



Learning Lessons from Instruction: Descriptive Results from an Observational Study of Urban Elementary Classrooms

Heather C. Hill
Harvard University

Erica Litke
University of Delaware

Kathleen Lynch
Harvard University

Background: For nearly three decades, policy-makers and researchers in the United States have promoted more intellectually rigorous standards for mathematics teaching and learning. Yet, to date, we have limited descriptive evidence on the extent to which reform-oriented instruction has been enacted at scale.

Purpose: The purpose of the study is to examine the prevalence of reform-aligned mathematics instructional practices in five U.S. school districts. We also seek to describe the range of instruction students experience by presenting case studies of teachers at high, medium and low levels of reform alignment.

Participants: We draw on 1,735 video-recorded lessons from 329 elementary teachers in these five U.S. urban districts.

Research Design: We present descriptive analyses of lesson scores on a mathematics-focused classroom observation instrument. We also draw upon interviews with district personnel, rater-written lesson summaries, and lesson video in order to develop case studies of instructional practice.

Findings: We find that teachers in our sample do use reform-aligned instructional practices, but that they do so within the confines of traditional lesson formats. We also find that the implementation of these instructional practices varies in quality. Furthermore, the prevalence and strength of these practices corresponds to the coherence of district efforts at instructional reform.

Conclusions: Our findings suggest that unlike other studies in which reform-oriented instruction rarely occurred (e.g. Kane & Staiger, 2012), reform practices do appear to some degree in study classrooms. In addition, our analyses suggest that implementation of these reform practices corresponds to the strength and coherence of district efforts to change instruction.

VERSION: November 2020

Learning Lessons from Instruction:

Descriptive Results from an Observational Study of Urban Elementary Classrooms

Heather C. Hill

Harvard Graduate School of Education

Erica Litke

University of Delaware

Kathleen Lynch

Harvard Graduate School of Education

Citation:

Hill, H., Litke, E., & Lynch, K. (2018). Learning lessons from instruction: Descriptive results from an observational study of urban elementary mathematics classrooms. *Teachers College Record*.

Author Note

The research reported here was supported by the National Science Foundation (NSF) through Grants 108051 and 0918383 and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C090023 to the President and Fellows of Harvard College. The opinions expressed are those of the authors and do not represent the views of the NSF, the Institute, or the U.S. Department of Education. Correspondence concerning this article

should be addressed to Heather C. Hill, Harvard Graduate School of Education, 6 Appian Way,
Cambridge, MA 02138. Email: heather_hill@harvard.edu

STRUCTURED ABSTRACT

Background:

For nearly three decades, policy-makers and researchers in the United States have promoted more intellectually rigorous standards for mathematics teaching and learning. Yet, to date, we have limited descriptive evidence on the extent to which reform-oriented instruction has been enacted at scale.

Purpose:

The purpose of the study is to examine the prevalence of reform-aligned mathematics instructional practices in five U.S. school districts. We also seek to describe the range of instruction students experience by presenting case studies of teachers at high, medium and low levels of reform alignment.

Participants:

We draw on 1,735 video-recorded lessons from 329 elementary teachers in these five U.S. urban districts.

Research Design:

We present descriptive analyses of lesson scores on a mathematics-focused classroom observation instrument. We also draw upon interviews with district personnel, rater-written lesson summaries, and lesson video in order to develop case studies of instructional practice.

Findings:

We find that teachers in our sample do use reform-aligned instructional practices, but that they do so within the confines of traditional lesson formats. We also find that the implementation of these instructional practices varies in quality. Furthermore, the prevalence and strength of these practices corresponds to the coherence of district efforts at instructional reform.

Conclusions:

Our findings suggest that unlike other studies in which reform-oriented instruction rarely occurred (e.g. Kane & Staiger, 2012), reform practices do appear to some degree in study classrooms. In addition, our analyses suggest that implementation of these reform practices corresponds to the strength and coherence of district efforts to change instruction.

EXECUTIVE SUMMARY

BACKGROUND

For nearly three decades, policy-makers and researchers in the United States have promoted more intellectually rigorous standards for teaching and learning in mathematics. These standards ask that students reason, communicate about mathematical ideas, and problem-solve; in response, teachers must develop instructional practices that encourage these student activities. Yet, to date, we have limited descriptive evidence on the extent to which reform-oriented instruction is being enacted at scale.

RESEARCH QUESTIONS

This study examines the prevalence of reform-aligned mathematics instructional practices in fourth and fifth grade classrooms from five U.S. school districts. We also describe the range of instruction students experience by presenting case studies of teachers at high, medium and low levels of reform alignment.

SETTING AND PARTICIPANTS

We draw on 1,735 video-recorded lessons from 329 fourth and fifth-grade elementary teachers in five U.S. urban districts on the East coast. Two districts were large and three were small to mid-sized. Though not random or representative, our district sample contains variation in districts' commitment to and efforts at implementation of instructional reforms. For example, three of the districts had adopted reform-oriented curriculum materials and provided targeted professional development and coaching around reform-oriented practices. The other two districts, while committed to the idea of reform, engaged in less coherent instructional support around these ideas. Teachers in our sample were generally representative of the teacher population in their respective districts. Participating districts enrolled at least 70% nonwhite students, and between 60% and 80% of students qualified for free or reduced price lunch across study districts.

RESEARCH DESIGN

The study design is largely descriptive, and includes quantitative and qualitative components. We first describe the lessons in the sample using item-level scores from a mathematics-focused classroom observation instrument, the Mathematical Quality of Instruction (MQI). These scores depict the extent to which instruction includes specific reform-aligned instructional practices, such as mathematical explanations, student questioning and reasoning, and teacher use of student thinking, across the sampled lessons. We also draw upon interviews with district personnel, lesson summaries written by raters, and the video-recorded lessons themselves to develop case studies of instructional practice.

FINDINGS

We find that teachers in our sample do use reform-aligned instructional practices, but that they do so within the confines of traditional lesson formats. For instance, even high-quality instruction as measured by MQI, in which there are numerous instances of student mathematical thinking and reasoning and relatively rich presentation of the mathematical content, typically proceeds in a teacher-directed format, with generous student-to-teacher talk but little student

discussion or inquiry. Typical instruction features few mathematical errors, but only occasional elements of rich mathematics. Tasks seldom challenge students to think deeply about the mathematics, and instruction generally proceeds without significant student explanations or mathematical reasoning. Instruction at the lower end of the quality range features thin and often confusing mathematics, as well as frequent classroom distractions. Finally, a cross-district analysis suggests that the prevalence and strength of reform practices is related to the coherence of district instructional reforms. Stronger average instructional quality corresponded to the concurrent and sustained presence of reform-aligned curriculum materials, professional development, and leadership support.

CONCLUSIONS AND RECOMMENDATIONS

Unlike other pictures of contemporary classrooms, in which reform-oriented instruction rarely occurred, reform practices do appear in sampled classrooms, although with variable quality and in largely teacher-controlled formats. This suggests several pathways for future reform. First, the finding that cognitively challenging, mathematically rich instruction occurred in teacher-controlled classrooms may indicate that teachers are more amenable to reforms that build incrementally upon existing practice, rather than require a wholesale revision of instruction. Second, while most teachers appeared to embrace reform goals and instructional practices, missing from some implementation was technical skill – the commonplace but often detail-laden competencies that enable, for instance, teachers to correctly and completely present a mathematical explanation, foster connections between two representations or ideas, or launch a task in a way that does not devolve its cognitive demand. Finally, our inability to generalize results past our five districts illustrates the need for nationally representative studies of mathematics instruction.

Learning Lessons from Instruction:

Descriptive Results from an Observational Study of Urban Elementary Classrooms

For almost three decades, U.S. policymakers and professional associations have formulated standards for more intellectually rigorous mathematics teaching and learning in the nation's schools. The National Council of Teachers of Mathematics (NCTM) led in 1989 with a set of standards whose ideas were later reaffirmed and extended (NCTM, 2000; 2014); many states followed by aligning their own instructional guidelines to these documents (Loveless, 2001). More recently, the Common Core State Standards (CCSS) have called for more student engagement in key mathematical practices, such as mathematical modeling, the use of mathematical structure to solve problems, and quantitative reasoning and argument (National Governors Association Center for Best Practices, Council of Chief State School Officers [NGA], 2010). Cumulatively, these efforts share several goals, including transforming classrooms from teacher-centered environments to ones that feature student engagement in discussion and collaboration; increasing the cognitive challenge of student tasks; increasing the presence of conceptually-focused mathematics, including more mathematical explanations, sense-making and reasoning, and emphasizing the structure of the discipline and connections between topics.

Evidence regarding the extent to which these reform ideals have materialized in classrooms over this time period has been sobering, however. Observational studies of U.S. classrooms broadly suggest that instruction is largely teacher-centered, procedurally-oriented, and features low-demand student activities such as listening to presentations and practicing procedures (Hiebert et al., 2005; Weiss, Pasley, Smith, Banilower, & Heck, 2003). This has been particularly true of urban schools and schools serving large percentages of students of color or students living in poverty (Boston & Wilhelm, 2015; Diamond, 2007; Kane & Staiger, 2012).

Conditions experienced by such schools—such as strong accountability pressures or underprepared or novice teachers—may enable the persistence of these traditional instructional techniques (Diamond, 2007; Milner, 2013).

Most recent research in this field, however, has relied upon purely quantitative descriptions of practice, potentially missing key subtleties in teachers' enactment of reformers' ideas. Recent work, as well, attends mainly to the cognitive demand of student work and student participation in the lesson, rather than attempting to assess the extent to which teachers explicitly attend to mathematical concepts (Hiebert & Grouws, 2007), regardless of instructional format. To provide an in-depth view of teachers' practice, and to estimate the extent to which teachers attend to mathematical concepts, we conducted a video-based observational study of elementary mathematics instruction in five urban districts. With data from 329 teachers and 1,735 lessons, we asked:

- How *frequently* does instruction include reform-oriented practices? Does this vary by district?
- What is the *range* of overall instructional quality experienced by students in these five districts, and what does the instruction look like?

Our results present scores from a classroom observation rubric, then contextualize that quantitative data via case studies of teachers in the 10th, 50th and 90th percentile of instructional quality. We also describe the association between district reform strategies and teachers' instructional practices (see also Boston & Wilhelm, 2015).

Literature Review

Mathematics education and policy has a 30-year history of calls for more student thinking and reasoning, more meaning-centered mathematics, and more resemblance between the

mathematics in school and the mathematics practiced in the discipline itself (NCTM, 1989; 2000; 2014). Researchers have followed these calls with correlational evidence that curriculum materials oriented toward student inquiry (Saxe, Gearhart, & Nasir, 2001) and instructional practices that maintain a focus on problem-solving and student cognitive challenge (Stein & Lane, 1996; Tarr et al., 2008) predict better student outcomes. Current U.S. mathematics standards continue this focus on students' perseverance, problem-solving and argumentation.

Looking at existing studies of mathematics instruction, we see some evidence that teachers have embraced such practices. For example, in a nationally representative survey of teachers, Banilower and colleagues (2013) found that 97% of elementary school teachers believed that students should have frequent opportunities to share their mathematical thinking. Case studies and additional qualitative evidence suggest that in urban schools in particular, some teachers utilize reform-oriented instructional practices in ways that promote higher-level student engagement and learning. For instance, the QUASAR project worked with teachers in six schools over a several-year period and found that teachers in the program offered more conceptually-based and cognitively challenging instruction. Through an in-depth multiple case study analysis, Boaler and Staples (2008) described mathematics instruction at one high-poverty urban high school that was rich in cognitive demand and challenge, emphasized open-ended problems, and encouraged multiple solution strategies.

However, larger-scale inquiry into instruction has yielded less evidence that reform-aligned practice has spread widely into both U.S. classrooms generally and to urban classrooms in particular. The most detailed evidence on this front comes from nationally representative studies, reviewed below. We organize this review based on the conceptual framework proposed by Hiebert and Grouws (2007), who distinguish between instruction with explicit attention to

mathematical concepts and instruction that fosters student struggle with key mathematical ideas. Like these authors, we argue that these two aspects of mathematics instruction may vary independently of one another; teachers may attend to mathematical concepts through explanation and representations, for instance, but do so mainly in a direct instruction format. Following those who advocate a recentering of classroom mathematical discourse away from teachers and toward students (Stein, Engle, Smith, & Hughes, 2008), we also review the research literature on instructional formats. Reviewing the research in these categories sets the stage for our own investigation, which explored these dimensions of instruction in urban districts.

We begin with the *format of instruction*, which encompasses the extent of teacher control of the lesson, student participation in mathematical discussions, and student work on applied problems. Interest in such formats dates to the 1970s, when research in the process-product tradition measured the amount of class time spent in each format and correlated time spent with students' mathematical and other outcomes (see, e.g., Good & Grouws, 1977 and Brophy & Good, 1986 for an excellent review of this research). Later, many scholars promoted mathematical discourse within classrooms (e.g., Lampert, Rittenhouse, & Crumbaugh, 1996; Nathan & Knuth, 2003; O'Connor, 2001; Walshaw & Anthony, 2008) and a focus on student problem-solving over teacher-directed instruction (Hiebert et al., 1996). Yet existing survey evidence seems to suggest that most classrooms have remained traditional in format. For instance, Weiss and colleagues (2003) observed a geographically dispersed sample of 57 teachers, finding that 84% of K–5 mathematics instructional time was spent in whole class or individual formats and only 16% was spent on collaborative student group work. The Third International Mathematics and Science Study (TIMSS) captured a slightly larger (83 classrooms) but nationally representative sample of U.S. 8th grade mathematics lessons (Hiebert et al., 2005).

Using lesson transcripts, the authors calculated a ratio of eight teacher words to every one student word, and also noted that most student utterances (71%) contained four or fewer words, suggesting that TIMSS teachers dominated classroom discourse.

These studies paint a similar picture regarding the *cognitive challenge of student tasks*. High-cognitive demand tasks require complex reasoning and problem-solving, as well as comprehension, synthesis, or interpretation, such as by solving problems using multiple methods and/or explaining and justifying mathematical ideas (Doyle, 1998; Stein, Grover, & Henningsen, 1996). Hiebert and colleagues (2005) documented that U.S. TIMSS lessons contained low levels of mathematical challenge, with most tasks enacted either briefly or procedurally, and found that the majority of student work time was spent repeating procedures. Tasks that were initially posed at high levels of cognitive demand were ultimately devolved by teachers to require little more than student listening and note-taking. The classroom observations conducted by Weiss and colleagues (2003) in elementary classrooms also returned low ratings for student intellectual engagement and investigation.

Information on the presence of *conceptually-focused features of instruction* at scale is more scant, but similarly indicates that instruction may not be as intellectually rigorous as reformers would hope. By conceptual features, we refer to instructional moves that allow teachers to attend to either the meaning of mathematical concepts or mathematics practices – regardless of whether the lesson content is cognitively challenging for students or not (Hiebert & Grouws, 2007). These moves include providing mathematical explanations and making connections across mathematical representations. TIMSS raters (Hiebert et al., 2005) judged only 1% of enacted U.S. problems as making visible mathematical connections and relationships, despite the fact that 17% of problems had the potential for such connections. Similarly, only 30%

of U.S. TIMSS lessons had partial or complete mathematical reasons or justifications attached to the problems or exercises used. Weiss and colleagues (2003) offered a similar depiction in U.S. elementary schools in their sample: Of nine indicators of K–5 mathematics instruction, sense-making around content and the portrayal of mathematics as an investigation-based discipline received the lowest ratings.

Finally, numerous scholars during this period noted problematic features of mathematics classrooms. The early process-product literature identified teacher clarity as an important component of effective instruction (see, e.g., Good & Grouws, 1977), and by 2000, influential critiques of mathematics reform focused on the mathematical precision and clarity of what occurs in classroom (Wilson, 2008; Wu, 1997). Both early and later work also found that time on task predicts students' mathematics outcomes (Brophy & Good, 1986; Stronge, Ward, & Grant, 2011). Large-scale research indicates that both features—time on task and mathematical precision—could be improved in U.S. classrooms. Weiss and colleagues (2003) estimated that 8% of observed class time was spent on non-instructional activities. TIMSS (Hiebert et al., 2005) found that 22% of U.S. lessons contained a non-mathematical segment of at least 30 seconds or longer. Furthermore, the accuracy of mathematics instruction appeared generally strong but not ideal; Weiss et al. (2003) reported that lessons scored a 3.73 out of 5.0 on this indicator, and Hill et al. (2008) also suggested that some instruction among teachers is mathematically problematic.

Examinations into urban classrooms reflect the trends described above. For example, Diamond (2007) presented results from a 1999–2000 investigation into the course of mathematics reform in eight Chicago public schools. Despite policies that explicitly encouraged student critical thinking and problem-solving in the district, results showed that teachers mainly used didactic instructional techniques and failed to encourage and support student questioning.

This trend was most pronounced among schools serving large populations of African-American students. The Measures of Effective Teaching (MET) study, conducted over a decade later in six urban districts, found classroom instruction to be generally characterized by low cognitive demand and few conceptually-focused features such as mathematical explanations, analyses, generalizations, representations, and problem solving (Kane & Staiger, 2012). Boston and Wilhelm (2015) similarly found little rigorous mathematical discussion and low enacted task cognitive demand in the majority of lessons they observed in four urban districts.

Thus, both the general and urban-specific mathematics literatures suggest that despite teacher support and strong instruction in individual schools or classrooms, reform has not penetrated far into most American classrooms. However, we argue for a reassessment of this claim using a combination of quantitative and qualitative data from five urban districts engaged, in varying degrees, in mathematics reform. One goal of this work is to determine whether separately examining attention to concepts and student cognitive demand, rather than jointly as in many current observation instruments (e.g., Boston & Wilhelm, 2015), yields a more nuanced picture of reform implementation. Another goal is to extend prior quantitatively-focused studies by purposively sampling and then conducting case studies of teachers at the 10th, 50th and 90th percentiles of quality. Doing so allows us to demonstrate the range of instructional quality students experience, and also provides insight into the ways teachers implement reform practices. Finally following Boston and Wilhelm (2015), we also focus on potential connections between district mathematics reform strategies and teachers' instructional characteristics. We describe our sites and methodology in the section below.

Methods

This paper is largely descriptive in nature. As King, Keohane, and Verba (1994) argued, description is fundamentally important both by itself and as a critical first step toward developing causal explanations. By providing descriptive evidence, we can develop a picture of the extent to which reform-oriented practices are currently in use in our districts' classrooms, and inform next steps for those interested in promoting such practices in similar environments.

District and Teacher Sample

We conducted our study in five urban East Coast districts during the 2010–2011 and 2011–2012 school years. We sampled districts based on convenience, via mathematics-focused professional contacts (e.g., colleagues in the National Council of Supervisors of Mathematics) or through non-mathematics contacts in central district offices. Though neither random nor representative, this sample does contain some degree of variation in districts' commitment to and efforts at implementation of instructional reforms. For example, three of the districts had adopted reform-oriented curriculum materials and provided targeted professional development and coaching around reform-oriented practices. The other two districts, while committed to the idea of reform, engaged in less coherent instructional support around the reforms. Although a more representative sample would allow generalizations to the population of urban districts, such a plan would have required a large sample size (dozens if not hundreds of districts) and thus precluded the use of video and observational methods. Two districts were large, and three were small to moderate in size. Each district enrolled at least 70% non-white students, and across all five, students' eligibility for free or reduced price school lunch ranged from roughly 60% to 80% (see Table 1). Slightly over 60% of sampled teachers reported using a reform-oriented curriculum material as their primary teaching resource, roughly double what was reported by teachers answering a 2012 national survey (Banilower et al., 2013).

The teacher sample was comprised of fourth- and fifth-grade teachers from these five urban districts. In four districts we recruited schools based on district referrals and size, requiring that they have at least two regular education teachers in each of the two grades. A comparison of schools involved and not involved in the project suggests that the two were quite similar in terms of student demographics (see Table 1), and that prior to starting data collection, project schools actually performed slightly worse on state assessments than non-participating schools.¹

Insert Table 1 here

Following principal recruitment, we invited 583 teachers from these four districts to join the study; 318 did so, for an enrollment rate of 55%. Enrolled and non-enrolled teachers in these districts were similar on key characteristics and did not display statistically significant differences in pre-study value-added scores (available from authors on request). In our fifth district, we recruited teachers at 18 schools based on principal and teacher willingness to participate in a longitudinal study of professional development, enrolling 88 teachers and collecting video, by design, from roughly half (40).

Data Collection and Measures

We used the video recordings to answer our questions about the nature of and variability in instruction in our five districts. In addition, we collected teacher background information and conducted interviews with district personnel to establish each district's history of supports for the NCTM and Common Core standards. We describe each below.

Video recordings. The study design called for recording each teacher's instruction six times over the course of two years. Some teachers, however, left the study after the first year (typically due to changes in school or grade) and were replaced by new teachers. This left both of the latter groups with only three videos. Several teachers had missing videos due to health and technical issues. For the analysis below, we include in the sample all teachers who completed at least three video-recorded lessons across the two academic years (329). The final dataset includes 1,735 video recorded lessons.

Video capture was completed via a multi-camera unit; school researchers set up the cameras, then left the classroom in order to minimize interference in the teacher's lesson. Teachers chose the day and time to be video recorded, but were asked not to schedule recording for a day on which they would be giving an extended assessment. Prior research suggests that when teachers were given discretion to choose their 'best' classroom videos from a set, the chosen and not-chosen videos provided similar information about teachers' instructional quality (Ho & Kane, 2013). Lessons lasted approximately 45–60 minutes.

The content of each lesson was scored by trained, certified raters using the Mathematical Quality of Instruction (MQI) classroom observation instrument (Hill et al., 2008). Raters were required to complete online training modules including practice videos and automated feedback about submitted scores. Raters then passed a certification exam and participated in ongoing calibration seminars over the entire period of scoring.

The instrument organizes the rating of lessons by the categories highlighted in the review above, including separate sections for instructional formats (*Instructional Format*), conceptually-focused features of instruction (*Richness of the Mathematics*), and student cognitive demand (*Common Core-Aligned Student Practices*). Other instruments tend not to have such a clear-cut

structure, in some cases conflating conceptually-focused features and student cognitive demand, even though in practice teaching may be strong in conceptually-focused features and weaker in student cognitive demand (e.g., when a teacher lecture presents detailed mathematical explanations). The MQI also provides detailed information on teacher errors (*Teacher Errors and Imprecision*), shown in prior research to be a less frequent but significant feature of U.S. mathematics 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). Table 2 displays a list of relevant elements, organized by dimension, and also provides information about score points. Finally, raters were asked at the end of each lesson to do three additional activities: assign an overall MQI score on a scale of 1 (*low*) to 5 (*high*), score each lesson for overall Richness, Common Core-Aligned Student Practices and Errors on a scale of 1 (*low*) to 3 (*high*), and write a summary describing the content, strengths, and weaknesses of the lesson.

Table 2 about here

Lessons were broken into 7.5-minute segments for scoring; all segments from each lesson, save for segments less than one minute long at the end of lessons, were scored by two randomly assigned raters on the 15 MQI elements. Raters were blind to the district from which each lesson was derived and to the history of reform efforts in these districts generally. The adjusted intraclass correlations of the within-year item-level scores ranged from 0.20 (mathematical generalizations) to 0.72 (student explanations)—low by conventional standards for reliability, but typical for observational metrics (Bell et al., 2012; Kane & Staiger, 2012).

Internal analyses of this data showed reasonable interrater agreement (an average of 0.77 across elements), but strong variability across lessons within teachers. Because this study was intended to depict typical instruction across *the lessons* our sample, we are less concerned about the low reliabilities than we would be than if we intended to make inferences about a specific teacher's instruction, or to correlate instruction with teacher characteristics or student outcomes.

Teacher background. A teacher survey provided information about teachers' backgrounds, including years of experience and mathematics-related leadership activities. The survey also included an assessment of teachers' mathematical knowledge, containing both Mathematical Knowledge for Teaching items (Hill, Schilling, & Ball, 2004) as well as more conventional mathematics problems. We used item response theory to construct teacher scores on the mathematical knowledge assessment; scores are expressed in standard deviations from the average teacher in the sample. We use the background and mathematical knowledge measures to describe our case study teachers; a more thorough analysis of the general relationship between teacher characteristics and instructional quality is contained in Hill, Blazar, & Lynch (2015).

District interviews. Seven 45- to 75-minute telephone interviews were conducted with the individuals who served as district mathematics coordinators in the years leading up to and during the study. Interviews were conducted by one of the study authors and covered topics such as math curriculum adoption, the mathematics professional development available to teachers and principals, and the coherence of instructional guidance in the years during and prior to the study. Interviews were audio recorded and subsequently transcribed for analysis.

Analyses

In order to describe mathematics instructional quality in these urban districts, we calculated segment-level frequencies for the MQI elements. Averaging these features to the

lesson, teacher, and district level, also allows us to examine the presence of these elements at more aggregate levels of data. Although two raters scored each lesson, we selected one set of scores at random. This allows for the presentation of integer values in the tables and allows us to use the more intuitive phrasing “percent of lesson segments” rather than the more awkward “percent of ratings of lesson segments” in the results section without changing the overall average depiction of instruction in our sample. This procedure left us with 13,166 segments nested in the 1,735 lessons.

One limitation of presenting the frequency of reform practices is that these statistics do not fully convey the nature of mathematics lessons, both because the MQI omits elements of reform instruction that are hard to define in practice (e.g., contextualizing and decontextualizing; the use of structure) and because students’ experiences are more than an accumulation of specific practices. Frequency tables also provide little information about the ways teachers combine elements to implement reform practices, and the micro-quality of implementation. They also fail to provide much guidance for those looking to engage in the work of instructional improvement. To address these issues, we illustrate instructional practice at the center and tails of the MQI distribution by conducting a qualitative analysis of lessons from teachers identified as in approximately the 90th, 50th, and 10th percentiles of the overall MQI distribution. To identify teachers at or near these percentiles and to guard against outliers, we selected teachers whose aggregated and averaged scores matched scores from a hierarchical model, which accounts for missing data by shrinking teachers’ scores toward average (Raudenbush & Bryk, 2002). We extracted a group of 39 teachers (eight at the 90th, 16 at the 50th, and 15 at the 10th) consistent in their percentile ranking on both sets of scores.²

To describe instruction at each percentile, two authors independently read through all

lesson summaries at each of the three levels, noting themes observed in the instruction. The authors discussed their observations, then created short, general depictions of instruction at each level. We then selected one teacher from each group to illustrate the overall profile we articulated. Two authors watched these teachers' lessons, taking notes and (a) selecting instances illustrative of the instruction at that level, and (b) noting instances of divergence from the common themes. Finally, these authors reconciled notes and developed the profiles below.

To provide evidence on the extent of district efforts to enact the NCTM and Common Core standards, we conducted an analysis of district mathematics coordinator interviews. We looked for commonalities and differences among these districts along three major dimensions: in the curriculum materials and professional development available to teachers; in support for principal learning and principal support for the reforms; and in the coherence of reform initiatives with other instructional guidance, both within the district at the time of the study and across time since the 2000 NCTM standards. We begin with this evidence below.

Results

District Efforts to Support Reform-Aligned Instruction

All five districts had actively committed to supporting their teachers' use of reform-aligned instructional practice, although with varying levels of strength and internal coherence.³ In Districts A, B, and E, efforts to improve instruction proceeded via several methods: (a) with the use of federal dollars to provide professional development to teachers; (b) by similarly using federal dollars to provide 1:1 coaching for teachers; and (c) through principal professional development around reform-aligned instruction and ongoing updates regarding math office activities and goals. In Districts A and B, these efforts were further supported by a set of curriculum materials, *Investigations*, an NSF-funded curricula developed in order to encourage

student exploration, inquiry and meaning-making around mathematical ideas. Efforts in A and B were further supported by the use of federal and foundation dollars to provide large numbers of teachers with intensive mathematics- and pedagogy-focused professional development from nationally-known programs. District E used a set of curriculum materials, *Math Expressions*, that its mathematics coordinators described as taking a “middle ground” between reform and more traditional approaches, and provided more sporadic professional development around standards-based instruction from both nationally known programs and district staff.

Overall, the policies in these three districts yielded a coherence seldom observed in district instructional guidance; it is also notable that this coherence extended over a significant period of time. In District A, the district adopted a formal strategy laconically titled “The Math Plan” in the early 2000s; strong consistency in this strategy and in leadership support for the strategy continued for roughly a decade and persisted into the study, although in a slightly weakened form due to conflicting instructional guidance to teachers. In Districts B and E, coherence and consistency lasted a shorter period of time, through the 2008–2009 school year, when new superintendents shifted focus to other mathematics and non-mathematics interventions (e.g., Response to Intervention).

Districts C and D presented a different case. Although both had formally adopted standards similar to the NCTM and later the Common Core standards, and although both made available reform-aligned curricula (NSF-funded ThinkMath in C; Everyday Mathematics in D) the scope of math office efforts was significantly smaller. Professional development was mainly provided by district mathematics staff and local consultants, and reached a far smaller proportion of teachers than in A, B, and E. Other instructional guidance and improvement initiatives, for instance in reading or content-generic topics such as individualized instruction, competed for

elementary teachers' time and attention. In District C, the professional development office was in fact reorganized to be content-generic, and a content-generic teacher evaluation system was put in place with strong incentives for teachers to score well. In District D, math coordinators described teacher and principal resistance to the proposed adoption of NSF-supported curriculum materials such as *Investigations*. Thus, while the mathematics offices within these two districts formally supported NCTM and Common Core standards, their efforts enjoyed less support and often conflicted with other sources of instructional guidance.

Frequency of MQI Elements

Next we describe the observed frequency of the instructional formats by lesson segment (Table 3). We find that direct instruction dominated in our sample; 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, leaving only a small fraction (7%) solely for independent student work or discussion. Whole-class discussion (which required students to comment on one another's mathematical ideas or solutions rather than responding solely to the teacher) occurred in some or all of only 5% of segments. Working on applied problems, where such problems were defined as involving a contextualized situation (e.g., finding "total distance traveled" from two portions of a trip), occurred in 28% of segments.

Table 3 here

At the segment level, we thus see the instructional formats used in these districts as similar to traditional mathematics instruction: teacher-led, with little student discussion and with

only a modest number of applied/contextualized problems. But these descriptions may obscure other trends in the data, especially if teachers have integrated reform-aligned practices within these older formats.

Table 4 shows that this may be the case. Here, we present the percentage of segments at each score point on the specific conceptually-focused dimensions of the MQI. Over one 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, scoring a mid or high on the Explanations code. This indicates that in these segments, students and teachers engaged to some degree in making meaning of mathematical ideas, a key practice in improving student understanding in mathematics. Another 29% of segments contained connections among different representations of mathematical ideas or procedures (e.g., noting linearity in both a graph and a table of the same equation) 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 to a set of similar problems. Though these proportions seem modest, they represent the segment-level prevalence of these activities; aggregated to the lesson level, we find 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 at least one of these three features.

Table 4 here

Focusing on the Richness dimension, we find a non-trivial amount of meaning-oriented instruction occurring in classrooms, albeit meaning primarily conveyed via direct instruction.⁴ The MQI score points also make possible an investigation of 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 mathematical explanation refers to explanations that were partial, lack detail, or are specific to only one problem (e.g., why 5 is the answer to $25 \div 5$). 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 linking/connections and multiple methods elements, where a high score signifies that the element was marked by explicit and/or detailed connections or comparisons between the representations, topics, or methods. These score distinctions are supported by research showing the effectiveness of the practices noted under the high score point (e.g., for connections between multiple methods see Rittle-Johnson & Star, 2007). However, Table 4 shows these MQI instructional features were seldom enacted with such characteristics in sampled classrooms. About 5% of segments contained high-quality connections, and less than 3% of segments contained high-quality explanations. Detailed and explicit links between multiple methods occurred in 1% of segments.

Two other elements within the MQI that focus on disciplinary practices—teacher and students’ use of mathematical language and developed generalizations—returned more discouraging results. Frequencies for the former show that while mathematical terminology was often used in these segments (with 42% of segments rated above *low*), it was rarely used densely or with explicit attention to developing fluency and precision in classroom talk (5% of segments). Generalizations, which captures instances in which teachers and/or students examine

multiple instances or examples of a mathematical phenomenon then make a general statement, was coded at either mid or high in only 4% of segments (with 15% of lessons having at least one segment that scored mid or high). Both are central to the Common Core practice standards, but frequent high-level use appears uncommon in these classrooms.

Turning now to the MQI elements that concern student involvement in cognitively challenging work, we see some evidence that students are afforded opportunities to engage to some degree in such practices, particularly at the lesson level. For instance, in 19% of segments, students ventured 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. Likewise, in 18% 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, offering claims or counter-claims about mathematical phenomena, forming conclusions based on patterns, or engaging in reasoning in the general sense (e.g., asserting that “A cube has 24 right angles total on its faces, because each face has four right angles and there are six faces”). Thirty percent of segments had some form of student mathematical activity beyond listening to lectures and repeating procedures; however, the fraction of segments in which such cognitively challenging work dominated the segment was small (3%). When aggregated across segments within lessons, we find that 84% of lessons contained at least one of these three MQI elements, and 21% of lessons contained one of these elements at a high level. This suggests that students’ opportunities to engage in cognitively challenging work exist, but likely occur in isolated moments rather than consistently across lessons.

The MQI also tracks the frequency with which the lesson veered from mathematics, or the teacher presentation lacked mathematical accuracy. Ninety-four percent of segments contained work connected to mathematics (not shown); aggregated to the lesson level, 28% of lessons had at least one segment scored as not connected to mathematics; these segments, for instance, included teachers passing out materials, answering the classroom phone, or redirecting disruptive students. Outright mathematical errors—e.g., teachers solving problems or defining terms incorrectly—were few, at 6% of segments, though these instances occurred in 29% of lessons. By contrast, 16% of segments contained an instance of language imprecision on the part of the teacher, for instance saying statements such as “division makes numbers smaller” without specifying that this statement is true only for whole numbers. While this is a relatively small number of segments, 50% of lessons contained at least one segment with an instance of language imprecisions. Raters judged another 10% of segments to feature unclear mathematical presentations on the part of the teacher (occurring across 37% of lessons). At the lesson level, 68% of lessons featured at least one of the errors elements at a mid or high level.

Given the nested nature of the sample, we investigated the degree to which the lessons that featured reform-oriented practices were distributed across teachers and districts in the sample. Figure 1 plots, by teacher, the proportions of segments rated as mid or high for at least one element within the Richness, Common Core-Aligned Student Practices, and Errors dimensions. Observers recorded the median teacher as having 22% of segments with at least one mid or high richness score; the mean was slightly higher, reflecting the longer right-hand-side tail. For elements of Common Core Student Practices, the data was more dispersed but the median was the same (22%). For Errors, the median teacher received a score of mid or high, meaning an error was present, in 10% of segments.

Figure 1 here

The presence of rich mathematics instruction and cognitively challenging student tasks differed by district in ways that suggest a reflection of the influence of district support for mathematics reform. Figure 2 displays histograms of raters' assessments of overall Richness and Common Core-Aligned Student Practices for each lesson (1=no richness, 3=very rich), plotted by district. In District A, where reform efforts were most intense and extended, and where a reform-aligned mathematics curriculum was in place, raters observed more mathematical richness in lessons, and correspondingly stronger Common Core-Aligned Student Practices. In Districts, C, D and E, raters were more likely to assign an overall lesson score of "low" for these two dimensions, perhaps reflecting the less coherent policy environment and lower levels of professional development support. In District B, which was similar to A in its policies and curriculum materials until a change in superintendent two years prior to the study, we see Richness and Common Core-Aligned Student Practices typical of Districts C, D and E, although District B had relatively more lessons rated above 1.5 on the Common Core-Aligned Student Practices scale. While there is some variation across districts, we see that the general shape of the distribution is similar, with the notable exception of Common Core-Aligned Student Practices in District A.

Figure 2 here

Illustrations of Instruction

To provide a description of what these reform elements look like in classrooms, and to further our understanding of how reforms translate into instructional practice, we illustrate typical teaching near the 10th, 50th, and 90th percentiles of the sample MQI score distribution. To do so, we sampled teachers, rather than lessons, because examining multiple lessons per teacher allowed us to aggregate information across lessons and thus better understand the strengths and weaknesses of urban mathematics instruction in our sample at these different levels of MQI quality. We sampled teachers across all districts as our goal was to illustrate instructional practices rather than to discuss district differences. (For a discussion of how instruction differs by district in this sample, see Blazar, Litke, & Barmore, 2016). At each percentile, we begin with a brief overview of segment-level scores for MQI elements and themes evident in raters' lesson summaries; we then present a snapshot of instruction from the teacher chosen to represent instruction typical for teachers in that stratum.

Teaching in the 90th MQI percentile. Table 5 shows segment-level MQI item frequencies for the eight teachers at the 90th percentile.

Table 5 here

This table shows that richness elements were fairly common in 90th percentile teachers' lessons, particularly compared to the full sample. Qualitative analyses of lesson summaries suggested that lessons in this category generally included depth in one or two elements of the MQI and seldom included all of the features simultaneously.⁵ In addition, lessons contained only occasional minor errors such as imprecisions or inaccuracies 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 a lesson.

Table 5 shows that more than half of segments recorded either a mid or high for Common Core Aligned Student Practices. Qualitative analyses of raters' lesson summaries suggested that students offered alternative solution methods, noticed mathematical patterns, and provided explanations and reasons. For example, in one lesson in which students multiplied a fraction by $2/2$ to obtain a common denominator, the teacher insisted that students provide a detailed oral explanation for why this process is considered multiplying by a whole. However, seldom did the lesson summaries suggest classrooms fully turned over to student investigation and discussion; instead, student contributions were most often invited and controlled by teachers, as the quantitative results for instructional formats suggest.

Thomasina. After reviewing the lesson summaries, we selected one fifth grade teacher, Thomasina, to illustrate instruction among teachers in the 90th percentile. Thomasina taught in a high-poverty school; in the second year of data collection, one third of her students were English language learners. At the time of the study, Thomasina had been teaching for just a few years. She reported using both a traditional mathematics textbook published by Harcourt as well as the Everyday Mathematics curriculum as supplemental material. Thomasina's mathematical knowledge for teaching was over a standard deviation above the average study teacher's, suggesting a strong command of the discipline.

Thomasina's lessons contained both richness elements and student involvement in mathematical thinking and reasoning. She incorporated these elements into quickly-paced lessons that consistently required students to solve what we call "small-sized" problems—not lengthy investigations, as per some reform visions, but nonetheless substantive tasks that

required students to engage in the practices found in Common Core and other standards. Often, these tasks demanded that students attach meaning to the mathematics they engaged in.

Thomasina's lessons also featured attention to fostering student fluency in mathematical language, another goal of the Common Core. We present Thomasina's lesson as an illustration of teaching that incorporates reform oriented practices at some depth, while doing so in the context of largely teacher-driven instruction.

A lesson on algebraic equations illustrates these themes. In this lesson, Thomasina gave the students situations such as, "Robert had 15 pieces of chocolate. He ate some and now has eight pieces left. How much chocolate did he eat?" Thomasina told students:

All right, we've worked on expressions and we talked briefly about equations, but we haven't really worked on writing them, right?... So, now, you're going to take what you've learned writing expressions [and write equations]. I don't want you to try and solve this equation. What I want you to do is to *write* the equation. I want you to tell me—don't answer the question, you're going to write an equation that fits this situation and tell what the variable stands for, okay?

In launching the task, Thomasina emphasized the focus on meaning-making, highlighting for students that they would be writing equations and making sense of the unknown in the given situation. She also explicitly required that students give explanations for how each element of the equation matched the situation, an element of meaning-centered mathematics instruction.

As the lesson progressed, Thomasina wrote out numerous situations, allowed students independent work time for each, then asked them to share their equations. As she did so, she repeatedly reminded students to "make sure you tell me what [the unknown] stands for." Thomasina also sequenced material in ways that kept the cognitive demand constant across the

lesson. For example, she varied the role of the unknown in the situation (e.g., the starting quantity, the amount added or subtracted, or the sum/difference) then later inverted the task, giving students an equation for which they had to write a situation (e.g., $8 - n = 5$). Toward the end of the lesson, Thomasina requested a context for the equation $(5 + 2) - n = 2$, requiring that students think about the order of operations. Wrapping up this example, Thomasina had the following exchange with her students:

Thomasina: You're going to listen because you're going to have to decide whether or not your classmate has come up with an acceptable equation or a situation for this equation...[Student M]?

Student M: Elmo had five crayons and gave Cookie Monster two more crayons. Well, Cookie Monster gave Elmo two more crayons but then, uh—

Thomasina: Uh-oh, she got stuck.

Student M: But then Elmo broke some.

Thomasina: Almost. Elmo had five—Elmo had five crayons, Cookie Monster gave him two more, but then he broke some; what should be next?

Student M: He had two left.

Thomasina: Now he has two left.

Multiple Students: How much does he have—

Student M: No, how many did he break?

Thomasina: How many did he break.

In this exchange, we can see several central elements of Thomasina's work with students. She assigned a relatively short but nevertheless demanding problem, one that resulted in a student getting, as she put it above, "stuck." This problem required students to contextualize an equation, a focus of the Common Core. Thomasina also urged other students to remain present in the discussion, and those students appeared to do so, often calling out answers and contributing ideas. These elements were present to some degree throughout this lesson and others.

Thomasina's efforts to maintain cognitive press extended into the time that students worked independently. Rather than devolving the cognitive demand of the tasks by intervening and directing students how to write situations or equations, Thomasina circulated through the classroom and asked pointed questions about students' thinking. Even in response to direct student requests for help, Thomasina often demurred, turning the question back to the students. For instance, one student showed Thomasina an incorrect situation for the equation $8 - n = 5$, and asked for her approval. Rather than correct the student's error, Thomasina asked, "Okay. What is it that you're trying to find out?... If this was the situation (points to student work), what is it you would have the person [answering your question] try and find out?" Thomasina waited while the student thought for a bit, allowing her to realize her mistake and revise her situation on her own.

Publicly engaging the class in understanding and correcting errors was another way Thomasina both maintained cognitive demand and focused students on making sense of content. For instance, one student (Student L) wrote the following situation for the equation $8 - n = 5$: "Maria has some dolls. She gave eight away. Now she has five. How many did she give away?" This student had written a situation that represents the equation $n - 8 = 5$, mistakenly assigning the number of dolls Maria started with to be the unknown. Rather than simply correcting her,

Thomasina engaged the class in writing the equation that fit Student L's situation correctly and reasoning about why it would be different than the equation given:

Thomasina: Okay. What equation should go with *that* situation? Maria...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?

Student L: Oh.

Thomasina: [Student L], what's the equation?... Tell me what you did. Tell me what equation it is.

Student L: Oh, I did "N minus eight equals five."

Thomasina: [Student L] did "N minus eight equals five."... Does that make a difference?

Multiple Students: Yes.

Thomasina: 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: A lower number than eight.

Thomasina: It has to be lower than eight. But in this case [*points at* $n - 5 = 8$], it has to be...?

Multiple Students: Greater than.

At this juncture, Thomasina asked the whole class to use number sense to check the reasonableness of their answers. Thomasina then asked Student L to state a new story problem that would match the equation $8 - n = 5$, and she responded, “Maria had eight dolls and she gave some away. Now she has five. How many did she give away?” By both articulating an equation from the student’s incorrect response and reasoning out why $8 - n = 5$ is not the same as $n - 8 = 5$, Thomasina engaged the class in sense-making and reasoning around the student’s error and the correct solution.

Finally, another characteristic of Thomasina’s instruction was her own generally precise use of mathematical language, and her encouragement of students to follow suit. For example, in a lesson on the area of a circle, Thomasina told students, “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.” She then spent a significant portion of the lesson developing precise definitions of radius and, subsequently, diameter:

Thomasina: 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 K: Like, it’s half of a circle?

Thomasina: 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....

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

Thomasina: I like the “line segment” and I like “that travels.” Does it always have to go—you can look at [our definition of radius]. Does it always have to go from the top to the bottom?

Multiple Students: No. [*Inaudible*].

Thomasina: ... There's something else it has to do....

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

Thomasina: 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.

Thomasina engaged in several language-related activities in this lesson excerpt: she encouraged students to offer a “full, complete” definition, and she also did not immediately correct an imprecise or incorrect response (“half of a circle”), instead referring the student back to an already-developed definition for radius. When the student again offered an imprecise response (“...travels from the top to the bottom”), Thomasina selected the correct parts of the student utterance (“I like...”) but pushed for a more accurate definition. In many ways, Thomasina’s instruction is deeply mathematical; she references already-developed definitions to help formulate others, salvages workable parts of the definition while discarding others, and pushes for complete and precise definitions (e.g., the definition does not include horizontal or vertical).

However, similar to the patterns described quantitatively above, these reform-oriented features of Thomasina’s instruction lie squarely within the traditional, teacher-centered format.

In the above example, for instance, Thomasina co-constructed a definition of diameter with students, directing the conversation and elaborating key ideas at the end of the excerpt.

Examinations of her conversations with students in the equation/expression lesson also suggest that Thomasina does, by far, the lion's share of talking in her classroom. Yet when students do offer observations, those observations are likely to be substantive, reflecting explanations, partial definitions, and problem solutions. This illuminates the ways in which her instruction, we argue, has adopted important reform-oriented elements.

Thus, Thomasina's instruction, as well as the instruction of other teachers in the 90th percentile, does afford students some aspects of reformers' vision such as engaging students in meaning making and explanations of mathematical ideas, but couches those opportunities in a largely traditional format. Like Thomasina's lessons, most instruction in this group featured the teacher introducing new material and allowing students time individually or in small groups to work on problems related to that content. This contrasts with visions of mathematics classrooms in which students develop understandings by first working on rich problems or participating in discussions, then solidifying that knowledge with the help of the teacher.⁶ Despite this, teachers in this category afforded students the opportunity to make meaning of mathematics within these more teacher-centered instructional formats. Students in these classrooms frequently engaged with mathematical ideas, asked mathematically motivated questions, and engaged in tasks that presented at least moderate cognitive demand. Teachers often made use of student ideas to further the mathematics. Teachers also injected disciplinary features into their instruction by focusing on meaning, precision in the use of language, and mathematical explanations. While this instruction may fall short of the idealized vision articulated by reform, we argue that it represents evidence that reform-oriented practices *have* made inroads into some classrooms and

that moving beyond format of instruction allows for a more nuanced depiction of instruction. Indeed, this picture of instruction may represent an integration of reform-oriented practices in ways that both benefit students and work as a leverage point, building on formats teachers typically engage in and moving teachers toward the more ambitious instruction called for in the literature.

Teaching in the 50th MQI percentile. Table 6 shows segment-level MQI frequencies for teachers at the 50th percentile.

Table 6 here

This table shows that the depiction of teachers at the 50th percentile differs little from the depiction of the dataset on average. Given that we selected teachers at the 50th percentile of the distribution, this is to be expected, as these teachers were purposively selected. Analyses of lesson summaries from teachers near the median revealed two general profiles. In one, a teacher's lesson featured few positive or negative elements; 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. Tasks were not enacted in a way that challenged students to think 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 25 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. Often, such lessons consisted of a

problem or activity presented by the teacher, with students contributing only occasional words or phrases.

The second profile at the 50th percentile combined strong and weak features. In some summaries, raters noted one or two richness elements—meaning-making around a specific problem, occasionally developed from real-life tasks; multiple procedures for solving these problems; and precise mathematical language. Less common among these lesson summaries, however, were notes describing extensive student participation in building mathematical ideas. In addition, observers also noted mathematical errors that may reveal a lack of teacher mathematical knowledge—for instance, one teacher’s brief comment that trapezoids are “irregular” and therefore do not have a formula for area. 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 in the example above).

Robert. From these summaries, we selected one fourth-grade teacher, Robert, to illustrate this profile. Robert taught in a school where roughly half the students received free or reduced-price lunch; in each year of data collection, a handful of students were categorized as English language learners in his classroom. At the time data collection began, Robert had over 10 years of teaching experience. His mathematical knowledge for teaching score placed him almost exactly at the study average.

Some of Robert’s lessons were relatively strong, particularly in their reference to the meaning of mathematical quantities and operations. In one, for instance, Robert placed “ $1/2$ ”, “ $2/3$,” and “ $3/4$ ” on a number line, each time dividing the space between 0 and 1 into the number of equal segments indicated by the denominator, counting off the number of segments indicated by the numerator, and marking the endpoint of that counting as the location of the fraction. This

work emphasized the meaning of both numerator and denominator, and reinforced the notion of fractions as numeric quantities (Wu, 2009). Then, Robert presented a more unfamiliar fraction, $\frac{8}{12}$, and asked students to place it on the number line:

Robert: What could I do? What could I do to eight-twelfths to make it a little bit easier?

Student: Break it up into six groups?

Robert: Okay, yep.

Student: And then you take every sixth group, and cut it in half.

Robert: Okay, that's a good strategy. What's another strategy we could do?

As his request for a second strategy (ultimately, simplifying $\frac{8}{12}$ to $\frac{2}{3}$) indicates, Robert was looking for a specific approach; however, he recognized the student's method for partitioning the number line as both viable and sensible before moving on. Later in the lesson, he also took advantage of a benchmark number for placing $\frac{7}{10}$ on the line:

Robert: Let's think about [placing $\frac{7}{10}$ on the line].

Student: Three fourths.

Robert: No, seven tenths *is* close to three fourths. But. *Seven tenths*. Tenths is a pretty friendly number, right?

Student: Yeah.

Robert: So where would—who could come up and show me where five tenths is? Where's five tenths going to be?

A student walked to the board and placed $5/10$ on the number line.

Robert: Nice. We know that five tenths—we know that five tenths is equal to a half, right? So you know that if we—we don't have to break it up into ten sections. We already found the halfway point. So this is five. Where do you think seven's going to be? So all you have to do is, yeah. You just do it with your hand. Six. Seven. Eight. Nine. Ten. Right around here. Maybe a little bit farther over, but you know, it depends on your line. But that's in the right ballpark. So when you have a fraction like seven tenths, think about what the halfway point would be. Think about five tenths, okay?

Robert thus recommended that students work strategically with a benchmark fraction to place a more difficult fraction on the number line. Although the format was direct instruction, his presentation referenced the size of the quantity $7/10$, and it allowed students to see his reasoning process. This attention to meaning within traditional instructional formats characterized some of the lessons we observed Robert teach.

Other lessons were marked by more problematic features, including the proceduralization of complex tasks and ambiguity in Robert's mathematical language; 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 12×12) 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. 2×1 , 2×2 , 2×3 , etc.) and finding the largest product that could fit into the dividend (in this case, 2×8):

Robert: If we have to do 17 divided by two, 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 two go into 17 without going over?

Student: Eight.

In this first moment of instruction, the meaning of division—breaking a quantity into equal-sized groups—was not broached. Instead, an answer was found via a procedure: going down the multiples of two until 17 had been reached but not passed. Robert continued by adding a context to the problem:

Robert: Eight. So, let's talk about one of the division strategies we were doing the other day. So we have 17 divided by two. 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: Nine.

Multiple Students: Eight.

Here, Robert used the partitive definition for division to provide some meaning for the problem he had given; a number of pencils partitioned into two groups for two children. However, he did not specify that each student must receive an equal amount, leading to confusion in the passage above and in the next few minutes of instruction:

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

Student J: Um.

Robert: How many can one person get?

Student J: Eight.

Robert: Eight. And how many pencils are left over?

Student J: Eight.

Robert: No. How many pencils are left over?

At this point, students talked for a moment; many appeared confused about whether Robert was asking about how many were left after one student received his share, or how many were left after the two students received their share. This confusion was only resolved after Robert again restated the question and a student finally provided the answer Robert was looking for: one.

The exchange over this problem continued until it reached a total length of eight minutes of instructional time. Throughout, Robert's questions to students appeared to cause student confusion. Thus his efforts to attach meaning to this division problem—a feature that mathematics reformers desire more of in classrooms—were not enacted cleanly. He never explicitly stated the connection between his hundreds-chart illustration and the mathematical principle (dividing a set into two equal groups), he omitted a key condition (equal groups) throughout his discussion, and, in both this and the additional examples he provided to students, he proceduralized the connection between multiplication and division by specifying a set of steps students must apply to the hundreds chart. Further, this lesson excerpt highlights another salient

feature of Robert's lessons: he did most of the talking. Students contributed answers to computational questions ("Eight." "Nine."), but said little else.

What we observed in Robert's lessons was common among other teachers at or near the 50th percentile in our data. Seldom did we see teachers present straightforward, traditional instruction, in which an idea or procedure is delivered and students then practice deploying that new knowledge. Instead, we found that teachers worked with more complex tasks, and we saw corresponding instances of mathematical richness or student participation in the lesson. But in many cases, these elements occurred only sporadically and without the level of detail and precision found in 90th-percentile classrooms like Thomasina's. In many cases, teachers also failed to leverage potentially meaningful tasks to successfully develop strong disciplinary insight or encourage students into a more active classroom role. In addition, lessons in this category often featured mathematical imprecisions; while most of these errors did not degrade the quality of the mathematics, they occasionally caused student confusion. This illustration of instruction provides insight for those aiming to improve typical instruction in many schools and classrooms.

Teaching in the 10th MQI percentile. Table 7 shows segment-level MQI frequencies for teachers in the 10th percentile.

Table 7 here

Instruction at this rank in our distribution still included some Richness and Common Core-Aligned Student Practice elements, though less frequently than did instruction from teachers at the 50th percentile. Segments were also slightly more likely to feature major mathematical errors, imprecision in the use of mathematical language, and a lack of clarity in their presentation. Our

analysis of summaries describing lessons from teachers at or near the 10th percentile also revealed further differences. First, students generally listened to or watched their teachers demonstrate how to solve problems without much input, similar to the way Robert conveyed the division lesson. However, unlike Robert's lesson and the lessons of 50th percentile teachers generally, the mathematics in these classrooms had only fleeting moments of richness. Further, teachers rarely sought to engage students in rigorous mathematical thinking, even for the short amounts of time seen in 50th percentile teachers. Finally, teachers in the 10th percentile made mathematical errors that frequently obscured the mathematics of the lesson.

In addition to the shared features of teachers at the 10th percentile, qualitative analyses revealed two distinct lesson profiles that further displayed concerning patterns. In one, 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., making an art piece using four colors on a 10x10 grid with no discernable mathematical content). In the other profile, lesson activities did connect to mathematics, but in ways that distorted the mathematics students were to learn.

Arlene. To illustrate these themes, we selected one teacher, Arlene, a 10-year veteran who worked in the same district as Thomasina. At Arlene's school, less than half of the students received free or reduced-price lunch, and few students were classified as English language learners. In her work in another district, Arlene had served on district math committees, as a peer mentor or coach in the area of mathematics, and also taught 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 year she reported a more diverse set of supports, including

NSF-funded curricula. Arlene's mathematical knowledge for teaching was a half-standard deviation below the study average.

Arlene's lessons were remarkably consistent both within and across years. She typically divided students into small groups for instruction, and presented a lesson to four to five individuals while other students worked elsewhere in the room. These other students served as a frequent source of distraction, either because they were off task and required redirection or because they interrupted Arlene's small-group lesson to ask clarifying questions, to note that materials were not available, or to ask permission to leave the classroom. Most often, Arlene repeated identical mini-lessons to each small group, essentially teaching the material several times within a 50- to 60-minute time span. Student participation in these short lessons was mostly confined to taking notes 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 by asking "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 students had finished, Arlene continued:

Arlene: So show me with your hands what a vertical line would look like.
[*Students hold arms akimbo.*] 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. [*To a student from another group*] I told you to get something else, to do one of the other activities. [*To the group receiving instruction*] So, here's my window.

Arlene and her 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 additional interruptions and a review of different kinds of graphs later, she drew students' attention back to the Y- (vertical) axis:

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

Multiple Students: Small to large.

Arlene: Small to large. So they're counting up, right?

Multiple Students: Yes.

Arlene: 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, Arlene 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 Y-axes 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 another set of students.

Mathematically, this instruction appeared quite superficial. There was 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 given why an interval might be five in one graph and 25 in another. In fact, the content of the mini-lesson primarily focused on the terms "scale" and "interval," and there was no evidence that the tasks students completed after leaving Arlene's center deepened the mathematics. This thinness was characteristic across Arlene's lessons. Also consistent with her other lessons, Arlene presented all the information to students, asking them only the most bounded questions (e.g., "Small to large. So they're counting up,

right?”). Finally, Arlene took no advantage of the small-group format to customize instruction or allow students more generous opportunities to participate.

Also striking in Arlene’s instruction were frequent mathematical imprecisions and a lack of connection between lesson activities and larger mathematical ideas. These themes are illustrated in a lesson that used a picture of a pan balance and shapes representing different weights (Figure 3). In her introduction to the material with the first small group, Arlene launched the activity:

Arlene: 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....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 contains numerous imprecisions. For instance, numbers and letters were said to represent each other, rather than letters standing for an unknown number. In addition, Arlene’s reference to “balance something out” was not specific enough to convey the actual activity or why the activity would be of mathematical importance. She continued:

Arlene: A mobile has a total weight of 30, circles weigh five, squares weigh 10, and triangles weigh three. So the whole thing itself is going to be level at 30.... [*Repeats weight amounts*]. [The mobile] only has two shapes. If only circles are on one side, can it be balanced?

Arlene's switch between "total weight of 30" and "level at 30" led some students to think that the entire weight should be 60 for the few minutes of instruction that followed.

Figure 3 here

As stated, this problem has the potential to be cognitively demanding for students, who must leverage the ideas of equivalence and common multiples (the left- and right-side quantities must have 15 as a multiple). However, Arlene frequently devolved the cognitive demand of the task through heavy hinting. Rather than giving students time to work on a solution to the first problem or consider the algebraic idea of balance, she told students:

Arlene: My circles weigh five. 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 three Post-it Notes of circles 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?

Here, she directed students toward the solution with little time or opportunity for them to make sense of the idea of balance or equivalence. After another diversion regarding whether the total weight would be 30 or 60, she continued to move students toward the answer:

Arlene: Can I use my square or my triangle to equal out 15?

Student E: Yes, triangle.

Arlene: Alright, tell me, how many triangles?

Student E: Five.

Arlene: All right, Student E says five triangles...[*Another diversion about the total weight on the balance*].

Arlene: ...Okay, so let's see if it's equal now. Okay, so I have 15 on this side, three, six, nine, 12, 15 on this side. Does the whole thing weigh 30?

Multiple Students: Yes.

Arlene: Is the whole thing balanced?

Multiple Students: Yes.

This pattern of teacher commentary followed by one- or two-word student responses continued for subsequent problems. At the end of each problem, she instructed students to write an explanation for their answer in their math journals; in most cases, however, Arlene had already given these explanations verbally in solving the problem with the group.

As the lesson progressed, a larger problem emerged. This task 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 the value of an unknown (in this case, the number and weight of the blocks per side) must maintain balance. However, in the task's enactment, Arlene repeatedly stated that the *shapes themselves* were the variable—that the circle, for instance, was standing for an unknown. In fact, in this lesson the circle stands for a fixed quantity. This imprecision substantially obscured the mathematical point of the lesson; although Arlene repeatedly told students that the objective was to use balances to demonstrate how to find a variable, she never explicitly named the number of shapes as the variable in question. Together with Arlene's step-by-step structuring of the task, this lack of clarity rendered the lesson at best unconnected to, and at worst misinformative about, important mathematical ideas.

Thus, Arlene unifies both threads within our analysis of 10th-percentile videos: little student participation or richness together with thin and often confusing mathematics and frequent distractions. Although the algebra lesson she taught was likely inspired by NCTM and similar reforms,⁷ imprecisions and lack of clarity obscured the mathematical substance. Further, the small-group instruction was not put to use in any significant way, as Arlene taught and retaught the same material to several groups over the course of a lesson without adjustment to student needs or enhanced student input.

Conclusion

This study is limited in its conclusion in several regards. Our sample includes only five urban districts, and thus is not generally representative of districts or teachers across the United States. Teachers responded modestly to our invitation to join the study; although these teachers looked similar to non-participating teachers demographically and in terms of value-added scores, they may have participated because of their comfort with and competence in mathematics instruction. Teachers chose their own lesson, which might also inflate the estimate of reform practices in these urban districts. Finally, we collected and analyzed a relatively small number of lessons per teacher, leaving open the possibility that we missed or mischaracterized important features of instruction.

Nevertheless, we believe this study holds lessons for the promise of reform. To start, our analysis showed that many classrooms devote at least some attention to mathematical concepts. Instruction at both the 50th and 90th percentiles contained features such as providing explanations for mathematical facts and procedures, the use of multiple solution methods, and encouraging sense-making around problems. However, on average, these features were often enacted either briefly or without the level of detail and precision called for in the research literature. Students

did participate in mathematical reasoning and accessed cognitively demanding tasks, but only in a very small percentage of classrooms was the mathematics developed by students, for instance through investigation or discussion. More often these cognitively challenging elements appeared as brief moments in teacher-led lessons.

These findings also suggest that unlike other pictures of contemporary classrooms in which reform-oriented instruction rarely occurred (e.g. Kane & Staiger, 2012), reform practices do appear to a variable degree in study classrooms. By separately reporting attention to concepts and student cognitive demand, we see that attention to mathematical meaning and practices appears to be more frequently implemented than the enactment of challenging cognitive tasks. Future reports should continue to track separately on these two aspects of reform vision.

In addition, our analyses suggest that implementation of these reform practices corresponds to the strength and coherence of district efforts to change instruction. Although we lack both information on initial mathematics instructional quality and evidence that would help rule out competing explanations for this phenomenon, this notion accords with recent literature. For instance, Stein and Coburn (2008) compared two urban districts' efforts to implement new, ambitious math 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.

The strengths and weaknesses of the instruction in this sample provide important considerations for those interested in continuing to advance mathematics instruction. One immediate implication is that a great deal of this instruction could be improved. That so many

teachers attempted so many of the reform practices suggests that teachers' will does not stand in the way of such improvements; in both observed lessons from our data and in national surveys (Banilower et al., 2013), teachers appear to embrace reformers' ideals. Furthermore, in our sample, teachers used the reform-based curricula provided to them by their districts. What seems to us to be missing is technical skill—the commonplace but often detail-laden competencies that enable, for instance, teachers to correctly and completely present a mathematical explanation, foster connections between two representations or ideas, or launch a task in a way that does not devolve its cognitive demand. Some in teacher education are beginning to investigate ways to foster teachers' skills in these practices (e.g., Ball & Forzani, 2009; Grossman, Hammerness, & McDonald, 2009; Lampert, Beasley, Ghouseini, Kazemi, & Franke, 2010). Improvement efforts for in-service teachers might consider similar ideas.

The instruction we observed at the tenth percentile, on the other hand, shows the challenges efforts at mathematics reforms face in helping all teachers meet the NCTM and Common Core standards. Some of the instruction in our sample was consistently poor, consonant with past concerns about the role of mathematical precision in prior reforms (Wu, 1997) as well as past evidence from large-scale observational studies (e.g. Weiss et al., 2003). It would likely take intensive intervention for this instruction to reach even an adequate level. Our case analysis uncovered further patterns in the instruction at this level. Although the frequency of MQI Richness and Common Core Student Practice elements looked similar across the 50th and 10th percentile teachers, lessons among the latter group were extremely slow-paced, did not address mathematical ideas in depth, and lacked connections between mathematical activities and larger mathematical ideas. For instance, while we believe it likely that Arlene, our case study teacher representing lessons at the 10th percentile, intended her lessons to meet NCTM and Common

Core standards, her instruction would have to become more mathematically precise and connected to big ideas in order to provide students significant opportunities to learn rigorous content. Patterns at the 50th percentile similarly indicate possible focus for improvement efforts.

More broadly and somewhat provocatively, another implication of this study is that cognitively challenging, mathematically rich instruction need not only occur in classrooms devoted to student discussion or extended investigations of complex tasks. Teachers in the 90th percentile rarely engaged in this instructional format, but did manage to integrate student participation and mathematical practices into more familiar, traditional instructional formats. Thomasina, for instance, required students to solve challenging but “small-sized” problems – modeling short situations using numbers and a variable. Teachers may see his hybrid approach as a better fit with the reality of often-conventional state assessments and pacing guides that demand a quick progression through material. This suggests, in turn, that teachers may be more amenable to reforms that build incrementally upon existing practice, rather than require a wholesale revision of instruction.

Finally, we comment on future research directions. Our district sample is unlikely to be similar to the average in the United States, and our findings are thus non-generalizable to instruction more broadly. Yet there is a pressing need to assess the extent to which standards-based instruction occurs across the United States, not only because many believe that such instruction benefits students, but also because we are 30 years into a major reform, and it is time to take stock. Nationally-representative and longitudinal samples, beginning in the near future, would help us understand these issues in more systematic ways.

Endnotes:

1. In District E, we received only information about study schools.
2. The smaller number of teachers in the 90th percentile is due to tied scores when creating the percentile ranks. There were 329 teachers from which the ranks were created – which should result in about roughly 3 teachers per rank category. However, teachers with the same MQI score were placed in the same percentile rank. Because more teachers had lower MQI scores, there were more ties in the lower ranges – resulting in bins of teachers at the lower percentiles greater than 3. At the higher ranges, there were fewer ties.
3. Some identifying details of districts and teachers have been changed to protect their anonymity.
4. We note here that a large emphasis on reforming instruction in mathematics has focused on moving away from direct instruction toward a more student-centered classroom. We are not claiming that this should not be an emphasis, rather we note the importance of looking at instructional features even in the context of direct instruction as we will the degree to which students are provided opportunities to engage in practices such as meaning making.
5. This may make sense given the interaction between lesson content and the MQI richness categories. Some lessons are likely to lend themselves to multiple solution methods and others not; the same can be said for mathematical explanations, linking, and generalizations.
6. In only a small percentage of lessons at this percentile did such inquiry-oriented activities occur (e.g. generating definitions for polygons and geometric solids by exploring and categorizing their characteristics).

7. Many reformers have argued for the inclusion of algebra in the elementary curriculum.

Often, these lessons include an emphasis on equivalence and algebraic reasoning, similar to Arlene's.

References

- Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511.
- Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 National Survey of Science and Mathematics Education*. Chapel Hill, NC: Horizon Research.
- Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17*(2–3), 62–87.
- Blazar, D., Litke, E., & Barmore, J. (2016). What does it mean to be ranked a “high” or “low” value-added teacher? Observing differences in instructional quality across districts. *American Educational Research Journal, 53*(2), 324–359.
- Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers College Record, 110*(3), 608–645.
- Boston, M. D., & Wilhelm, A. G. (2015). Middle school mathematics instruction in instructionally focused urban districts. *Urban Education, 1*–33.
- Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York: Macmillan.
- Diamond, J. B. (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.
- Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist, 23*(2), 167–180.

- Good, T. L., & Grouws, D. A. (1977). Teaching effects: A process-product study in fourth grade mathematics classrooms. *Journal of Teacher Education, 28*(3), 49–54.
- Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice, 15*(2), 273–289.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte, NC: Information Age.
- Hiebert, J., Stigler, J., Jacobs, J., Givvin, K., 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.
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., ... & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher, 25*(4), 12–21.
- Hill, H. C., Blazar, D., & Lynch, K. (2015). Resources for teaching. *AERA Open, 1*(4), 1–23.
- 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., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher, 41*(2), 56–64.
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*(3), 794–831.

- Hill, H.C., Schilling, S.G., & Ball, D.L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, 105(1), 11–30.
- Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel* (MET Project Research Paper). Seattle, WA: Bill & Melinda Gates Foundation.
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (MET Project Research Paper). Seattle, WA: Bill & Melinda Gates Foundation.
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- Lampert, M., Beasley, H., Ghouseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. In M. K. Stein & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 129–141). New York: Springer.
- Lampert, M., Rittenhouse, P., & Crumbaugh, C. (1996). Agreeing to disagree: Developing sociable mathematical discourse. In D. R. Olson & N. Torrance, N. (Eds.), *The handbook of education and human development: New models of learning, teaching and schooling* (pp. 731–764). Cambridge: Blackwell.
- Loveless, T. (2001). The tale of two math reforms: The politics of the new math and NCTM standards. In T. Loveless (Ed.), *The great curriculum debate: How would we teach reading and math?* (pp. 184–209). Washington, DC: Brookings.
- Milner, H. R. (2013). Analyzing poverty, learning, and teaching through a critical race theory lens. *Review of Research in Education*, 37(1), 1–53.

- Nathan, M. J., & Knuth, E. J. (2003). A study of whole classroom mathematical discourse and teacher change. *Cognition and Instruction, 21*(2), 175–207.
- 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 Council of Teachers of Mathematics (NCTM). (2014). *Principles to actions: Ensuring mathematical success for all*. 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.
- O'Connor, M. C. (2001). "Can any fraction be turned into a decimal?" A case study of a mathematical group discussion. *Educational Studies in Mathematics, 46*(1–3), 143–185.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574.
- Saxe, G. B., Gearhart, M., & Nasir, N. S. (2001). Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education, 4*(1), 55–79.
- 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.

- Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313–340.
- Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455–488.
- Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation, 2*(1), 50–80.
- Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education, 62*(4), 339–355.
- Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247–280.
- Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research, 78*(3), 516–551.
- 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*. Chapel Hill, NC: Horizon Research Associates.
- Wilson, S. M. (2008). *California dreaming: Reforming mathematics education*. New Haven: Yale University Press.

Wu, H. (1997). The mathematics education reform: Why you should be concerned and what you can do. *The American Mathematical Monthly*, 104(10), 946–954.

Wu, H. (2009). What's sophisticated about elementary mathematics? *American Educator*, 33(3), 4–14.

Table 1: Student Characteristics in Study and Non-Study Schools, by District

	District A		District B		District C		District D	
	Study	Non Study						
N Schools	16	69	9	2	16	69	13	66
% Male	52.00	52.24	51.92	53.13	42.20	41.02	51.32	50.72
% White	7.22	15.31	26.13	23.61	5.46	7.61	27.47	31.31
% African American	41.43	35.35	51.88	56.99	78.78	76.56	32.86	29.75
% Hispanic	34.56	39.80	14.97	12.35	13.91	13.20	27.84	24.39
% Asian	11.96	4.05	2.96	3.12	1.36	1.68	7.90	10.16
% Other race/ethnicity	4.83	5.49	4.06	3.92	0.49	0.95	3.93	4.39
% FRPL-eligible	84.20	78.40	76.47	80.26	78.16	75.93	58.94	54.15
% SPED	21.57	26.30	16.95	10.88	12.64	9.52	13.35	13.99
% ELL	36.72	29.35	25.48	31.86	5.89	6.59	17.27	16.01
Prior state mathematics achievement	-0.05	0.04	-0.08	0.12	-0.08	0.03	-0.02	0.02

NOTE: Comparisons to non-study students not available for District E.

Table 2

*MQI Dimensions and Elements***Instructional Formats**

This dimension captures the pedagogical features of classroom work.

- *Direct instruction*: Teacher is in control of delivery of mathematical content. Mathematical content can be correct or incorrect; major feature is high amount of teacher talk and/or control relative to other activities (e.g., student talking, practice, etc.). (Score points: None, Some, All)
- *Whole-class discussion*: The teacher is in charge of the class, just as in direct instruction. However, the teacher is not primarily engaged in delivering information. Rather, he or she has students share their thinking, explain the steps in their reasoning, and build on one another's contributions. Key feature is that students comment on mathematics of one another's contributions. (Score points: None, Some, All)
- *Applied problems*: Teacher and/or students work on applied problems—for instance, figuring out among four recipes the proportion of orange juice and water that makes a mixture “more orangey.” (Score points: None, Some, All)
- *Classroom work connected to mathematics*: Focus is on mathematics for a majority of the segment rather than other topics (e.g., passing out materials, behavior management, off-topic discussion). (Score points: Not Connected to Mathematics, Connected to Mathematics)

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 elements:

- *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.

(Score points: Low [element not present], Mid [brief or pro forma enactment], High [substantive and more detailed enactment])

Common Core-Aligned Student Practices

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

-
- *Providing mathematical 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.*

(Score points: Low [element not present], Mid [brief or pro forma enactment], High [substantive and more detailed enactment])

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 elements:

- *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.

(Score points: Low [element not present], Mid [brief or pro forma enactment], High [substantive and more detailed enactment])

Table 3

Percentage of Segments at Each Score Point on Instructional Format

	Score Point		
	None	Some	Most/All
Direct Instruction	7	32	61
Whole Class Discussion	95	4	1
Working on Applied Problems	72	10	18

Table 4

Percentage of Segments at Each Score Level on MQI Elements

	Score Point		
	Low	Mid	High
RICHNESS			
Linking and Connections	71	24	5
Explanations	73	25	2
Multiple Methods	85	13	1
Generalizations	96	3	1
Mathematical Language	58	37	5
COMMON CORE-ALIGNED STUDENT PRACTICES			
Student Explanations	80	18	1
Student Math Questioning and Reasoning	81	16	2
Enacted Task Cognitive Activation	69	27	3
ERRORS			
Major Errors	94	5	1
Language Imprecision	84	15	1
Lack of Clarity	90	9	2

Table 5: Percentage of Segments at Each Score Level on MQI Elements for Teachers at the 90th Percentile (n = 8 teachers)

	Score Point		
	Low	Mid	High
RICHNESS			
Linking and Connections	46	36	18
Explanations	55	41	4
Multiple Methods	81	18	2
Generalizations	95	4	1
Mathematical Language	37	49	14
COMMON CORE-ALIGNED STUDENT PRACTICES			
Student Explanations	62	34	4
Student Math Questioning and Reasoning	68	24	8
Enacted Task Cognitive Activation	42	50	8
ERRORS			
Major Errors	93	7	0
Language Imprecision	91	9	0
Lack of Clarity	97	3	0

Table 6: Percentage of Segments at Each Score Level on MQI Elements for Teachers at the 50th Percentile (n =16 teachers)

	Score Point		
	Low	Mid	High
RICHNESS			
Linking and Connections	73	24	3
Explanations	76	23	1
Multiple Methods	90	10	1
Generalizations	95	4	1
Mathematical Language	63	33	3
COMMON CORE STUDENT PRACTICES			
Student Explanations	84	15	1
Student Math Questioning and Reasoning	81	16	3
Enacted Task Cognitive Activation	74	23	2
ERRORS			
Major Errors	92	6	1
Language Imprecision	79	18	2
Lack of Clarity	91	9	1

Table 7: Percentage of Segments at Each Score Level on MQI Elements for Teachers at the 10th Percentile (n = 15 teachers)

	Score Point		
	Low	Mid	High
RICHNESS			
Linking and Connections	79	19	2
Explanations	77	19	4
Multiple Methods	92	7	1
Generalizations	98	2	1
Mathematical Language	64	32	4
COMMON CORE STUDENT PRACTICES			
Student Explanations	86	12	2
Student Math Questioning and Reasoning	82	16	2
Enacted Task Cognitive Activation	81	18	2
ERRORS			
Major Errors	91	7	2
Language Imprecision	80	19	2
Lack of Clarity	88	11	1

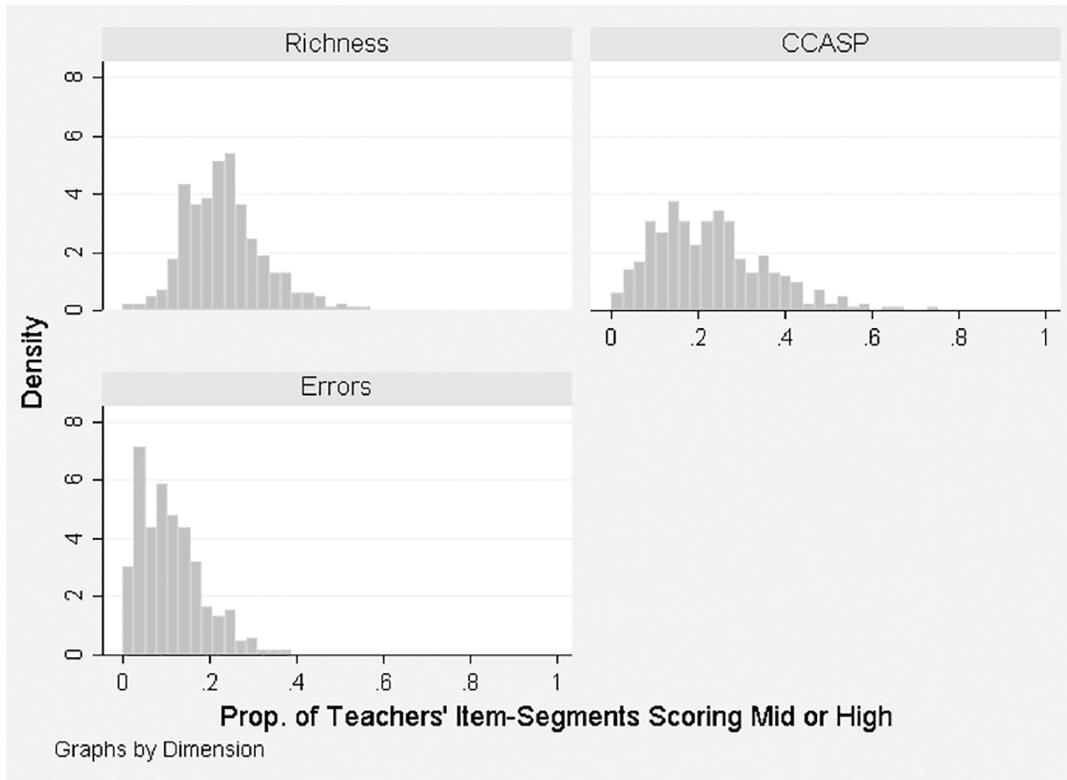


Figure 1: Teacher-level proportions of segments rated as mid or high for at least one item within the dimension

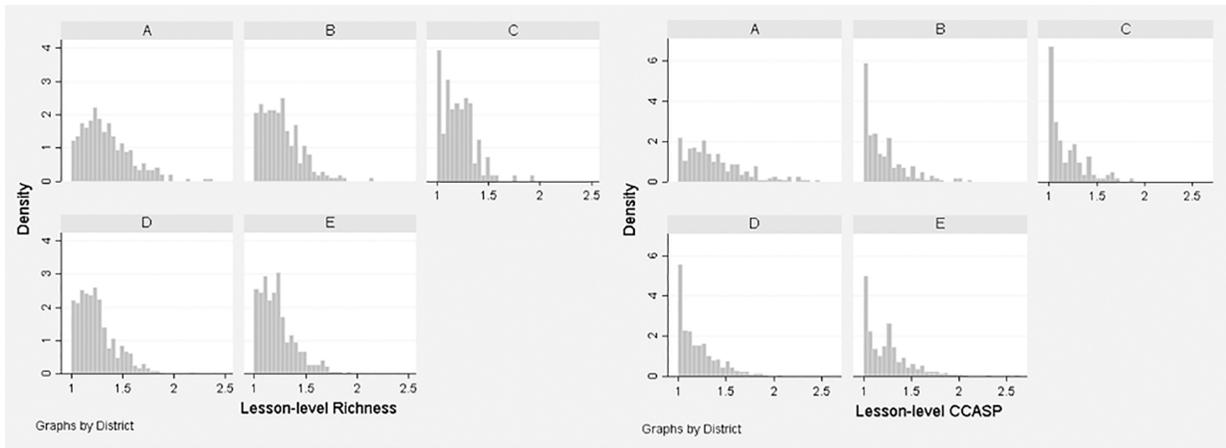


Figure 2. District differences in the MQI Richness and Common Core Student Practices measures



Figure 3. Arlene's pan balance and weights.