lesson on writing algebraic equations, Thomas told students not to try to solve for the unknown, but rather to focus on writing an equation for a given situation and identifying what the unknown represented. The nature of this work was careful and precise, allowing students the opportunity to understand the components and structure of a formula or equation in a meaningful way.

In contrast to teachers in the 50^{th} and 10^{th} percentiles, students in Thomas' class were the main mathematical actors in the classroom. He routinely encouraged student participation in meaning-making around the mathematical topics, and students did a good deal of mathematical thinking with his occasional prompting. For example, in a lesson on conversion of units of measurement, as part of a warm up the students had previously established that 1 metric cup was equivalent to 250 milliliters. Given this, Thomas asked the students to then determine how many metric cups were equivalent to 1,000 milliliters. Students were unsure and many guessed answers, most of which were incorrect. One student responded that 4 metric cups was equal to 1000 milliliters because 4 times 250 was 1,000. Thomas did not end the interaction once the correct answer was found. He asked the students, "What's another way we could have done it? Because we didn't know it was four." He next facilitated a short discussion about multiple methods to solve the problem as well as the relationship between multiplication and division.

Teacher:	But since we didn't know 4 - and we don't want to guess and check, how could I have used what I knew already to figure out it was 4.
Student:	You could have used 250 divided by 1,000.
Teacher:	250 divided by 1,000? 250 divided by 1,000?
Student:	Oh, 1,000 divided by 250.
Teacher:	1,000 divided by 250 equals?
Student:	Four.

At this point, a student remarked that she had guessed her original answer, which had happened to have been the correct answer, but she didn't know why. Thomas encouraged students to determine a way to know the answer for sure, clarifying that guess and check is an appropriate strategy as long as you do the check in addition to the guessing. He continued:

Teacher:	How would you guess and check what we just did if you were doing 250 times 4. If you said 4 metric cups, how could you guess and check it? Yes?
Student:	By dividing?
Teacher:	You'd multiply. And then go back and divide and see if you get the same thing. Double check yourself. …
Teacher:	How come you can check your multiplication by division?
Student:	*[Inaudible]*. Cause they are…
Teacher:	What's the difference [inaudible]?
Student:	The number is *[Inaudible]*.
Teacher:	What's the difference between multiplication and division? Student M?
Student:	In division it's – in multiplication the problem is four times four equals sixteen. It's just – division is just the problem flipped around. Division is just 16 divided by four.

In this brief exchange, Thomas highlighted three methods for answering the original question (multiplying 250 by 4, dividing 1000 by 250, and guess and check). He also asked questions that elicited a student explanation about the larger relationship between multiplication and division.

Another way in which Thomas encouraged student thinking and meaning-making was through the sequence and nature of the questions and tasks he posed. In a lesson on algebraic equations, he began by giving students a situation containing an unknown and asked them to write the equation that would match it. He gave several more examples, varying the role of the unknown in the situation (the starting quantity, the amount added or subtracted, or the sum/difference). Thomas then reversed the process, giving students an equation for which they had to write a situation. He continued with more examples of this nature, increasing their difficulty, for instance asking the students to write a situation for (5+2) – n = 2 necessitating that the students think about the order of operations in their situations. By planfully varying the content and increasing the difficulty of the problems presented to students, Thomas pressed students to think more deeply about the mathematical content.

In addition to tasks that engaged students in the mathematics, Thomas frequently interacted with students in a way that encouraged them to do the mathematical work of the task. While students worked, Thomas circulated the classroom and asked questions designed to help move students forward, but did not devolve the cognitive demand of the task by intervening directly and telling students how to solve problems. Rather, he engaged students in understanding and correcting their own errors or engaged the class in working with a student error. For example, when students shared their answers in the task in which they were required to write an algebraic equation to represent a situation, Thomas made a point of highlighting the correct parts of incorrect answers and then asked the class to identify and fix any errors. In one segment of instruction, Thomas asked students for a situation that fit the equation 8 – N = 5. Lily responded with the following situation: "Maria has some dolls. She gave eight away. Now she has five. How many did she give away?" This is incorrect because Lily has written a situation that represents the equation N – 8 = 5, mistakenly assigning the number of dolls Maria started with to be the unknown. This is a common student error in writing algebraic equations from situations. Rather than simply correcting Lily, Thomas engaged the class in writing the equation that fit her situation and reasoning why it would be different than the equation given.

Teacher: Okay. What equation should go with *that* situation? Who was the person, was it Maria?

Lily: Yeah.

Teacher: Has some dolls. She gave eight away, now she has five. How many did she start with? What equation is that? I agree with the people who said that's not really eight minus N equals five. But we need to figure out what it *is*.

She has some dolls, Maria did, she gave some – she gave eight away and now she has five. How many did she start with?

Lily: Oh.

Teacher: Lily, what's the equation?

Student: [Inaudible]

Teacher: No, no, no. Tell me what you did; tell me what equation it is.

Student: Oh, I did add five and eight.

Teacher: I'm sorry?

Student: Equals five.

Teacher: Lily did N minus eight equals five. Now, let's see if that – does that make a difference?

Multiple Students: Yes.

Teacher: Because in *this* [points at 8 – N = 5] – don't tell me what number it is. You just know that N in this case has to be what? Think.

Student: A lower number than eight.

Teacher: It has to be lower than eight. But in this case [points at N – 5 = 8], it has to be…?

Multiple Students: Greater than.

Teacher: Or it has to – well, it has to be greater than because she's got five left, okay? So these are going to be two different answers. So now Lily, can you switch it around a little bit so that yours fits eight minus N equals five?

Student: Okay. Maria had eight dolls and she gave some away, and now she has five. How many did she give away?

By both articulating an equation from Lily's incorrect response and reasoning out why 8 – N = 5 is not the same as N – 8 = 5, Lily was able to eventually remedy her own error.

Another feature that characterized Thomas' instruction was a focus on precision in mathematical terms and definitions. For example, in a lesson defining the parts of a circle, he pressed students to create a complete definition for radius. When asked for the definition of a radius, a student responded: "The radius is the center to the other side of the circle." Thomas acknowledged that the student was more or less correct, but pressed for a more precise definition. Another student offered: "It's – radius is like the – is like in the center, but it's not yet half – it's not yet – it's like half of the circle." Thomas replied, "Mmm – I know what you're saying. We're gonna try to fix a word in a little bit." A third student offered, "The radius is how big the circle is, like how wide it is?" Thomas acknowledged that this was descriptive, but not the definition, and told students that they were missing a critical piece. He pointed to the radius and said:

Teacher : But what is the radius? It's something. It's a geometric figure. It is something. It is a? Student T?

Student: A line?

Teacher: A *line*?

Student: Segment?

Teacher: Very good.

Student: A line segment.

Thomas pushed students for a complete, precise definition of radius, not settling for what students were simply able to describe. He remarked, "That's the official definition. So a lot of us were describing it and yes, your descriptions were really accurate, so we want to make sure we can then define it." Later, in defining diameter, he returned to his definition of radius to encourage students to be complete as well as to build upon their earlier work:

Teacher: Can someone give me a definition of a diameter? Give me a full, complete definition. Try to put the words together. All right, Student K.

Student: Like it's half of a circle?

Teacher: Give me a full definition. Look at your definition for radius that we have. Try to formulate something that might relate to it but is more appropriate for the diameter. There might need to be some changes made. Some parts might be able to stay the same.

Student: A line segment that travels from the top to the bottom or from the bottom to the top?

Teacher: I like the "line segment" and I like "that travels." Does it always have to go – you can look at ours. Does it always have to go from the top to the bottom?

Multiple Students: No. [*Inaudible*].

Teacher: From one edge to another edge. There's something else it has to do. …

Student: It has to go – it has to come from the center, it has to pass the center.

Teacher: That it goes from one edge to the other edge and passes through the center. Very good. A diameter must pass through the center. Okay? And it must touch opposite edges. It does not have to be horizontal. It can be vertical, it can be diagonal, it can be shifted

> to the side diagonal. As long as it goes from one end to the other and passes through the center, it is a diameter.

In another lesson, he emphasized the difference between equations and expressions: "When we go to equations there's going to be a major difference. Anyone think they remember what the major difference is going to be when we go from – move from expressions to equations? What do you think the major difference is going to be?" A student replied, "You can use an equal sign with those?" Thomas' reply to the student is: "Okay. Not just you *can* use an equal sign within the equation; you *have* to use an equal sign." Again, he is here emphasizing a key distinction in the definition that students often miss. At one point, he explicitly told a student who had given an incomplete answer, "We're refining. We don't want to say 'wrong.' No, four is right, but there's just more. We're refining."

Thomas' instruction was indicative of many of the teachers near the 90th percentile. Instruction at this level frequently included student engagement with the mathematical content alongside elements of Richness such as comparing across multiple procedures or solution methods, explanations and/or strong mathematical language. Teachers frequently pressed students - as did Thomas - to be precise in their use of mathematics vocabulary and to make sense of the concepts presented. In Thomas' lessons, as well as in many of the other lessons in this category, students' voices were present and students were active participants in their mathematical learning.

Teachers at the 50^{th} percentile

A scan of lesson summaries from teachers near the median revealed two general profiles. In one, a teacher's lesson featured few positive or negative elements – a problem or activity was presented by the teacher, with students contributing an occasional word or phrase. While the teacher did not make mathematical errors, lessons had only occasional richness elements, and these were enacted in a pro forma manner, with no special features that would lead to raters categorizing them as strong (e.g. having multiple solution methods for a given problem but not comparing them explicitly). Tasks were not enacted in a way that challenged students to think deeply about the mathematics, nor did teachers elicit significant student explanations or mathematical reasoning during instruction. Interestingly, many of the tasks presented in these lessons were not procedural in nature. For instance, one teacher provided students 25 counters and demonstrated that when dividing that number by 2, 3 and 4 there would be a remainder; however, the motivation behind this task, and its connection to larger mathematical ideas, was not explored.

The second profile combined strong and weak lesson features. In many summaries, raters noted one or two richness elements – meaning-making around a specific problem, occasionally developed from real-life tasks or the use of concrete representations; multiple procedures for solving these problems; and precise mathematical language. Less common among these lesson summaries were notes describing extensive student participation in building mathematical ideas. Notably, observers also noted mathematical errors that belied a lack of specific mathematical knowledge, such as whether one can model decimal operations with manipulatives (one can, despite one teacher's claim) and whether trapezoids are "irregular" and therefore do not have a formula for their area (as another teacher claimed). However, these errors tended not to get in the way of the main content conveyed to the student (e.g., trapezoids were not the major focus of instruction).

Robert

From these summaries we selected one fourth-grade teacher, Robert, to illustrate the second profile. Robert taught in a small urban district, one marked by high poverty and a large immigrant population. In Robert's school, half the students received free or reduced-price lunch, and in each year of data collection, a handful of students were categorized as English Language Learners in his classroom. The district had long had a progressive orientation to mathematics, selecting TERC's *Investigations* as the primary curriculum materials and providing generous professional development over the years in both mathematics and inquiry-oriented pedagogical strategies. At the time data collection began, Robert had 14 years of experience teaching, and had come out of a traditional teacher education program.

Robert's lessons were marked by several features. First, the tasks were grade-level appropriate (placing fractions on a number line, finding fractional parts of a group), but were largely teacher-directed and also not terribly efficiently enacted, with extensive amounts of time spent on problems that did not warrant that investment. Second, Robert proceduralized many of the unusual and interesting tasks he presented students by breaking them down into step-by-step instructions and doing much of the mathematical work himself. Third, some of his lessons were marked by substantial mathematical sloppiness.

An excerpt from a lesson on division illustrates these issues. With the comment "Some of you are still having trouble understanding that in order to divide you have to be able to multiply," Robert placed a multiplication chart (showing the products of all numbers to 12 x 12) on the board. He began the lesson by asking students how they would divide 17 by 2, then outlined his preferred procedure: inspecting the twos row on the multiplication chart (e.g. 2x1, 2x2, 2x3, etc.) and finding the largest product that could fit into the dividend (in this case 2x8):

Teacher:	If we have to do 17 divided by 2, what we can do is we can use our multiplication table to help us. Our goal is to get as close to 17 without going over. Very good. Okay, and we know it has to be 16. But how many times can 2 go into 17 without going over?
Student:	8.

In this first moment of instruction, the meaning of division – breaking a quantity into equal-sized groups – is not broached. Instead, an answer is found via a procedure – going down the multiples of two until 17 has been reached but not passed. Robert continued by attempting to add some meaning to the situation:

Teacher:	8. So let's talk about one of the division strategies we were doing the other day. So we have 17 divided by 2. So I want everybody to put it on their paper. Then I want you to do …So we're gonna practice it two ways. So we have 17 pencils. We're gonna break them up into two groups, because we're gonna give the pencils to two kids. Now looking at your multiplication chart, how many pencils? What is the maximum number of pencils I can give to each student?
Student:	9.
Multiple Students:	8.

Students again arrive at the correct answer (8), but with a significant issue: the teacher has not specified equal groups. Though this is often assumed when treating division, in this particular case his phrasing suggested that both students who answered 9 and 8 were correct. Nevertheless, he continued, repeating the procedure on the multiplication chart again:

Teacher:	8. Let's prove it. Let's see. Let's take a look at our hundreds chart. Let's take a look at our hundreds chart. Take a look at your hundreds chart. How many pencils do we have?
Student:	8.
Teacher:	No.
Multiple Students:	17.
Teacher:	We have 17 pencils. We don't have 18 pencils. How can I give 18 if I don't have 18? So we can't go up to 18. How many pencils can I give away total without going over?
Student:	16.
Teacher:	I can give away 16 pencils. How many can each child get?
Student:	8.

By "proving it," Robert seems to mean returning to the original procedure.

The exchange over this problem continued until it reached a total length of eight minutes of instructional time. Throughout, Robert's attempts to make sense of the problem were poorly stated and appeared to cause student confusion. In this exchange with students, for instance, Robert attempted to restate the question:

Teacher:	If we're doing 17 divided by 2, what we're trying to do is figure out how many one person can get. So how many can one person get, Student J?

Because Robert emphasized the amount "one person" gets, the student targeted by this question appeared confused:

Student:	Um.
Teacher:	How many can one person get?
Student:	8.
Teacher:	8. And how many pencils are left over?
Student:	8.
Teacher:	No. How many pencils are left over?

At this point, students talked for a moment; many appeared confused about whether he was asking about how many are left after one student receives his share, or how many are left after the two students receive their share. Robert restated the question:

Teacher:	Look. How many pencils are left over?
Student:	One.
Teacher:	One. You know what? 17 minus 16, dear. You have to think about what you're doing and what makes sense. So the answer is 8 remainder 1. Okay.

Thus Robert's efforts to attach meaning to this division problem – a feature that mathematics reformers would credit as an asset to his teaching – is not enacted cleanly. He never explicitly stated the connection between his illustration and the mathematical principle (dividing a set into two equal groups), omitted a key condition (equal groups) throughout his discussion, and in both this and the additional examples he provided to students, proceduralized the connection between multiplication and division using a set of steps students must apply to the hundreds chart. As well, despite saying at the beginning of this passage that the students were "gonna practice it two ways," he never returned to the other model of division (eight groups of two).

This lesson excerpt also demonstrates another salient feature of Robert's lessons: he does most of the talking. As seen above, students contributed answers to computational questions ("Eight. Nine.") but little else. Throughout his lessons, there were occasional challenges to students to think (e.g., "Why don't you explain it like it was money. So what would 5/10 be equal to?") and attendant student meaning-making or reasoning, but these examples were relatively infrequent. Instead, Robert's instructional time was spent walking students through solutions to the tasks he presents.

Some of Robert's instruction across lessons did feature attention to meaning and other elements of mathematical richness but these moments were infrequent and did not characterize his instruction. For example, in one lesson on placing fractions on the number line, Robert was careful to stipulate that the denominator specifies how many equal-sized portions to make, and the numerator counts off those equal-sized portions. He also noted that students should use benchmark fractions to make sense of unusual ones – for instance, 7/10ths is a bit larger than 5/10ths.

What we observed in Robert's lessons was common among other teachers at or near the 50^{th} percentile in our data. Seldom did we see teachers present straightforward, traditional lessons, in which an idea or procedure is delivered and students then practice deploying that new knowledge. This is quite different than depictions of instruction in past work. Instead, our study found that teachers worked with more interesting tasks – and we saw some corresponding instances of mathematical richness or student participation in the lesson. But in many cases, these aspects of the lesson occurred only sporadically. Rather than part of a coherent, well-sequenced series of events designed to invite student participation in meaning-making and reasoning, teachers might ask an occasional "why" question. Rather than carefully building connections between and among mathematical representations and ideas, and providing a sense of meaning behind mathematical procedures, these elements would pop up briefly or in isolation. In other cases, teachers failed to realize the promise of potentially meaningful tasks, neglecting to leverage these tasks to successfully develop strong disciplinary insight or encourage students into a more active role in the development of their own mathematical understanding. In addition, lessons in this category held numerous mathematical imprecisions; while most of these errors did not degrade the quality of the mathematics, they occasionally caused student confusion.

Teachers in the 10^{th} percentile

Similar to teachers at the 50^{th} percentile, our scan of summaries describing lessons from teachers at or near the 10^{th} percentile also revealed some common features. First, students generally listened to or watched their teachers demonstrate how to solve problems without much input, similar to the way Robert conveyed the 17 divided by 2 material. However, unlike Robert's lesson and the lessons of 50^{th} percentile teachers generally, the mathematics in these classrooms had few to no elements of richness described in Table 2. Teachers also rarely sought to engage students in rigorous mathematical thinking, even for the short amounts of time seen in 50^{th}-percentile teachers. Third, teachers in the 10^{th} percentile frequently made mathematical errors that obscured important mathematical features of the lesson.

In addition to the shared features of teachers at the 10th percentile, two distinct lesson profiles further displayed additional concerning patterns in instruction. For one group of teachers, the pacing of the material was extremely slow; teachers worked through only a few problems over long periods of time without doing so in mathematically rich ways, and behavioral and/or administrative issues took time away from instruction. In some cases, raters judged the activities contained in lessons to be non-mathematical (e.g. spending a large portion of a lesson making an art piece using 4 colors on a 10x10 grid with no discernable mathematical connection). For the other group of teachers, lessons contained occasional moderate to serious mathematical errors, or activities that led to mathematical dead ends.

Arlene

To illustrate these themes, we selected one teacher, Arlene, who displayed features of both profiles and of this type of instruction on the whole. Arlene worked in the same district as Thomas, in a school where about one-third of the student population receives free or reduced price lunch. Few students in either study year were classified as English Language Learners. By the first year of our study, Arlene had taught for 13 years, with a traditional certification and a relatively strong mathematics background, reporting between three and five mathematics courses during her college and post-college years. She also reported serving on district math committees, as a peer mentor or coach in the area of mathematics, and teaching in-service courses in the area of mathematics. In the first year of the study, she reported drawing heavily upon state-supplied curriculum materials; in the second, she reported more diverse set of supports, including Everyday Mathematics, Investigations, Harcourt Mathematics (the district-adopted text) as well as other supplemental materials.

Arlene's lessons were remarkably consistent both within and across years. As was the recommended practice in her district, she divided students into small groups for instruction, presenting the lesson to a group of four to five individuals while other students worked on other tasks in centers throughout the room. These students served as a frequent source of distraction for Arlene, either by interrupting (to ask mathematical or clarifying questions, to note that materials were not available, to ask permission to the leave the classroom), or because they were off-task. Typically, Arlene repeated the exact same mini-lesson to each small group, essentially teaching the material several times over within a 50 to 60-minute time span. With one exception, student participation in these short lessons was mostly confined to taking notes (students kept math journals for definitions and data displays) and helping the teacher execute procedures.

One striking aspect of Arlene's lessons was the relative thinness of the mathematical content. In a lesson on graphs, for instance, she called three successive groups to her desk to go over two ideas about the Y axis. She began the lesson with a question:

Teacher: What information is on the vertical axis of a bar graph and a line graph?

Students spent almost three minutes copying this question into their math journal. After it appeared that all students had finished, she continued:

Teacher: So show me with your hands what a vertical line would look like.

Students held up their arms akimbo; seeing no consensus, the teacher continued:

Teacher: My vertical line is like my vertical blinds. They go up and down. So let's put a little code there for us so we know, vertical blinds. I told you to get something else, to do one of the other activities. So here's my window.

Arlene and the students then spent another three minutes drawing a window with vertical blinds in their journal, followed by a horizon (to help remember horizontal lines). Several interruptions and a review of different kinds of graphs later, she drew students' attention to the Y axis:

Teacher: What do you notice about the numbers? Are they going big to small or small to large?

Multiple Students: Small to large.

Teacher: Small to large. So they're counting up. Right?

Multiple Students: Yes.

Teacher: And they're counting up by fives. So this is called my scale, S-C-A-L-E.

Students copied this information into their math journals; after a time, she asked and answered her own question: "Do you know what they call it, what it counts by? That is my interval. My interval on my scale is what they are counting by." Students examined scales that had intervals of five and twenty, then drew a graph with an interval of three. The group then dispersed and Arlene began the process anew with a new set of students.

Mathematically, this instruction appeared quite superficial. There is no connection between the interval and the actual bar or line graphs, no discussion of how the interval influences how you read a graph, and no reason why an interval would not be 1 for all graphs. In fact, the content of the mini-lesson is just the vocabulary "scale" and "interval" and there is no evidence that the other tasks students complete once leaving Arlene's center deepen the mathematics. This feature was characteristic across Arlene's lessons. Also consistent with her other lessons, Arlene presented all the information to students, asking students only the most obvious question ("small to large. So they're counting up. Right?"). Throughout, there were numerous disruptions from other students seeking information or needing supervision. Finally, Arlene took no advantage of the small-group format to customize instruction or allow students more generous opportunities to participate in the lesson.

The exception to this overall pattern was a lesson that used a representation of a balance scale (a picture on the wall of a balance) and "blocks" to represent different weights. In this lesson, instruction was characterized both by more student involvement in problem-solving and an emphasis on students providing explanations, but also significant mathematical imprecision on the part of the teacher. In her introduction to the material with the first small group, for instance, Arlene launched the activity as follows:

> Okay, we are going to be doing something with a balance today and we're still talking about algebra. So far we've gone over what variables are, right? So we know what the numbers are and letters are to represent each other, and we also had been talking about growing patterns, like we did yesterday with the iguana. So today is going to be something a little bit different where you're still using variables, which is a symbol or letter to match the number, but we're going to try to balance something out.

This passage is highly imprecise – letters are said represent each other, rather than an unknown number. In fact, the unknowns in this lesson represent constant values. And the reference to "balance something out" is not specific enough to convey the actual activity or why the activity would be of mathematical importance. The insertion of the growing pattern reference isn't relevant to this lesson at all, and further clouds the problems with her launch.

Her next set of instructions did little to elucidate the task:

> A mobile has a total weight of 30, circles weigh 5, squares weigh 10, and triangles weigh 3. So the whole thing itself is going to be level at 30. … [The mobile] only has two shapes. If only circles are on one side, can it be balanced?

Watching the subsequent minutes of instruction, it became clear that students are not sure whether all shapes should total 30, or the shapes on each side should total 30. The teacher also partially devolves the cognitive demand of the task through heavy hinting:

> My circles weigh 5. Is there any way that I can use circles on one side – and I made a bunch of circles but this may not necessarily be correct [*Arlene puts up post-it notes representing circles of weight five on one side of the balance*]– is there any way I can put circles on one side and balance it out with my square or my triangle?

Within these constraints, however, students actually do a bit of mathematical thinking, noting that three five-unit circles (15) could be balanced by five three-unit triangles. Students solve several problems of this variety, and then are dismissed back to their seats.

This lesson is intended to motivate the idea of balance in an algebraic equation – the notion that the value on one side of the equal sign is equivalent to the value on the other and that you can find an unknown value by maintaining balance. However, in its enactment this mathematical point is lost. Arlene directed students to pay more attention to the shapes themselves than their weight (for example, at one point she emphasizes "If only circles are on one side can it be balanced?" and then added to the confusion when she repeated this condition differently, "If circles are only on one side can it be balanced?"). She seemed to lose sight of the purpose of the use of the shapes, hinting at the relationship between the weight of the shapes and their role as coefficients when she tells students that 4 written next to circle is the same as 4 times circle. However, she neglects to emphasize the weight of the circle in that context.

Thus Arlene unifies both threads within our analysis of 10th-percentile videos: little student participation or richness, thin mathematics, frequent distractions. When more substantial mathematical topics are addressed, imprecisions and lack of clarity obscure the mathematical substance.

District differences

We next turn to whether the five districts in this study differ in their instructional profile. If so, it would suggest searching for explanations for those differences. Boxplots in Figure 1 and averages in Table 6 show that instruction does appear to differ substantially by district. For richness, District 11 has both a higher average and more variability across teachers; for Common Core Aligned Student Practices, District 11 likewise has higher scores and more variability across teachers. For errors, District 13 has a higher average and more variability across teachers. Table 7 shows that many of these differences between districts are statistically significant, especially those between District 11 and the rest of the sample.

In light of these differences, the next section unpacks the structures that were in place in Districts 11 and 14 in the years during and leading up to the study that could have influenced the quality of elementary mathematics instruction.

District 11

Description and Introduction

Located on the Eastern Seaboard of the United States, District 11 is a mid-sized urban district enrolling 57,000 students and employing 4,000 teachers in100 schools. 65% of children are eligible for subsidized lunches, and the majority of students are identified as belonging to a minority (75% Black or Hispanic). 12% of students are designated as having special needs and 30% of students are English language learners.

As discussed, our study revealed that the average quality of mathematics instruction from teachers in District 11 exceeded that of other districts by more than a standard deviation. Anecdotally, study researchers also found that lessons from District 11 teachers were most likely to provide examples of ambitious instruction in line with Common Core ideals. This disparity is particularly interesting given that District 11 had been plagued by underperforming schools for many years. Interviews with district personnel, however, revealed other, structural factors in the years leading up to and during our study that may have influenced the quality of elementary math instruction in District 11. We learned that our study captured elementary math instruction in District 11 at a time in which those individuals in the district who best understand mathematics instruction, curricula, and teachers were empowered to carry out a systemic, coherent, and long-term plan to promote conceptual and standards-based reforms in mathematics instruction, which was referred to the District 11 math office as the "Math Plan."

New Leadership, New Plan

Plagued by chronically underperforming schools and inner turmoil, District 11 hired a superintendent in the mid-1990s who was backed by the governing board, the teachers' union, the business community, and the mayor. This unusually broad showing of support across stakeholders positioned the superintendent to implement initiatives geared toward systemic reform of this struggling mid-sized urban district. His background in education and previous experience with urban district reform made him a strong believer in instructional improvement as the most powerful lever to improve student learning. He was therefore supportive of the Math Plan proposed by the math office, even defending it early on when it initially by school principals. As a believer in the importance of instruction to improve learning, the superintendent also provided the office of curriculum and instruction—and each academic team—relative autonomy in the design and implementation of its work, thereby leaving those individuals most familiar with the content area to design curriculum and professional development.

This focused and consistent leadership enjoyed by District 11 has extended down to the math office. The current math coordinator was hired in 2000 and brought with her a wealth of experience as a middle school and secondary mathematics teacher in an urban district, and later as a professor of mathematics education. Over the past 15 years she has been a steadfast leader and champion of the Math Plan, and also supervised a team of mathematics curriculum and instruction specialists and coaches, who together contributed to the design and execution of the Math Plan in schools and at the central office.

Dedicated Funding

Knowing that achieving fundamental and district-wide reforms to math instruction would require adequate support and resources, the math director worked with the district leadership to write a proposal for a sizable National Science Foundation (NSF) grant aimed at improving mathematics instruction through the adoption of a standards-based curriculum and the development of school-based capacity and instructional leadership. This dedicated line of funding enabled the math office to initiate its planned work without competing with other offices for resources, and thus was integral to its success.

The Math Plan

When we asked the math director about the "package" that the district provided its teachers to improve instruction, the math coordinator explained that the Math Plan was based upon three key elements: the adoption of high-quality, standards-based curricula, a portfolio of opportunities for teachers to engage in professional development, and the development of school-based instructional support for teachers and principals. These efforts were envisioned in concert with one another, and together were as "the cohesive tool for the work."

Adoption of a standards-based curriculum.

Curriculum materials in District 11 were traditionally adopted by individual schools after choosing between three district-recommended curricula. The math director, wanting to avoid the complications and disjointedness that this policy had historically fostered, however, worked with district leadership to choose a single curriculum, *Investigations Through Time and Space*, a reform-oriented curriculum developed with NSF funding, which already had a strong reputation among mathematics educators and researchers for its focus on deep conceptual understanding of mathematics. *Investigations* was also aligned to the new state math test and later, the Common Core State Standards. The math director explained, "We chose this text because it is the best standards-based option. When something comes out that is said to be better, we look at it carefully. But I have yet to see anything that's better."

Moreover, the math coordinator shared that adoption of a single mathematics curriculum district-wide sent a message about the superintendent's support of the Math Plan, and also about the math office's goal of cohesion among schools in the district. While there was some initial resistance to the new curriculum during its phased roll-out over two years, the director shared that the materials began to "speak for itself."

Professional development.

A second component of the Math Plan was a portfolio of high quality professional development opportunities for teachers. Among the offerings were:

- Curriculum institutes (30 hours) and seminars (up to 10 hours) designed to support teachers in full use of the curriculum as well as the professional development resources embedded in the curriculum itself.
- Targeted seminars (24 hours per seminar), developed by and with a set of highly trained teachers, provided deeper focus on selected topics designed to deepen teachers' math content knowledge for teaching.
- School-based workshops that focused on pedagogical issues such as purposeful planning, facilitating classroom discussions, and looking at student work.
- Specially designed courses (up to 30 hours over the year) for principals and administrators designed to promote a deeper understanding of mathematics teaching and learning and to help leaders support this vision in their particular schools.

The mathematics director shared that it was important that the professional development options were aligned and mutually reinforced one another, and were designed with this cohesion in mind. She shared,

> We generally use the same model for all the PD. We engage teachers in some mathematical work, and then attend to students' mathematical thinking around that same work. In almost any PD session we offer, we look at teachers' mathematics but also students' mathematics, and implications for practice. We then ask [teachers], 'with your curriculum materials, how do you address how students are thinking, and what they need to learn?'

When asked about the reach that these opportunities had across the district, the director indicated that each year at least on teacher per grade level, per school participated in seminars, with the intent of bringing learning back to the building which would then be reinforced by a math coach assigned to the school.

Developing school-based instructional support
The third component of the Math Plan was the deployment of a dedicated cadre of math coaches, each with a caseload of schools to provide individualized instructional support. These coaches also worked at the math office designing and facilitating professional development, thereby creating consistency between the professional learning provided by the math office and the support that teachers received in their schools. In this way, math coaches were essential to keeping the professional development offered by the math office "grounded in … the real needs of schools." The math director explained,

> "coaches take the substance of what is being explored in professional development and then work very specifically with teachers. Teachers can learn a lot during professional development, but what really helps is having a coach who then helps you translate that into what this going to look like in your classroom."

In the early 2000s, there were 12 – 15 dedicated math coaches who were key to creating a "feedback loop" between schools and the math office. The math coordinator also recalled, however, that this streamlined communication and coherence encouraged by math coaches were threatened when external funding ended, and due to funding constraints, the number of coaches were greatly reduced or they were supervised at the school level by principals. The math director shared that, during this time

> "there wasn't much support for real strengthening of classroom practice. You can imagine that a principal trying to run a building might have different priorities than the math office might--covering a class if a sub doesn't show up,

> all sorts of things. Also, it's also hard to push on teachers if they are your colleagues and you're all being supervised and evaluated by the same administrator. So one thing we learned was that it's more effective to have coaches who are supervised by someone who works outside of the building—they can have a more focused, coherent, plan of action that lets coaches do their work."

Another important component to the development of school-based instructional support was the Math Plan's effort to build math leadership capacity in every building by developing "math teacher leaders," highly trained full-time teachers who also worked as math facilitators for the teachers in their schools. Like the district-supervised math coaches, teacher leaders offered both instructional support and common messaging from the math office.

Alignment With the State

Because the initial work of the Math Plan coincided with a time of new test-based accountability measures from the No Child Left Behind Act, we were interested in learning if and how emphasis on the state test either helped or hindered the work of the math office. The math coordinator indicated that she was fortunate that the state math test did not conflict with the work of the math office, and in fact, "doesn't include anything that we would say isn't important". In fact, she shared that part of the argument the math office made in 2000 when proposing the Math Plan was that "it would help students perform better on the test." Indeed, she shared that sample items from the state test would be brought into professional development workshops and seminars to stimulate "good discussion about what the mathematics is that the state test draws upon." Teachers were even encouraged to use sample items with their students as a means to understand how students thought about the mathematics presented on the test.

Additionally, the math coordinator shared that over the years, the math office has worked closely with the state to learn how it understood mathematics, and has even had opportunities to influence the content of state test through participation in item development committees and statewide meetings of district math coordinators. She shared that District 11 students have seen good gains on the state test over the years, and noted that they have also done so on the National Assessment of Educational Progress (NAEP) as compared to other urban districts.

Impediments to the Math Plan

When we shared the results from our study with the math director, she acknowledged that District 11 teachers demonstrated higher overall quality of mathematics instruction but also noted that our sample from District 11 included some teachers whose math instruction were in the lower ranges. This prompted our question of whether she viewed anything over the years as having been an obstruction to carrying out the work of the Math Plan. Again, she noted that the decrease in external funding has led to a reduction in the math office's primary human capital resource, math coaches, making the execution of the work in schools more challenging. She also stated,

> "if you look at the span of your study, 2010 – 13, not only do you come in at a stage where there's been this long-term plan that has been able to be carried out, we had this era of funding where we could target teacher professional development, and then, at least in the time period of 2010 – 13, we enjoyed the organization structure within the district such that the elementary side of the math office had fairly consistent face-time with schools..."

Another member of the math office offered that something that could be seen as an impediment to the work, "for better or for worse" was the implementation of the new teacher evaluation system in which "everyone's running off and doing their own thing and everyone's sort of grasping in a dark room trying to figure out what [they're] all supposed to be doing." She continued, "We're quite concerned—the entire academic office, not just the math office. What happens moving forward?"

Finally, the team worried about conflicting messages from the academic and evaluation offices alienating teachers, who might read the message as "be a learner, but be perfect." The math team expressed concern that potentially competing goals and messaging between offices may serve to unravel the cohesion it has worked long and hard to create for District 11 teachers.

District 14

Description and Introduction

As one of the fastest growing districts in the U.S., District 14 has well over 165,000 students. A large, urban district in the South, District 14 experienced a dramatic population shift over the last two decades as a surplus of jobs and affordable housing attracted young families to the area. This rapid growth resulted in a significant demographic shift in the overall county population as the numbers of Whites dropped from 90.9 % to 53.34 % between 1990-2010. According to the U.S. Census (2014), during this same time period, African Americans increased from 5.2 to 23.6 %, Asians from 2.9 to 10.59%, and Hispanics from 2.4 to 20.12%. By 2012, 33% of households in the county spoke a language other than English and 29% were single parent families. The district was challenged on every front with the need to build and staff additional schools, develop curricular supports for increasing numbers of English as a Second Language (ESL) and socioeconomically disadvantaged students, and to provide appropriate professional development opportunities for teachers and district leadership.

The descriptive account that follows is a brief overview of District 14's comprehensive plan for school improvement. As noted earlier, District 14's active response to these concerns resulted in remarkably consistent instruction across participating teachers' elementary mathematics classrooms. However, we also were curious to understand why after more than a decade since the plan was implemented; the district's systemic approach to professional learning along with increased accountability has not significantly improved the quality of mathematics instruction.

A New Academic Standard

By the late 1990's the district had established a positive state-wide reputation for preparing students from traditionally White, affluent neighborhoods to move on to higher education and the workplace. A benefit of heighted accountability under No Child Left Behind (NCLB), increased visibility to student level data, highlighted a gap between the districts' top students and the growing numbers of students who needed more intense instructional supports than were available under the existing system. However the superintendent, district administration, school board, and local residents were at odds for how to solve this problem. It was only after several years of contentious public debates, that the greater school community agreed there needed to be a clear set of accountability measures that were aligned to the curriculum, and which would serve as gateway assessments for students moving through the system.

The district initially turned to the state for help in revamping the district standards and curriculum, however, the development team soon determined that the state-suggested curriculum was unwieldy and not well aligned to the state assessment. Under the eventual guidance of a new superintendent, a team made up of teachers, parents and community members created its own set of benchmark assessments and curriculum standards that would be implemented in every school and reviewed annually. As part of a comprehensive improvement plan, the benchmark assessments were to be administered every six weeks to help building principals and teachers track student achievement. Parents and students were given a standards checklist that would allow them to track where students were in relation to the instructional calendar. Students who failed to meet the standards were referred to summer school and allowed to retake the state assessment. The district's stated goal was to put an end to social promotion and demystify the instructional process, to ensure systemic coherence and make learning visible to all stakeholders; and to hold every district administrator, principal and teacher in the district accountable for student achievement.

Supports for Improved Mathematics Instruction

Traditionally, District 14 has invested heavily in professional development for both teachers and building administrators. Throughout the 1990's, district staff enjoyed easy access to emerging research in mathematics, well trained mathematics coaches, and professional learning opportunities such as the Summer Leadership Conference, the Summer Math Institute and were encouraged to attend outside trainings hosted by leading researchers in the field. However, as the full effects of the faltering U.S. economy were felt in District 14, the district administration was forced to make deep budget cuts and slowly began to pull back essential mathematics (and other) supports and resources. By 2010, principals were looking closely at their building budgets and needed to consider carefully which services they could afford to keep. The following describes how the district adapted the suite of mathematics resources and supports to fit the new budget constraints.

District Math Office

The math office team is under the direction of the executive director of curriculum and instruction and is charged with developing a coherent plan for district-wide math instruction. Perhaps the biggest challenge facing the team has been the fact that over the past four years there have been three different district mathematics directors and three different perspectives on what math instruction should look like in the district. Although each of the directors has brought significant administrative experience to the position, none of the three have a strong background in mathematics. Two of the three directors have since moved on to become building principals. In addition to the director, two district coaches (one of the coaches works half time) provide support for over 70 elementary schools.

We wondered why there have been so many turnovers in the district math office. In part, this may be due to the fact that past directors lack the necessary mathematics background to engage in the work deeply. Further, with few staff and a minimal budget, the math team has limited capacity to influence and support district-wide change. We also speculate that mathematics has not been an instructional priority for the district executive team as evidenced by the lack of district support staff assigned to the math program.

Building-Level Math Coaches

When the new standards were initially rolled out, a cadre of math coaches helped support teacher learning. In a systematic effort to raise instructional standards, mathematics coaches were trained to go into buildings and work one-on-one with teachers. The coaches provided curriculum materials, instructional strategies and even modeled classroom lessons. However, when faced with the need to reduce building budgets, including the number of full-time employees (FTEs), some principals chose to give up their math coach. When asked, the district math director noted,

> We had two instructional coaches at the district level that were designated primarily to the elementary schools. Each of the local schools had the wherewithal to determine whether they would allocate some of their staff points for math coaches, and I would say the vast majority of our elementary schools had a math coach [during the years of the study].

However, another district staff member who worked closely at the building level noted that principals who did not view themselves as instructional leaders in mathematics –as “math principals”—sometimes failed to see the value in having a math coach in the building and opted to focus on reading or other, more comfortable instructional areas. And even when the math coach was retained, principal-supervisors sometimes assigned the coach “other duties” not related to mathematics instruction.

Curriculum Adoption

As noted earlier, the choice of a reform-based (e.g. NSF-funded curriculum) can help support more ambitious mathematics instruction. Thus, the process for district curriculum adoption matters. In District 14, a mathematics curriculum adoption occurs every six years. As part of the adoption process, the district gathers a list of recommended texts, narrows the selection to three, which are then piloted in district classrooms. After the pilot, district teachers vote on the curriculum-of-choice. This process has tended to favor more traditional texts that are: familiar to most district mathematics teachers, aligned to the district’s standards, and which encourage more traditional –often more comfortable for teachers-- modes of instruction. McGraw Hill was selected during the most recent adoption and Harcourt Trade was adopted previously through a similar process.

It is important to note that the recent decision to adopt a traditional curriculum was not undisputed nor had the district math team missed the fact that they would soon be shifting over to the Common Core standards and would benefit from a reform curriculum. The district math director commented,

> “We knew that the Common Core aligned standards were on the horizon, so our first goal was to begin preparing our teachers for, again, not necessarily more rigorous standards, but certainly a more conceptual approach and deeper understanding of the content standards. And so our professional development was pretty much geared towards that as well as trying to provide resources that would also promote that type of instruction.”

Some teachers and coaches also remarked they would have preferred a reform-based curriculum such as *Investigations* or *Everyday Math* that supports a conceptual teaching approach advocated by the Common Core. However, a district coach remarked that the majority of the committee was made up of a group of very traditional teachers who were worried that the district did not have sufficient resources to train district teachers to use a reform-based curriculum that would require them to learn a significantly different approach to teaching mathematics. The committee included too few independent thinkers with strong math knowledge –“the movers and shakers”-- to influence the choice of curriculum. And while the math director could have influenced the decision, he was “brand new and didn’t want to make waves.”

In the end, the deciding factor for the current adoption was the fact that McGraw Hill includes an extensive digital platform with online resources. A district math coach and strong advocate for reform curriculum was frustrated with the choice, “Are we really choosing delivery model over content?” While the district math team acknowledged it was the district’s responsibility to lead the move towards reform mathematics, they were not well positioned to do so. They simply did not have the resources to support that effort. Nor were they willing to take a political risk and overrule the vote cast by a majority of teachers.

The math director was quick to point out that the district does provide curricular support for the Common Core rollout in the form of a digital platform referred to as the MOCC or the Math Online Communications Center. This online platform allows teachers to upload and share “best lessons.” (Although it is not clear how those lessons are to be vetted or implemented into practice.) The director further pointed out that “we also had think maps, we had exemplars that the district paid for to be developed --Super Source-- we also had hands-on standards which promoted this notion of a conceptual approach to math instruction.”

Professional Learning

For the last decade, the primary source of professional development for the district’s elementary math teachers has been the Summer Math Institute. The brainchild of a former math director and one of the current math coaches, the Institute has the strong support of the superintendent –who continues to pick up the annual cost of the summer training from his discretionary budget. The professional learning activities for the institute were closely aligned to the district standards and end-of-unit assessments as well as to a number of initiatives being promoted through the Curriculum and Instruction office (e.g. Quality Plus teaching strategies).

The Summer Math Institute applies a train-the-trainer model to develop teachers’ understanding of conceptual teaching and has trained as many as 700 teachers during a single summer. During the institute, “Master” teachers, often coaches or highly skilled teachers from across the district, deliver instruction to a group of struggling summer school students who are preparing to retake state assessment.. A second group of teachers (Cluster Trainers) observe the instruction and also have opportunities to engage with students. The theory behind this “fishbowl” approach is that teachers will be able to replicate quality instruction as a result of viewing examples of conceptual teaching and having hands-on experience in an actual classroom. After observing and debriefing the first, four-day institute, Cluster Trainers return to their own building’s summer school and repeat the training with a new group of students and teacher-observers. There also is a general expectation that the Cluster trainers will share this learning once back in their buildings.

During the institute, participants also are given instructions for how to prepare a building-level presentation that they are encouraged to share with their peers in vertical team meetings or other building forums. However, there is not a formal process for helping teachers articulate their learning back in their buildings. A district math coach commented that this dissemination effort sometimes falls flat because it is dependent on whether the principal is focused on improving math instruction --and not reading or another core content area—and will allow building meeting time to be used for engaging teachers around the mathematics.

Professional learning opportunities for mathematics teachers are limited. Budget cuts have clearly impacted what had been at one time, a coherent professional learning plan for principals and teachers. Whereas in the past, district and building administrators were required to attend a multiple day Summer Leadership Conference that helped ensure principals grounded their work around district initiatives. Now, however, principal trainings are held periodically and attendance is optional. In general, there are far fewer professional learning opportunities for principals and teachers and according to a district math coach, available trainings are "loosely coupled" to building-level work.

Sending a Consistent Message

There has been a surfeit of change in District 14 since the mid 1990's. The district was out in front of the rest of the educational community for establishing a rigorous set of instructional standards and then holding district principals and staff accountable for student learning. When asked to attribute a reason for the unusual degree of consistency across the District 14 mathematics classrooms in the study, the math director remarked that the district's emphasis on consistent messaging and professional development helped ensure uniformity.

> Consistent messaging, consistent professional development that's common for all of teachers, and probably a pretty systematic district-wide approach to any professional development, communications resources provided and those sorts things even though we're a large district

The district is confident in its ability to design, implement, and assess rigorous standards. So confident in fact that District 14 has rejected the PARCC assessment as a Common Core measure and plans to develop its own assessment. Based on past experience, there is little doubt that when the assessment is complete, it will be messaged and pushed out to buildings in an extremely consistent manner.

Next Steps

Less clear is how District 14 will improve mathematics instruction to better support students' conceptual understanding of the mathematics –an essential requirement of the Common Core standards. The study findings highlight a concern with the quality of mathematics instruction across the district. Although there was uniformity across classrooms, the overall quality of instruction was lacking. And although the stated goal of the Summer Math Institute –the primary source of professional development for math teachers-- was to help teachers learn to design and teach lessons that will advance students' conceptual learning and mathematical understanding, we saw little evidence that this type of instruction was occurring in District 14's classrooms. The math director pointed to the state test as one possible reason,

> I would say that the [state test] is a minimum standards assessment. It tests –it's a multiple choice assessment. It also tests really the bare minimum procedural level in almost every case, and so you would almost see a reward in a procedural approach to the instruction.

The director argued that teachers may have aligned their instruction to the test because "they were being held accountable for standards other than the ones we were attempting to promote through professional development." Teachers simply lacked the motivation to make significant changes in their practice. He further suggested that when the district writes their own Common Core assessment as a more accurate measure of students' conceptual understanding, teachers may be more highly motivated to adopt the types of instructional strategies promoted through the district professional development.

Two Districts, Two Approaches

The stories of District 11 and 14's work around elementary mathematics in the years during and leading up to our study, along with their respective results, may demonstrate relative successes and challenges faced by two urban districts striving to improve mathematics instruction and learning outcomes. District 14's steadfast adherence to its own curriculum standards and benchmark assessments has all but ensured that its teachers will exhibit a minimum level of instructional quality, but, in our sample, were unlikely to push beyond the traditional instruction that these same standards promoted. District 11 has also long-adhered to a plan of its own, but in contrast to District 14, one that has slowly pushed its teachers toward more ambitious and Common Core-like instructional practices. This distinction is further exemplified in the districts curricula and state tests which encouraged and rewarded a traditional instruction in District 14 and reform-oriented instruction in District 11. District 11's math office also enjoyed strong political backing from the superintendent, dedicated funding for coaching and professional development, and math expertise and unchanging leadership. At present, both districts may be seen to be at a crossroads. For District 14, the adoption of the Common Core may motivate a self-examination of how it can support its teachers to move beyond traditional, albeit consistent, instruction, to ambitious and Common Core aligned practices. In District 11, with its funding sources greatly depleted and facing interruptions to its work from potentially competing initiatives, the math office will need to think resourcefully about how it can continue to support its teachers to achieve the ideals of the Common Core, and avoid losing the ground it has made over the years.

Conclusion

References

Cohen, D. K., & Ball, D. L. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis*, *12*(3), 233-239.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, *12*(3), 311–329.

Cohen, D. K. (2011). *Teaching and its predicaments*. Harvard University Press

Cohen, D. K., & Hill, H. C. (2008). *Learning policy: When state education reform works*. Yale University Press.

Cohen, D. K., McLaughlin, M., & Talbert, J. (1993). Teaching for understanding: Challenges for practice, research and policy.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1890-1990*. Teachers College Press.

Cuban, L. (2013). *Inside the black box of classroom practice: Change without reform in American education*. Harvard Education Press.

Diamond, J. (2007). Where the rubber meets the road: Rethinking the connection between high-stakes testing policy and classroom instruction. *Sociology of Education*, *80*(4), 285–313.

Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, *106*(6), 1145-1176.

Heaton, R. M. (1992). Who is minding the mathematics content? A case study of a fifth-grade teacher. *The Elementary School Journal*, 153-162.

Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education*, 3-19.

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., ... & Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis*, *27*(2), 111-132.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430-511.

Hill, H. C., & Lubienski, S. (2007). Teachers' mathematics knowledge for teaching and school context: A study of California teachers. *Educational Policy*.

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *The Future of Children*, *17*(1), 129-153.

Kane, T. J., & Staiger, D. O. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project. *Bill & Melinda Gates Foundation*.

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

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37-62.

National Council for Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council for Teachers of Mathematics (2000). Principles and Standards for School Mathematics. Reston, VA: Author.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Author.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. US Department of Education.

Smith, T. M., Desimone, L. M., & Ueno, K. (2005). “Highly qualified” to do what? The relationship between NCLB teacher quality mandates and the use of reform-oriented instruction in middle school mathematics. *Educational Evaluation and Policy Analysis*, *27*(1), 75-109.

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

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). Looking inside the classroom: A study of K-12 mathematics and science education in the United States.

Table 1: Student characteristics

				Student Characteristics					
District	N	% African American	% Asian	% Hispanic	% White	% Other Race	% FRPL	% ELL	%
Overall	75745	36.93	8.29	26.71	24.02	3.86	63.81	20.31	
11	14422	35.43	8.57	39.06	11.79	5.15	79.27	32.93	2
12	4127	51.61	3.39	13.30	27.36	4.34	75.24	25.05	
13	9020	72.55	2.24	13.07	11.03	1.11	68.10	6.12	
14	47387	29.08	9.88	27.06	29.96	4.02	57.14	19.09	
15	789	51.71	2.66	6.21	21.67	0.00	73.45	0.76	

Table 2

MQI Dimensions and Items

Richness of the mathematics*:* This dimension captures the depth of the mathematics offered to students. Rich mathematics focus either on the meaning of facts and procedures or on key mathematical practices. The dimension consists of the following items:

- *Linking and Connections*: Linking and connecting mathematical representations, ideas, and procedures*.*
- *Explanations*: Giving mathematical meaning to ideas, procedures, steps, or solution methods.
- *Multiple Procedures or Solution Methods*: Considering multiple solution methods or procedures for a single problem.
- *Developing Generalizations*: Using specific examples to develop generalizations of mathematical facts or procedures.
- *Mathematical Language*: Using dense and precise language fluently and consistently during the lesson.

Student participation in meaning-making and reasoning: This dimension captures evidence of students' involvement in cognitively activating classroom work. Attention here focuses on student participation in activities such as:

- *Providing explanations*.
- *Posing mathematically motivated questions or offering mathematical claims or counterclaims*.
- *Engaging in reasoning and cognitively demanding activities,* such as drawing connections among different representations, concepts, or solution methods; identifying and explaining patterns.

Errors and Imprecision*:* This dimension is intended to capture teacher errors or imprecision of language and notation, uncorrected student errors, or the lack of clarity/precision in the teacher's presentation of the content. This dimension consists of the following items:

- *Major mathematical errors or serious mathematical oversights*, such as solving? problems incorrectly; defining terms incorrectly; forgetting a key condition in a definition; equating two non-identical mathematical terms.
- *Imprecision in language or notation*: Imprecision in use of mathematical symbols (notation), use of technical mathematical language, and use of general language when discussing mathematical ideas.
- *Lack of clarity* in teachers' launching of tasks or presentation of the content.

Table 3: Instructional Format Frequencies

	Score Point		
Code	Active	Small Group	Both
Instructional Format	46	34	20
Modes of Instruction	None	Some	Most/All
Direct Instruction	7	32	61
Whole Class Discussion	95	4	1
Working on Applied Problems	72	10	18

Table 4: Frequency of MQI elements (in precentages)

	Score Point		
Code	Low	Mid	High
Richness			
Linking	71	24	5
Explanations	73	25	2
Multiple Methods	85	13	1
Generalizations	96	3	1
Math Language	58	37	5
CCASP			
Student Explanations	80	18	1
Student Math Questioning and Reasoning	81	16	2
Enacted Task Cognitive Activation	69	27	3
Errors and Imprecision			
Major Errors	94	5	1
Language Imprecision	84	15	1
Lack of Clarity	90	21	2

Table 5: Regression of classroom characteristics on MQI dimensions

	MQI Dimension		
Classroom-Level Characteristic	Richness	Errors	CCASP
% FRPL	0.32 (0.24)	0.17 (0.23)	0.30 (0.23)
% ELL	0.80** (0.23)	-0.08 (0.23)	0.73*** (0.21)
% SPED	0.25 (0.38)	-0.71** (0.22)	0.63* (0.32)
Prior Math	0.31** (0.12)	-0.08 (0.11)	0.56*** (0.12)

Scores are standardized and classroom-level statistics are either in decimals or SD (prior math). i.e., 10% difference in FRPL classroom is associated with .03 SD difference in Richness Score.

*** p<.001
** p<.01
* p<.05
~ p<.10

Table 6: District differences in MQI score

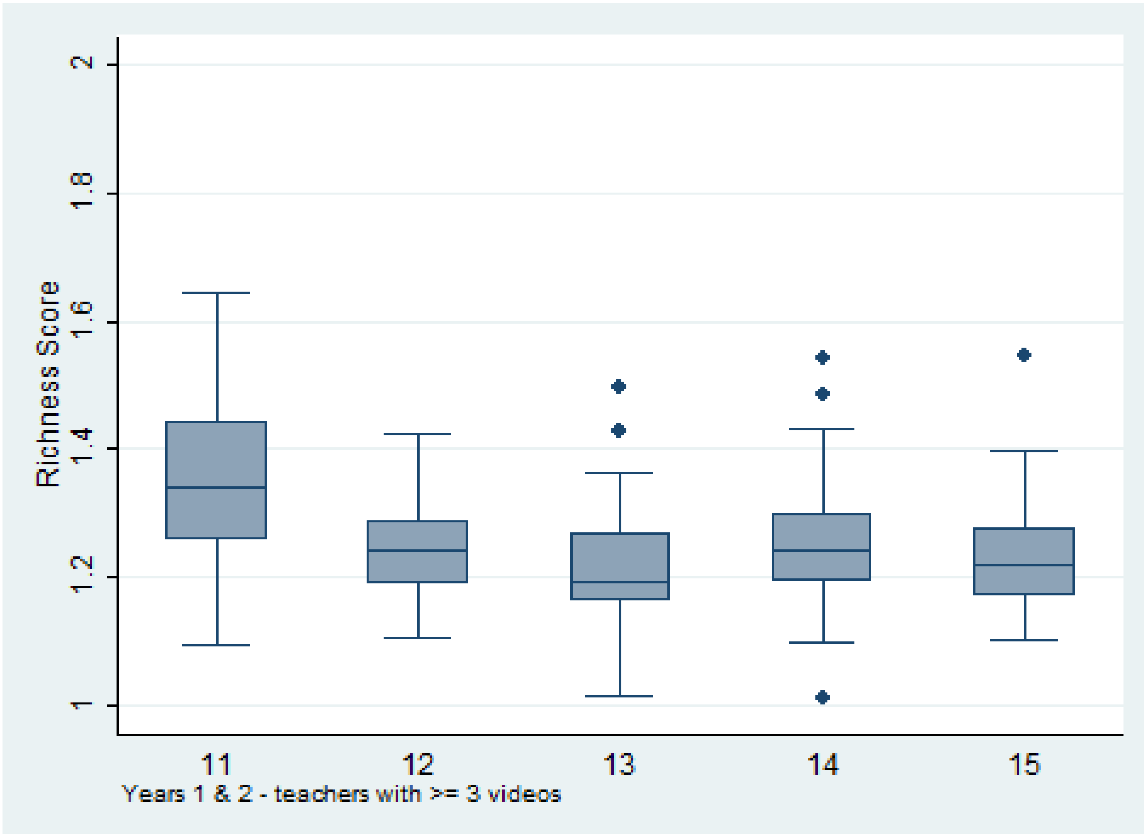
	MQI Dimension Score, Range (1, 3)			
District	Richness	Errors	CCASP	Lesson MQI
Overall, N=329	1.28 (0.11)	1.12 (0.08)	1.24 (0.14)	2.90 (0.48)
District 11, N=72	1.37 (0.13)	1.11 (0.10)	1.37 (0.16)	3.20 (0.58)
District 12, N=51	1.26 (0.07)	1.09 (0.06)	1.19 (0.08)	2.93 (0.28)
District 13, N=34	1.22 (0.10)	1.14 (0.08)	1.17 (0.12)	2.53 (0.53)
District 14, N=123	1.27 (0.08)	1.13 (0.08)	1.21 (0.11)	2.86 (0.39)
District 15, N=49	1.25 (0.08)	1.13 (0.07)	1.24 (0.11)	2.79 (0.39)

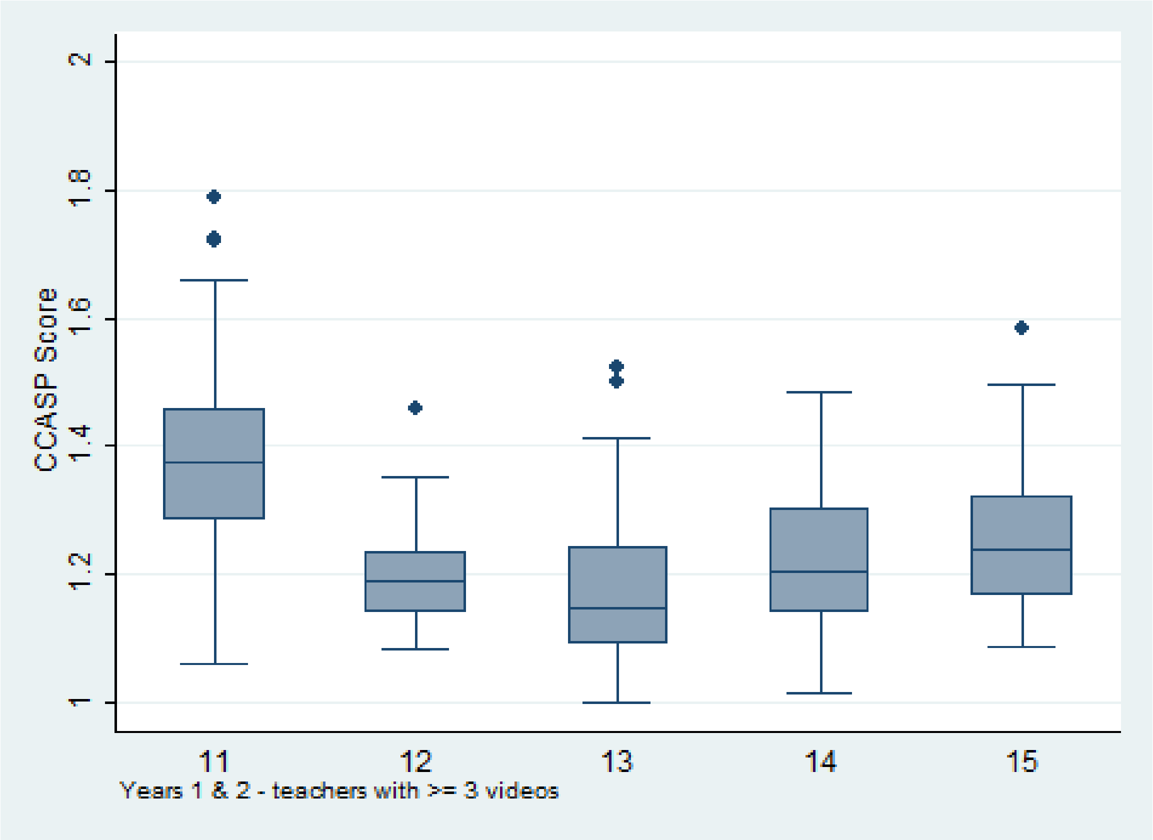
Table 7: Significance of between-district comparisons

	MQI Dimension Score			
District Comparison	Richness	Errors	CCASP	Lesson MQI5
11 = 12	***		***	**
11 = 13	***		***	***
11 = 14	***		***	***
11 = 15	***		***	***
12 = 13	*	**		***
12 = 14		***		
12 = 15		**	**	*
13 = 14	*		~	***
13 = 15			**	*
14 = 15			~	

*** p<.001
** p<.01
* p<.05
~ p<.10

Figure 1: Boxplots




CCASP Score
2
1.8
1.6
1.4
1.2
1
11
12
13
14
15
Years 1 & 2 - teachers with >= 3 videos