Exploring Relationships Between Setting Up Complex Tasks and Opportunities to Learn in Concluding Whole-Class Discussions in Middle-Grades Mathematics Instruction

Author(s): Kara Jackson, Anne Garrison, Jonee Wilson, Lynsey Gibbons and Emily Shahan

Source: *Journal for Research in Mathematics Education*, Vol. 44, No. 4 (July 2013), pp. 646-682
Published by: National Council of Teachers of Mathematics
Stable URL: http://www.jstor.org/stable/10.5951/jresematheduc.44.4.0646
Accessed: 08/09/2013 17:39

Journal for Research in Mathematics Education
2013, Vol. 44, No. 4, 646–682

Exploring Relationships Between Setting Up Complex Tasks and Opportunities to Learn in Concluding Whole-Class Discussions in Middle-Grades Mathematics Instruction

Kara Jackson
McGill University

Anne Garrison and Jonee Wilson
Vanderbilt University

Lynsey Gibbons
University of Washington

Emily Shahan
Vanderbilt University

This article specifies how the setup, or introduction, of cognitively demanding tasks is a crucial phase of middle-grades mathematics instruction. We report on an empirical study of 165 middle-grades mathematics teachers' instruction that focused on how they introduced tasks and the relationship between how they introduced tasks and the nature of students' opportunities to learn mathematics in the concluding whole-class discussion. Findings suggest that in lessons in which (a) the setup supported students to develop common language to describe contextual features and mathematical relationships specific to the task and (b) the cognitive demand of the task was maintained in the setup, concluding whole-class discussions were characterized by higher quality opportunities to learn.

Key words: Cognitive demand; Instruction; Opportunity to learn; Tasks

Over the past several decades, mathematics education researchers have achieved broad consensus regarding a set of goals for students' mathematical learning, which are represented in documents like the National Council of Teachers of Mathematics' *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics, 2000), the National Research Council's *Adding It Up: Helping Children Learn Mathematics* (Kilpatrick, Swafford, & Findell, 2001), and the more recent *Common Core State Standards* for mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). These documents describe a concrete set of learning goals that encompass both conceptual understanding and procedural fluency in a range of mathematical domains as well as the development of productive problem-solving capabilities and dispositions.

At the same time, mathematics education researchers have worked to specify what should happen between teachers and students in classrooms in order to accomplish these goals for students' learning (Franke, Kazemi, & Battey, 2007). For example, research suggests that in order to accomplish such learning goals, instruction should include frequent opportunities for students to solve challenging mathematical tasks, to articulate their mathematical reasoning, and to make connections between mathematical ideas and representations (Franke et al., 2007; Hiebert et al., 1997). This kind of instruction has been called *ambitious mathematics teaching* (Lampert, Beasley, Ghousseini, Kazemi, & Franke, 2010) because of the knowledge and skill involved in supporting each student to develop an increasingly sophisticated understanding of central mathematical ideas; it necessarily requires that teachers teach in response to what students do as they engage in solving mathematical tasks (Kazemi, Franke, & Lampert, 2009; Lampert & Graziani, 2009). Ambitious mathematics teaching can take a variety of forms. One common lesson structure in reform-oriented middle-grades mathematics curricula (e.g., Connected Mathematics Project) is the three-phase lesson. A complex task is introduced, students work on solving the task, and the teacher orchestrates a concluding whole-class discussion (Van de Walle, Folk, Karp, & Bay-Williams, 2010).

Prior research has suggested that in an effective first phase of lessons with this structure, the teacher clarifies expectations for the final work product and how students should work (e.g., individually or in groups) to solve the task (Boaler & Staples, 2008; Smith, Bill, & Hughes, 2008). During the second phase of instruction, while students are solving the task either individually or in groups, the teacher circulates among the students, paying close attention to what students are doing as they complete the task so that he or she can decide what mathematical ideas and whose solutions to make a focal point during the whole-class discussion that follows (Lampert, 2001; Stein, Engle, Smith, & Hughes, 2008). The third phase of

---

An earlier version of this article was presented at the National Council of Teachers of Mathematics Research Presession, April 12, 2011, Indianapolis, IN. The empirical work reported in this article has been supported by the National Science Foundation (NSF) under grants DRL-0830029 and ESI-0554535. In addition, the National Academy of Education/Spencer Postdoctoral Fellowship Program supported Kara Jackson's contributions to the article. Anne Garrison's and Jonee Wilson's contributions to the article were supported by the Institute of Education Sciences (IES) predoctoral research training program, grant number R305B080025. The opinions expressed do not necessarily reflect the views of the National Academy of Education/Spencer Postdoctoral Fellowship Program, NSF, or IES. We would like to thank Chuck Munter for his useful comments on a previous version of this manuscript, and Melissa Boston, Paul Cobb, Glenn Colby, Robert Jiménez, Lindsay Clare Matsumura, Rich Milner, Rebecca Schmidt, and Thomas Smith for their contributions to this work. We thank Cynthia Langrall, the editor, and the editorial staff for their incredibly helpful guidance in revising the manuscript, as well as the five anonymous reviewers and statistical consultant for their thorough critiques and suggestions. Additionally, we thank the participating school districts, and especially the teachers, for allowing us to learn from their practice.

instruction refers to when the teacher orchestrates a concluding whole-class discussion aimed at developing all students' increasingly sophisticated understanding of the key mathematical ideas (Stein et al., 2008). In order to orchestrate a *productive* concluding whole-class discussion (i.e., one that advances the learning of all students), the teacher has identified and sequenced particular students' solutions to ensure that the discussion advances his or her instructional agenda (Stein et al., 2008). During the discussion, the teacher presses students to explain and justify their solutions, evaluate their peers' solutions, and make connections between different solutions (Ball & Bass, 2000; Chapin, O'Connor, & Anderson, 2003; Hiebert et al., 1997; Stein, Smith, Henningsen, & Silver, 2000). The teacher plays a crucial role in mediating the communication between students to help them understand each other's explanations (McClain, 2002) and in supporting students to link student-generated solution methods to disciplinary methods and important mathematical ideas (Stein et al., 2008).

A central goal of this article is to elaborate on the *how* of ambitious mathematics teaching by identifying *high-leverage practices* (Ball, Sleep, Boerst, & Bass, 2009) that teachers can develop. Such practices have the potential to increase student participation and learning as they engage in mathematical activity aimed at rigorous learning goals. Here, we focus on the first phase of a lesson, in particular the practice of setting up complex tasks so that all students are able to productively engage in solving the task and thereby participate in and learn from the third phase of the lesson, the concluding whole-class discussion.

Motivation for the Study

Our concern with specifying aspects of a high-quality setup initially arose from watching video recordings of the instruction of middle-grades mathematics teachers, all of whom were attempting to implement reform-oriented curricula. The video recordings were collected as part of a 4-year research project focused on middle-grades mathematics teaching and learning (Cobb & Jackson, 2011; Cobb & Smith, 2008). Our goal in watching the videos was to identify forms of practice that were not yet specified in the mathematics education research literature and that appeared important for supporting all students to substantially participate in and learn through solving complex tasks. As we watched the videos, we identified the first phase of instruction as important for two reasons.

First, we noticed that how the task was set up appeared to affect which students could participate in solving the task and in what ways. At the start of a lesson, it is not reasonable to expect that all students will have requisite or similar understanding of a given task statement. Some tasks, particularly those associated with middle-grades reform-oriented curricula, often embed problem solving in scenarios. Tasks that embed mathematics problem solving in a scenario pose a challenge; it is unlikely that all students will be equally familiar with the scenario described in a task statement (Ball, Goffney, & Bass, 2005; Boaler, 2002; Lubienski, 2000; Silver, Smith, & Nelson, 1995; Tate, 1995).

For example, consider the Dollars for Dancing task in Figure 1. In this task, students are asked to represent and interpret various ways of earning money in a dance marathon. Scenarios such as this one provide a context to support students' initial engagement in the problem and their mathematization (e.g., Gravemeijer & Doorman, 1999) of key ideas and relationships. In this case, the dance marathon scenario provides a context about which students can reason to develop meaning for linear relationships. However, if a student does not understand key aspects of the scenario, it is unlikely that he or she will engage in solving the task. And without substantially engaging in solving the task, it is unlikely that a student will be able to participate in or learn much from a concluding whole-class discussion.

Dollars for Dancing

Three students at a school are raising dollars for the school's Valentine's Dance. All three decide to raise their money by having a dance marathon in the cafeteria the week before the real dance. They will collect pledges for the number of hours that they dance, and then they will give the money to the student council to get a good DJ for the Valentine's Dance.

Rosalba's plan is to ask teachers to pledge $3 per hour that she dances. Nathan's plan is to ask teachers to give $5 plus $1 for every hour he dances. James's plan is to ask teachers to give $8 plus $0.50 for every hour he dances.

Part A. Create at least three different ways to show how to compare the amounts of money that the students can earn from their plans if they each get one teacher to pledge.

Part B. Explain how the hourly pledge amount is represented in each of your ways from Part A.

Part C. For each of your ways in Part A, explain how the fixed amount in Nathan's plan and in James's plans is represented.

Part D. For each of the ways in Part A, show how you could find the amount of money collected by each student if they could dance for 24 hours.

Part E. Who has the best plan? Justify your answer.

Figure 1. Dollars for Dancing. Modeled after Task 1.3, "Raising Money," in *Moving Straight Ahead: Linear Relationships* by Lappan, Fey, Fitzgerald, Friel, and Phillips (2009).

Second, we noticed that how a task was introduced appeared to affect the work of teachers in subsequent phases of the lesson. When students are not supported to understand key aspects of the task statement, teachers often spend the next phase of instruction reintroducing the task to individuals or groups of students while others begin to solve the task. This is not necessarily bad—it may be that some students need additional or different information about the task to begin to solve it in productive ways. However, reintroducing tasks for an extended period is a loss of valuable time. If a teacher spends the second phase of instruction reintroducing the task, he or she is unlikely to be able to carefully plan for a concluding whole-class discussion. This, then, makes it less likely that the teacher will support students to make important connections between student-generated solutions and key mathematical ideas.

Having identified that the setup appeared important for student learning, we engaged in a qualitative analysis of the video recordings of 40 middle-grades mathematics teachers' instruction, collected in Year 1 of the project. Our goal was to identify key aspects of setting up complex tasks in middle-grades mathematics lessons that were likely to support students' subsequent participation in instruction. The four key aspects that we identified pertained to contextual features, mathematical ideas and relationships, development of common language, and maintenance of the cognitive demand of the task. Each of these aspects is described in detail in the conceptual framework section, which follows.

We became interested in whether—and if so, how—the aspects we identified were related to students' opportunities to learn, specifically in concluding whole-class discussions. We therefore conducted a quantitative study of the video recordings from 165 middle-grades mathematics teachers in Year 3 and Year 4 of the project, in which we aimed to answer the following questions:

1. What is the nature of the setup phase of instruction? In particular:
   a. To what extent do teachers attend to contextual features and the mathematical relationships of a task statement?
   b. To what extent do teachers maintain the cognitive demand of the task during the setup, especially when they attend to the contextual features and the mathematical relationships of the task?
2. How is the quality of the setup related to students' opportunities to learn mathematics in the concluding whole-class discussion?

Conceptual Framework

Mathematical Tasks Framework

Our research builds on Stein and her colleagues' efforts to conceptualize key relationships between features of classroom instruction and students' opportunities to develop significant mathematical understanding (e.g., Stein, Grover, & Henningsen, 1996; Stein et al., 2000). A central aspect of instruction is the cognitive demand of the activity in which students participate. Cognitive demand refers to what students need to do (e.g., the nature of reasoning) in order to solve a particular problem or, at a broader level, participate in a given activity (Doyle, 1988). One determining factor of the cognitive demand of classroom activity, and thus the nature of students' learning opportunities, is the nature of the task that a teacher chooses to use in instruction, or the task as it appears in instructional or curricular materials (Stein & Lane, 1996). Stein, Grover, and Henningsen (1996) systematically identified characteristics of mathematics tasks with low and high cognitive demand. Tasks with low cognitive demand require students to memorize or reproduce facts or to perform relatively routine procedures without making connections to the underlying mathematical ideas. Tasks with high cognitive demand tend to be open-ended (i.e., a solution strategy is not immediately apparent); require students to make connections to the underlying mathematical ideas; and engage students in disciplinary activities of explanation, justification, and generalization. Based on analyses of middle-grades mathematics instruction aimed at ambitious learning goals, Stein and Lane (1996) found that the use of tasks with high cognitive demand was related to greater student gains on an assessment requiring high levels of mathematical thinking and reasoning.

However, as Stein et al. (1996) illustrated, selecting a task with high cognitive demand does not ensure that students will be provided opportunities to engage in rigorous mathematical activity. Instead, tasks have to be understood as part of classroom activity, and interactions between the teacher, students, and the task determine the extent to which the cognitive demand is maintained across the course of a lesson (Stein et al., 2000). The cognitive demand of a high-level task can be lowered if a teacher or student suggests a solution path before students begin to solve a problem or if a teacher alters the directions in the task such that students are no longer required to justify their thinking or solve the more challenging aspects of the problem.

Stein and colleagues identified two phases of instruction that were influential when considering the extent to which the cognitive demand of a task as it appeared in materials is maintained: the *setup phase* and the *implementation phase* (Stein & Lane, 1996). In Stein, Smith, Henningsen, and Silver's (2000) terms, the setup phase of instruction refers to how a teacher introduces the task, or "the teacher's communication to students regarding what they are expected to do, how they are expected to do it, and with what resources" (p. 25). All activity that occurs after the setup constitutes the implementation phase, in which students begin to work to solve the task. Therefore, it includes both what happens when students work on solving the task and any concluding whole-class discussion.

Stein et al. (1996) found that in classrooms where tasks with the potential for high levels of cognitive demand were assigned, teachers and students often decreased the cognitive demand over the course of the lesson. They identified three factors that appeared to influence whether the teacher and students maintained the challenge of the task through the setup: the teacher's goals for instruction, the teacher's subject matter knowledge, and the teacher's knowledge of students. For example, a task of high cognitive demand could be viewed by a teacher as supporting the development of procedural understanding of a particular skill because of her instructional goals or her knowledge of mathematics. Additionally, Stein et al. (1996) wrote that a major factor contributing to the decline of the cognitive demand of high-level tasks was "student failure to engage in high-level activities due to lack of interest, motivation, or prior knowledge" (p. 480). They argued that this stems from teachers' lack of knowledge of their students—teachers did not choose an appropriate task for their students.

Key Aspects of High-Quality Setups

Stein and colleagues identified the setup as an important phase of middle-grades mathematics instruction in lessons aimed at rigorous goals for students' learning. However, there has been little work done to identify what teachers might do during this phase of instruction to support student participation in solving tasks of high cognitive demand.[1] As stated above, in our qualitative study of 40 videos collected in Year 1 of the research project, we identified four key aspects of high-quality setups, which we conjectured would support student participation in solving complex tasks. The four aspects are as follows:

1. Key contextual features of the task scenario are explicitly discussed.
2. Key mathematical ideas and relationships, as represented in the task statement, are explicitly discussed.
3. Common language is developed to describe contextual features, mathematical ideas and relationships, and any other vocabulary central to the task statement that might be confusing or unfamiliar to students.
4. The cognitive demand of the task is maintained over the course of the setup.

We elaborate these four aspects here in order to illustrate what we investigated in teachers' instruction. The explanation of these aspects is grounded in an actual setup from a seventh-grade classroom that was video recorded in Year 1 of the research project. This setup is also described in Jackson, Shahan, Gibbons, and Cobb (2012).

The teacher, Mr. Smith,[2] introduced the Dollars for Dancing task (shown in Figure 1) about midway through the school year. In prior lessons, the students used tables, graphs, and equations to solve problems involving linear relationships with y-intercepts of zero. This lesson was their first encounter with a linear relationship with a non-0 y-intercept. Mr. Smith's goal was to leverage the dance marathon scenario to support students in developing understanding of the y-intercept as an initial value and its relationship to slope. Mr. Smith devoted 8 minutes of an hour-long lesson to setting up this task.

Key contextual features. As suggested above, the extent to which a student is familiar with a task scenario will impact whether the student can productively engage in solving the task. Thus, one key aspect of an effective setup is the explicit

[1] Franke, Kazemi, and colleagues (Franke, 2006; Kazemi et al., 2009) have identified the practice of "problem posing" as critical for supporting all students' learning in elementary mathematics teaching. Their work provides useful images of what high-quality problem posing, or setting up tasks, looks like in the context of elementary mathematics.

[2] Note that all names are pseudonyms and quotations are taken from transcripts of video recordings of classroom instruction.

discussion of the key contextual features of the task scenario. Key contextual features are aspects of the scenario that students would not understand unless they had prior experience with it. For example, key contextual features of the Dollars for Dancing scenario include knowing what a dance marathon might involve and why people organize and participate in dance marathons to raise funds.

To develop his students' understanding of the key contextual features in Dollars for Dancing, Mr. Smith elicited students' prior knowledge about dance marathons by projecting pictures of dance marathons from the Internet and asking students to discuss what they saw. For example, one student replied "dance," while another student said "dance marathon." He took advantage of students' contributions to develop an initial description of dance marathons as "groups of people who dance for a certain amount of time." Mr. Smith then pressed students to explain why people might hold a dance marathon. He built on several of their proposals to explain that the task they were going to solve involved holding a dance marathon to raise money in order to hire a deejay for the school's upcoming Valentine's Dance.

Key mathematical ideas and relationships. A second critical aspect of high-quality setups involves the explicit discussion of how the key mathematical ideas and relationships are represented in the task statement. In Thompson's (1996) and McClain and Cobb's (1998) terms, students' initial understanding of the mathematical relationships described in the task statement provide a basis for any mathematizations they might make as they attempt to solve a task. For example, in Dollars for Dancing, the students were expected to use tables, graphs, and equations to represent the accumulation of money over time in three different plans. In order to do so, it was essential that the students understand that money accumulates as a participant continues to dance for a greater number of hours. Furthermore, there are different ways of accumulating money—starting with a fixed amount and/or earning a fixed amount of money per hour of dancing. Absent *situation-specific imagery*, or an understanding of how these key mathematical relationships are represented in the statement, students' efforts to solve tasks typically become decoupled from their interpretations of problem situations (McClain & Cobb, 1998; Thompson, 1996). In the case of Dollars for Dancing, without an image of how money is accumulating over time, students are unlikely to make connections between accumulating quantities of money, slope, and *y*-intercept, as the task intends. It is also probable that some students will struggle to create appropriate tables, graphs, and equations themselves, or to understand their peers' representations.

Mr. Smith's setup illustrates this second aspect of high-quality setups. In addition to discussing key contextual features of the scenario, Mr. Smith also supported his students' understanding of how a key mathematical relationship (the accumulation of money) was represented in the statement. He began with the difference between an up-front amount and an hourly amount.

> There [are] two ways you can raise money in a dance marathon that we're going to talk about. One way is to dance for a long time.… So if you dance for a long time, and

> let's say I give you 50 cents every hour, you're going to make a lot of money. But there's another way that you could raise money, and that is to ask for a pledge. Not per hour, but just a donation. Okay, we call that a donation. And you might go up to your teacher and say, "Can you give me $6 for being in the dance marathon?" Now that's different. Can anybody explain how that is different if I say, "Can you give me $6?" or instead, "Can you give me 50 cents an hour?"

Mr. Smith called on a number of students to explain the distinction. For example, one student, Jasmine, responded, "Either they pay you up front or you continue, so like they continue to pay you for however long you dance." As students shared their ideas, Mr. Smith asked students to restate what others said (e.g., "Can you say what Jasmine said in your own words?"). Mr. Smith also praised students' ideas and adopted students' ways of describing the distinction. For example, another student offered, "Well, like one of them you already start with it and the other one you have to kind of work for it to get more." Mr. Smith responded, "Exactly. I like the way that's worded. One of them you start with it, you just have it. The other one you got to work for it to get the money." Once it was clear that the majority of students could explain the distinction between the two ways to accumulate money, Mr. Smith handed out the task sheet and briefly explained students' responsibilities for working in their small groups. The students then began to solve the task.

Development of common language. A third aspect of a high-quality setup relates to the nature of teacher and student talk. In the effective setups that we identified, teachers did not simply talk to students about the key features of the task. Instead, they solicited input from multiple students and asked questions that required more than a *yes* or *no* response. Broad and active student participation helps the teacher assess students' understanding of the key features of the task in order to determine the level of support that students may need (Boaler, 2002). Mr. Smith's setup is illustrative in this regard because he elicited students' understanding of the key contextual features and mathematical relationships in the scenario.

In addition, high-quality setups aim at developing common language to describe key contextual features and mathematical ideas and relationships. The use of common language is an indicator that students have developed *taken-as-shared* understanding (Cobb, Wood, Yackel, & McNeal, 1992; Cobb, Yackel, & Wood, 1992) of the key features of the task. Given that one cannot determine that any two people have the same understanding, taken-as-shared refers to understanding for which there is evidence that it can be taken as, or reasonably assumed to be, compatible enough to enable students to communicate in consistent ways about the relevant ideas. Prior research has identified how supporting the development of taken-as-shared understanding among students serves as a basis for communication during instructional activity, for example when communicating and representing mathematical ideas in small groups (Cobb, Wood, et al., 1992; McClain & Cobb, 1998) and when supporting students with less developed understanding of the mathematical ideas to participate in and benefit from a whole-class discussion (McClain & Cobb, 1998).

Visual representations can support the establishment of a taken-as-shared understanding of an important idea. For example, Mr. Smith projected images of dance marathons, a key contextual feature of the Dollars for Dancing scenario. However, it is crucial that students are provided opportunities to develop common language to describe the representation. As Moschkovich (1999) clarified, "Objects do not provide 'extra-linguistic clues.' The objects and their meanings are not separate from language, but rather acquire meaning through being talked about and these meanings are negotiated through talk" (p. 14).

Mr. Smith's setup illustrates the establishment of taken-as-shared understanding of key contextual features and mathematical relationships. In particular, he used several *talk moves* (Chapin et al., 2003) to support students in developing common language. For example, he revoiced or adopted students' language for describing aspects of the scenario, asked students to state or restate ideas in their own words, asked students to add on to their peers' ideas, and marked particular ideas as important. Asking students to revoice and add on to their peers' ideas also provided Mr. Smith with an informal assessment of the extent to which the students had developed compatible understandings of key features of the task.

Maintenance of the cognitive demand. The aspects of effective setups described above are aimed at ensuring that all students can engage productively in solving complex tasks while maintaining the cognitive demand of the task during the setup. Here, we refer to Stein et al.'s (2000) use of cognitive demand as the nature of reasoning students are expected to engage in to participate in the given activity (e.g., solving the task, discussing their solutions to the task in a whole-class discussion). As Sztajn, Confrey, Wilson, and Edgington (2012) clarified, the cognitive demand of a task as experienced by students is the relationship between what is expected in a task and individual students' present conceptions and their "informal and previous instructional experiences" (p. 150). Thus, a central part of setting up a task is identifying what counts as cognitively demanding—given the instructional goal(s) of a lesson or sequence of lessons, students' prior instructional experiences, and students' current understanding—and planning how to engage students in solving the task without compromising students' opportunities to learn significant mathematics. In our estimation, providing students with access to the key ideas of complex tasks while maintaining the cognitive demand of the task is delicate work.

For example, Mr. Smith knew his students were already comfortable making tables, graphs, and equations to represent linear equations with y-intercepts of 0. Representing equations with y-intercepts not equal to 0 was a central aspect of the challenge of this particular task. Mr. Smith could have reduced the cognitive demand of the Dollars for Dancing task by suggesting to the students how to solve the problem (e.g., show the students how to use a graph or equation to represent a linear relation with a non-0 y-intercept). Another challenging aspect of the task entailed comparing the merits of the particular plans using mathematical justification. Therefore, Mr. Smith would have reduced the cognitive demand of the task had he told students to skip parts D and E. Instead, he used the setup to ensure that all students developed the requisite understanding to reason about significant mathematical ideas and maintained the cognitive demand of the task.

Summary. Above, we described four key aspects of setting up complex tasks that we conjectured would support all students to participate substantially in classroom activity. It is worth clarifying two points. First, our division of *contextual features* from *key mathematical ideas and relationships* is artificial; the key mathematical ideas and relationships as represented in a task scenario are also contextual features of the scenario. We have chosen to separate key mathematical ideas and relationships from contextual features because they capture two foci of high-quality setups. A teacher can attend to what we term contextual features without attending to key mathematical ideas, and vice versa.

Second, throughout we use the language of *key* features and mathematical ideas and relationships. Scenarios associated with cognitively demanding tasks are often quite dense, both contextually and mathematically. The scenarios could lend themselves to extended talk about a number of contextual features, which may or may not be critical to solving the task, and a given task could be used to develop several mathematical ideas and relationships. Clearly, time is of the essence in classroom instruction; therefore, teachers need to make judgments regarding what to focus on in the setup. These judgments must be made against a clear set of mathematical goals for instruction and knowledge of what is likely to be unfamiliar (contextually, mathematically, and linguistically) to students.

Methods

In the remainder of this article, we report on an empirical study of 165 middle-grades mathematics teachers' instruction. In this study, we sought to understand the nature of teachers' setups, and to explore relationships between how tasks were set up and students' opportunities to learn mathematics in the concluding whole-class discussion.

Research Context

To answer our research questions, we used data that were collected in Year 3 (2009–2010) and Year 4 (2010–2011) of the 4-year project (2007–2011) designed to address the question of what it takes to improve the quality of middle-grades mathematics teaching, and thus student achievement, at the scale of large, urban districts in the United States. Each year, several types of data were collected to test and refine a set of hypotheses and conjectures about district and school organizational arrangements, social relations, and material resources that might support mathematics teachers' development of high-quality instructional practices at scale (Cobb & Jackson, 2011; Cobb & Smith, 2008). In each of the four districts, approximately 30 teachers and their instructional leaders (principals, assistant principals, and coaches) participated in the project. Data collected include interviews with all participants, video recordings of classroom instruction, assessments of teachers' and coaches' mathematical knowledge for teaching (Hill, Schilling, & Ball, 2004), video or audio recordings of professional development sessions, and student achievement data. For the purposes of this study, we focused on one form of data, video recordings of classroom instruction, which we explain in further detail below.

Participating districts. Each of the four districts was purposively invited to participate in the research project for a few reasons. On the one hand, the districts faced challenges typical of most large, urban districts—limited resources, high teacher turnover, large numbers of students identified as low performing in mathematics, and disparities between subpopulations of students' performance on state assessments (Darling-Hammond, 2007). On the other hand, the districts were chosen because their response to high-stakes accountability pressures to improve students' performance in middle-grades mathematics was atypical. Namely, each district was attempting to achieve a vision of instruction in middle-grades mathematics classrooms that is broadly compatible with the National Council of Teachers of Mathematics' (2000) Standards. Three of the four districts (which we will call Districts A, B, and D) adopted the Connected Mathematics Project 2 (CMP2) curriculum materials. In District C, mathematics specialists created a curriculum that was a blend of CMP2 and a more conventional mathematics text. Districts B, C, and D adopted their respective curricula in Year 1 of the project; prior to 2007−2008, mathematics teachers in those districts used a conventional mathematics text. District A adopted CMP2 in Year 2 of the project; however, District A teachers had been using the first edition of the Connected Mathematics Project materials for approximately 10 years prior to the adoption of CMP2. Additionally, each district supported teachers in developing ambitious forms of instructional practices (e.g., curriculum frameworks, coaching, regularly scheduled time to collaborate with colleagues on issues of instruction, professional development for instructional leaders).

Participating teachers. Six to ten schools in each district were selected to participate in the project. Schools were purposively sampled to reflect variation in student performance and in capacity for improvement within each district. Our sample consists of 165 teachers across the four districts, located in the selected schools: 34 teachers from District A, 48 teachers from District B, 35 teachers from District C, and 48 teachers from District D. Teachers in our sample averaged 9.5 years of experience teaching mathematics, with a significantly more experienced group of teachers averaging 13.3 years of experience in District A. The number of teachers with 3 or fewer years of experience teaching mathematics varied by district: 4 teachers in District A, 25 teachers in District B, 11 teachers in District C, and 27 teachers in District D.

Data Source: Video Recordings of Classroom Instruction

Two days of instruction of the same class were video recorded in January, February, or March for each participating teacher in each year of the study. When possible, videographers recorded 2 consecutive days of instruction. This was done in order to account for the fact that a lesson might extend over more than 1 day. Teachers were expected to teach the content that they would normally teach; however, we asked that teachers include a problem-solving activity and a related whole-class discussion in their instruction. This was compatible with the lesson structures suggested in all the districts' curricula. To be clear, the goal of the video recordings was not to capture the nature of teachers' everyday practice but rather to assess the quality and extent to which a teacher might enact the particular kind of instruction articulated by district leaders as the goal of the instructional reform. Given our directions to include a problem-solving lesson and a whole-class discussion, it would be appropriate to think of what was video recorded as teachers' best shot at enacting reform-oriented instructional practices. Although the majority of the lessons were contained within a single class period, 44 of the lessons spanned 2 days. Therefore, we analyzed a total of 460 lessons for the 165 teachers (132 teachers participated in Year 3; 87 of those teachers remained in the study in Year 4, and 33 teachers joined the study in Year 4).

Measuring Students' Opportunities to Learn Mathematics

The video-recorded lessons for each teacher were coded using an expanded version of the Instructional Quality Assessment (Boston, 2012; Matsumura et al., 2006). The standard Instructional Quality Assessment (IQA) is based on the Mathematical Tasks Framework and is consistent with the districts' instructional visions and professional development programs. The IQA is designed to measure the cognitive demand of the task as it appears in curricular materials, the cognitive demand of the task as implemented, and the quality of the concluding whole-class discussion. Prior studies have shown the individual IQA observational rubrics to be sufficiently reliable and valid (Boston & Wolf, 2006; Matsumura, Garnier, Slater, & Boston, 2008). The IQA did not have a rubric specific to the setup phase of instruction, other than a measure of the clarity of expectations regarding a final work product.[3] We therefore developed a set of rubrics to measure the quality of the task setup. We refer to the use of the standard IQA and the additional task-as-set-up rubrics as the Expanded IQA. Figure 2 provides a brief description of each rubric, and Figure 3 illustrates how each of the rubrics maps onto the Mathematical Tasks Framework and a three-phase lesson structure.

The Expanded IQA focuses on what teachers and students do in the classroom; however, it does not directly measure what students actually learned via instruction. Therefore, in the analyses that follow, we explicitly refer to students' *opportunities to learn*, with the assumption that the higher the scores on the rubrics, the more likely it is that students were provided opportunities to learn significant mathematics.

---

[3]At the start of the larger project, the research team chose not to use the rubrics related to Clear Expectations for two reasons: (a) pilot studies revealed that they were the most challenging rubrics in producing reliable estimates (Matsumura et al., 2006), and (b) many of the elements of the Clear Expectations rubrics were designed to be used with student work samples, which we did not collect from the participating teachers.

	Rubric	Focal Aspect of Instruction
STANDARD IQA RUBRICS	Task Potential*	Cognitive demand of the task as it appears in the curricular materials
	Task Implementation	Cognitive demand of the task as it is implemented (after students start to work on solving the task through the end of the lesson)
	Academic Rigor of the Discussion*	Academic rigor of the whole-class discussion
	Participation	The percentage of students who participate in the whole-class discussion
	Teacher Linking	Teacher links between contributions within the whole-class discussion
	Student Linking*	Student links between contributions within the whole-class discussion
	Teacher Asking	Teacher press for conceptual explanations within the whole-class discussion
	Student Providing*	Student providing of conceptual explanations within the whole-class discussion
SETUP RUBRICS	Contextual Features*	Building a taken-as-shared understanding of the contextual features of the problem-solving scenario in the task statement
	Mathematical Relationships*	Building a taken-as-shared understanding of the mathematical relationships and ideas in the task statement
	Setup Maintenance*	Maintenance of the cognitive demand of the task specific to the setup phase of instruction
	Post Setup Task Potential	Cognitive demand of the task (based on the instructional materials) at the end of the setup
	Setup Participation	The percentage of students who participate in the setup discussion

Note. Rubrics marked with * are included in this analysis.

Figure 2. Expanded IQA rubrics and focal aspects of instruction.

Standard IQA measures. In our analyses, we used the following IQA measures of opportunities to learn: Task Potential; Academic Rigor of the Discussion; and two finer grained measures of the quality of discussion, Student Linking and Student Providing.[4] Each is explained below, and all Expanded IQA rubrics are available at http://peabody.vanderbilt.edu/departments/tl/teaching_and_learning_research/mist/mist_instruments.php.

[4] The standard IQA includes several other measures of the quality of discussion: teacher linking, teacher asking, and student participation. For the purposes of this analysis, we decided to use the measures related to the content of student talk because of our focus on how the setup might support students' participation in the concluding whole-class discussion. Additionally, in our data, teacher linking and teacher asking are highly correlated with the corresponding student measures. We did not

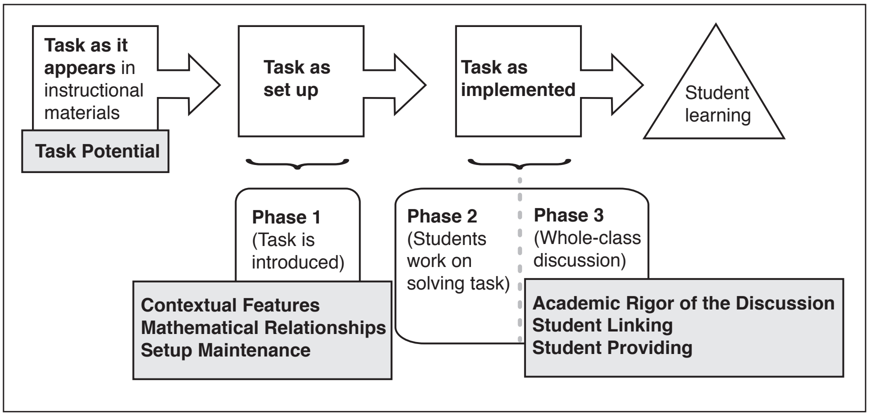

Figure 3. Mathematical Tasks Framework (Stein et al., 2000), three-phase lesson structure, and Expanded IQA rubrics (limited to those used in this analysis).

The task refers to the activity (usually written) that the majority of the students participated in for the majority of the time; there was only one task scored for each video-recorded lesson. *Task Potential* measures the cognitive demand of the task that is posed. For Task Potential, a score of 0 indicates the task is not mathematical in nature. Scores of 1 or 2 indicate that a task has relatively low cognitive demand. A score of 1 indicates that students are asked to memorize or reproduce facts. A score of 2 indicates that students are asked to apply a standard procedure to solve a relatively routine problem. Scores of 3 or 4 indicate that a task has relatively high cognitive demand. A score of 3 indicates that students are asked to engage in complex thinking (e.g., make mathematical connections, create meaning for a procedure), but not necessarily provide evidence for their reasoning. A score of 4 indicates that students are asked to solve a relatively nonroutine problem and to provide evidence of their mathematical reasoning (Stein et al., 2000). For example, the Dollars for Dancing task (Figure 1) would be scored a 4 because there are multiple solution pathways, students are expected to make connections between representations, and students are expected to explain their reasoning. Given that we did not have data on individual students' mathematical understanding or the nature of instruction that happened prior to the video recordings, we were unable to judge cognitive demand as students actually experienced a task. Instead, any measure of cognitive demand is based on the nature of reasoning expected in the classroom activity.

Academic Rigor of the Discussion (ARD) measures the quality and nature of the discourse used in the concluding whole-class discussion phase of the lesson; both

use measures of student participation because we found that with video recordings of instruction it was difficult to reliably count the exact number of student participants. The Post Setup Task Potential only differed in about 6% of lessons from the Task Potential, so we did not include that in our analyses. Because we chose to focus on the concluding whole-class discussion phase of the lesson, we did not include the overall Task Implementation score in our analyses.

teacher and student actions are taken into account when assigning a score. For ARD, a score of 0 indicates there was no concluding whole-class discussion. A score of 1 indicates that students provide brief or one-word answers. A score of 2 indicates that, in a whole-class format, students describe their written work for solving the task but do not engage in a discussion of their strategies, procedures, or mathematical ideas. This type of discussion often takes the form of a show-and-tell in which students take turns sharing solutions with little support to engage in conceptual discourse or to make connections between solution strategies and important mathematical ideas (Ball, 2001). A score of 3 indicates that students show or describe written work for solving a task and/or engage in a discussion of the important mathematical ideas in the task. During the discussion, students provide explanations of why their strategy, idea, or procedure is valid and/or students begin to make connections, but the explanations and connections are not complete and thorough. A score of 4 indicates that, in the discussion, students provide thorough explanations of why particular strategies are valid and make connections between strategies and the underlying mathematical ideas.

The standard IQA includes additional measures that provide a finer grained analysis of the quality of whole-class discussion. *Student Linking* measures the extent to which students' contributions link to and build upon each other; the focus is on what students do, not what the teacher does. Examples of Student Linking include revoicing ideas and relating an idea to someone else's idea. A score of 1 indicates that students do not make any effort to link to prior contributions. A score of 2 indicates that students make superficial efforts to connect ideas (e.g., there are indications of connections between ideas without explicit talk about how they are connected). A score of 3 indicates that students make efforts to connect ideas to each other with some talk of how they are connected; a score of 4 indicates that students consistently make explicit connections.

Student Providing measures the extent to which students support their contributions with evidence or reasoning; similar to Student Linking, the focus is on what students do, not what the teacher does. A key distinction in measuring Student Providing is whether the nature of student reasoning is *conceptual* or *calculational*. Research suggests that productive whole-class discussions, or those that support students' conceptual understanding of central mathematical ideas, are characterized by conceptual rather than calculational discourse (Cobb, Stephan, McClain, & Gravemeijer, 2001; A. G. Thompson, Philipp, Thompson, & Boyd, 1994). Calculational discourse refers to discussions that only emphasize how one arrived at a solution. In contrast, conceptual discourse refers to discussions that emphasize why one chose to solve a problem in a given way and how the problem was solved. A student provides an explicit rationale for using particular methods and grounds any talk of mathematical quantities or relationships in the context of the given task. Conceptual explanations are therefore more likely to support all students' learning, particularly the listening students, because the student provides an "explicit account of the task interpretations that underpin particular solution strategies" (Jackson & Cobb, 2010, p. 24). A score of 1 indicates that there are no efforts on the part of students to provide evidence for their contributions or to explain their thinking. A score of 2 indicates calculational explanations or insufficient evidence. A score of 3 or 4 indicates that students provide conceptual explanations, with a 4 indicating that students consistently provide conceptual explanations. A lesson receives a 0 for both Student Linking and Student Providing if there is no concluding whole-class discussion.

Task-as-set-up measures. A team of mathematics educators and doctoral students, including the first four authors of this paper, developed a set of rubrics to measure the quality of the task-as-set-up phase of instruction. As described above, we identified four key aspects of high-quality setups in a qualitative analysis of 40 video recordings collected in Year 1 of the project (2007–2008). We then developed a set of rubrics based on those observed aspects. The development of the rubrics was an iterative process carried out over two years. We used drafts of rubrics to code existing lessons from Years 1 and 2 of the project to both refine them conceptually (e.g., deciding how taken-as-shared understanding of particular features of a task might be developed) and pragmatically (e.g., deciding how to code setups of tasks without a problem-solving scenario). We regularly consulted with an expert in measuring the quality of instruction in mathematics education throughout the process of rubric development. Because we know of no instrument that attends specifically to the setup, we were unable to externally validate our measures with an existing instrument. However, one approach to testing construct validity is to look for relationships between the measures and existing evidence of opportunities to learn (Kane, 2006). In our case, we examined relationships between what happened in the setup and opportunities to learn in the concluding whole-class discussion. As we share in the results section, we found significant relationships between these two phases of instruction using our instrumentation.

The setup rubrics and measures were designed to complement the standard IQA rubrics, and as such, they employ compatible language and have a structure similar to the standard IQA rubrics described above. *Contextual Features* (CF) measures the extent to which students are supported to develop taken-as-shared understanding of the contextual features of the problem-solving scenario (PSS). The CF rubric is only used if the task has a PSS (i.e., the mathematical task is presented in the context of a story or scenario). For example, a naked-number task or a problem-solving task that does not include a scenario (e.g., find the angle of rotation of a set of objects and then conjecture about the relationship between angles of rotation and rotational symmetry) would not be scored using the CF rubric. A score of 0 indicates that no attention is given to the contextual features. A score of 1 indicates that the teacher is the only person providing information about the contextual features of the PSS; at best, students provide yes/no responses. A score of 2 indicates that the teacher elicits what students understand about the contextual features, but their ideas remain unconnected. A score of 3 or 4 indicates that the teacher and students connect ideas together regarding the contextual features, with a 4 indicating these connections are consistent or happen frequently. Therefore, a score of

3 or 4 indicates there is evidence that a taken-as-shared understanding of the contextual features of the PSS was likely developed.

Mathematical Relationships (MR) measures the extent to which students are supported to develop taken-as-shared understanding of mathematical ideas and relationships, as they are represented in the task statement. To be clear, coders are not asked to assess the relevance of the mathematical ideas that a teacher chooses to discuss in the setup. Instead, the coder is asked to assess the quality of teacher and student talk with respect to the mathematical ideas discussed. The MR rubric is used for any type of task. A score of 0 indicates that no attention is given to the mathematical relationships. A score of 1 indicates that the teacher is the only person providing information about the mathematical relationships; at best, students provide yes/no responses. A score of 2 indicates that the teacher elicits what students understand about the mathematical relationships, but their ideas remain unconnected. A score of 3 or 4 indicates that the teacher and students consistently connect ideas together regarding mathematical relationships, with a 4 indicating evidence of a connection that is conceptual in nature. Thus, a score of 3 or 4 indicates there is evidence that a taken-as-shared understanding of the mathematical relationships was likely developed.

We also measured whether the cognitive demand of the task was maintained, increased, or decreased during the task-as-set-up phase of instruction; we call this the *Setup Maintenance*. A lesson was scored as *maintain* if the setup did not alter the cognitive demand of the initial task the students were to complete. A lesson was scored as *decrease* if the teacher reduced the cognitive demand of the initial task. Such a decrease in cognitive demand can manifest itself subtly if a teacher suggests a particular approach to the problem, yet still requires students to carry out the task as described in the instructional materials. More dramatically, a teacher may decrease the cognitive demand during the setup by modifying the task as described in the instructional materials (e.g., the teacher eliminates task sections that required students to explain their mathematical reasoning). Our coding scheme allowed for an increase in the cognitive demand during the setup. However, this score category was used only one time in our sample, so we marked this case as *maintain* for the purpose of this analysis.

Coding of Video Recordings

Nine coders were trained to use the Expanded IQA in a reliable manner. Five coders were trained in Year 3 of the study and coded the Year 3 (2009–2010) data. Seven coders were trained in Year 4 of the study and coded the Year 4 (2010–2011) data; three of these coders had been trained the previous year and coded the Year 3 data as well. We followed the same coding process for both Year 3 and Year 4 data. Before actual coding began, coders were required to achieve 80% exact score agreement across the rubrics on a set of previously coded videos, which were chosen to represent the variety of anomalies that the coders would encounter. Each of the coders was randomly assigned a list of teachers to code using the Expanded IQA. The set of 2 class days for each teacher was coded chronologically, given that it was possible that the lesson from the first day might continue into the second day, which would result in just one set of scores for the lesson. It is important to note that, given this coder assignment process, the same coders scored the setup and concluding whole-class discussion phases of instruction, often in the same sitting. However, coders generally assigned scores for the setup phase of instruction at the end of the setup, prior to viewing the subsequent phases of the lesson. We mention this to acknowledge the fact that the same coder assessing the setup and the whole-class discussion could be a source of some bias in the tested relationships between the two phases of instruction; however, we do not think that it outweighs the potential benefits of this analysis.

Over the course of the coding period, one set of teacher scores for each coder was randomly checked for reliability once every 2 weeks by one of two reliability coders to account for rater drift, which resulted in double-coding of approximately 10% of the sample. Any discrepancies were consensus-coded to maximize the accuracy of the scores and to allow for ongoing learning on the part of the coders. The overall percent agreement for the coders (across Year 3 and Year 4 coding) was 70.5%, with an average Cohen's kappa score of 0.48. The percent agreement range for the Expanded IQA was from 60.1% to 82.0%, and the kappa scores ranged from 0.30 to 0.60. Because this is the first time the task-as-set-up rubrics were used on a large scale, we give reliability information for each rubric separately in Table 1. We also provide reliability information for the rubrics from the standard IQA that we use in our analyses. In general, the reliability scores for the task-as-set-up rubrics do not differ significantly from the reliability scores for the standard IQA rubrics.

Table 1
Reliability Scores for the Expanded IQA Rubrics

Rubric	Percent agreement	Kappa
Task Potential	67.1	.57
Academic Rigor of the Discussion (ARD)	69.2	.60
Student Linking	82.0	.42
Student Providing	65.3	.47
Setup Maintenance	77.4	.50
Contextual Features (CF)	72.1	.51
Mathematical Relationships (MR)	60.1	.30

Methods of Analyses

Nature of the setup phase of instruction. In our first set of analyses, we described the nature of setups across the classrooms participating in the study. It is possible that the nature of the setup of tasks might differ dramatically depending on whether a task included a PSS. For this reason, we generally examined these two types of lessons (lessons with PSS tasks, lessons with non-PSS tasks) separately as well as together so we could identify important differences. In our sample of 460 lessons, 267 involved PSS tasks while the other 193 lessons involved non-PSS tasks. Recall that the CF rubric only applies to PSS tasks, so the subsample of lessons with both MR and CF scores consists of the 267 lessons involving PSS tasks. We used the Wilcoxon rank-sum test when the samples were not matched to test the hypothesis that two independent samples (i.e., unmatched data) were from populations with the same distribution. We used the Wilcoxon signed-rank test when the samples were matched to test the equality of matched pairs of observations.

Relationships between the setup and concluding whole-class discussion. Our second set of analyses examined relationships between the quality of the setup and students' opportunities to learn in the concluding whole-class discussion. Given the nature of our data, we are only able to descriptively report relationships rather than make causal claims. Several sets of multilevel ordered logistic regression models allowed us to explore and describe the relationships between aspects of the setup and students' opportunities to learn within the concluding whole-class discussion. Scores were assigned at the lesson level (in total one to four lessons per teacher, depending on whether the lesson spanned 2 days of instruction and whether the teacher participated in both years of the study). We used a multilevel modeling approach to account for the fact that we treated multiple lessons for teachers as independent observations and because teachers were nested within schools.

For ease of interpretation of results, we describe a series of four models that were estimated for each of the three outcomes of interest (ARD, Student Linking, and Student Providing). We considered the possibility that the relationship between the setup and concluding whole-class phases of discussion may vary depending on whether a task had a PSS. For this reason, we controlled for the two types of tasks and tested interactions between whether the task involved a PSS and MR and Maintenance. We found no statistical difference. Therefore, in the results section, we generally present models in which lessons with both kinds of tasks are considered together. The exception is the models in which we tested the relationship between CF and the outcomes of interest; in those cases, we only consider lessons using a task with a PSS.

Across all models, we included scores for Task Potential, Maintenance, and MR, as well as a set of dummy variables to indicate district membership. Given that prior research suggests the level of challenge of the task could influence the quality of the whole-class discussion (Boston & Wolf, 2006), we dichotomized Task Potential as low (score of 1 or 2) or high (score of 3 or 4) and controlled for High Potential in each of the models.

In Models 1 and 2, we investigated the relationship between the outcome of interest (ARD, Student Linking, Student Providing) and attention to MR. Therefore, we regressed the outcome of interest on High Potential, Maintenance, MR, whether the task involved a PSS, and district controls. In Model 1, we used the combined measure of MR and did not explore the various MR levels; in Model 2, we explored the differential relationships of each MR score level. We did this by replacing the MR score in the models with dummy variables for MR score categories (MR 0, MR 2, MR 3, and MR 4) while keeping the rest of the model the same. We would have used the lowest score as the comparison category, but in this sample, only about 6% of the lessons were scored at 0 for MR (see Table 3), indicating no attention to the mathematical relationships. Therefore, we used a score of 1, indicating minimal attention to the mathematical relationships, as the comparison category. To control for the rare instances with no attention to mathematical relationships, we included MR 0 in the model.

Models 3 and 4 were limited to lessons with PSS tasks because they included attention to CF. Model 3 included the same covariates as Model 1, with the addition of CF. In Model 3, CF scores ranged from 0 to 4 (five possible values) and were modeled as a continuous variable. Model 4 used our approach of exploring differential relationships by score levels, as in Model 2, but this time, both MR score categories and CF score categories (CF 1, CF 2, and CF 3/4) were included as dummy variables. We combined 3s and 4s for CF because 3s or 4s only occurred in 27 lessons, with only 3 lessons scored at a 4. Additionally, we used the lowest score (0 for CF) as the comparison category; a score of 0 on CF was the most common CF score across all the lessons (see Table 3).

Given that results from ordered logistic regression are in the form of logits, we include odds ratios (OR) to aid in interpretation. In general, the odds ratio is an effect size that gives the ratio of the odds of an event occurring in one group compared with the odds of an event occurring in another group (Cohen, Cohen, West, & Aiken, 2003). Also, we report three different levels of significance in the tables of results for the models, but in the narrative descriptions of our results, we use $p < .10$ as our cutoff for marginal statistical significance.

Results

Descriptive Statistics of the Observed Lessons

We first provide descriptive statistics of the Standard IQA scores for the 460 lessons (i.e., each rubric's score range, mean, and standard deviation); see Table 2. The potential of the tasks used in 274 of the 460 lessons was rated as of high cognitive demand (i.e., 3 or 4 on Task Potential). All but 5 of the 460 lessons were rated a 2 or higher on Task Potential. With regard to scores for the concluding whole-class discussion, 98 of the 460 lessons (about 21%) did not include a concluding whole-class discussion and, therefore, received a 0 for ARD, Student Linking, and Student Providing.[5] Of the lessons with a whole-class discussion, the majority of the scores for ARD were low; only 73 lessons were scored at a 3 or 4. Scores for Student Linking and Student Providing were similarly low. The highest score for Student Linking was a 3, and that occurred in only 9 lessons. For Student Providing,

[5] This finding is fairly consistent with what Matsumura et al. (2006) reported in their pilot study of the IQA and middle-grades mathematics teaching in urban schools; they found that 30.8% of lessons included, at best, a nonmathematical whole-class discussion.

Table 2
Descriptive Statistics of Standard IQA Scores Across 460 Lessons

Rubric	Mean	0	1	2	3	4
Task Potential	2.77	0	5	181	190	84
Academic Rigor of the Discussion (ARD)	1.55	98	103	186	55	18
Student Linking	0.94	98	299	54	9	0
Student Providing	1.39	98	153	148	54	7

Table 3
Frequencies of Particular Scores for Contextual Features (*CF*) *and Mathematical Relationships* (*MR*) *by Use of Task With or Without Problem-Solving Scenario* (*PSS*)

Score for CF or MR	PSS task ($N = 267$)		Non-PSS task ($N = 193$)
	CF	MR	MR
4	3 (1%)	16 (6%)	11 (6%)
3	24 (9%)	38 (14%)	26 (13%)
2	61 (23%)	165 (62%)	119 (62%)
1	76 (28%)	33 (12%)	26 (13%)
0	103 (39%)	15 (6%)	11 (6%)
Mean	1.06	2.03	2.00
SD	1.04	.86	.85

Note. The percentage listed is the percentage of the total number of lessons in the type-of-task category (i.e., out of 267 lessons with PSS tasks and out of 193 lessons with non-PSS tasks).

61 lessons were scored at a 3 or higher. Overall, the majority of scores for quality of the whole-class discussion were at a level 1 or 2.

The Nature of the Setup Phase of Instruction

To what extent did teachers attend to contextual features and the mathematical relationships of a task statement during the setup phase of instruction? Table 3 gives frequencies of CF and MR scores for lessons with PSS tasks and non-PSS tasks. As represented in Table 3, in lessons with PSS tasks, the scores for MR are higher than scores for CF. A Wilcoxon signed-rank test suggests the difference in score distributions is statistically significant ($z = 11.27$, $p < .001$). In other words, teachers appear to attend in higher quality ways to the mathematical relationships than to the contextual features. Additionally, a two-sample Wilcoxon rank-sum test suggests that there is no statistical difference between lessons with PSS and those with non-PSS tasks in the distribution of the quality of attention to the MR ($z = 0.37$, $p = .71$).

Given that the quality of the attention to mathematical relationships tends to be higher than the quality of the attention to the contextual features, one might wonder if high-quality attention to mathematical relationships occurs at the expense of attention to contextual features. The cross-tabulation of scores for CF and MR given in Table 4 shows that this does not appear to be the case. In fact, scores for CF and MR are significantly positively correlated (Spearman's rank correlation, $\rho = .142$, $p < .05$), meaning that in lessons with PSS tasks, scores for one measure generally tend in the direction of scores for the other. Despite that trend, only 18 of the 267 lessons involve attention to the mathematical relationships and contextual features in taken-as-shared ways (i.e., at a level 3 or 4 on both rubrics).

A key aspect of setting up complex tasks is maintaining the cognitive demand of the task. Table 5 provides an overview of the extent to which the cognitive demand of the task was maintained during the setup. The cognitive demand of the task was maintained in the setup in less than half (36.1%) of the lessons. In addition, Fisher's exact test suggests significant differences in maintenance of the cognitive demand by Task Potential scores ($p < .05$). Further, the percentage of lessons in which the cognitive demand was maintained during the setup was higher for lessons with Task Potential 4 than for lessons with Task Potential 2 or 3.

We also examined the extent to which teachers maintained the cognitive demand while attending to the contextual features and the mathematical relationships of the task. Table 6 provides the overall percent of lessons within each MR and CF score category in which the lesson received a Setup Maintenance score of decrease. There was considerable variation in the percentages of lessons that decreased in cognitive demand in relation to attention to MR and CF during the setup. The fact that the cognitive demand of the task was maintained in about 55% of the setups in lessons with MR scores of 4 (i.e., the cognitive demand of the task was decreased in 44.4% of the setups in lessons with MR scores of 4) suggests it is possible to attend to the mathematical relationships of complex tasks in taken-as-shared ways without decreasing the cognitive demand (see Table 6). Also, the highest percentage of lessons with decreases in cognitive demand occurs in lessons in which the attention to MR is at a level 2. Lastly, we did not find the same trends in the percentages of lessons that decreased in cognitive demand with regard to the score categories for CF; the percentages do not differ dramatically by score category and range from 33.3% to 65.0%. Although setting up a task so that all students have access while simultaneously maintaining the cognitive demand is challenging, these findings suggest that it is possible.

Relationships Between the Setup and Students' Opportunities to Learn in the Concluding Whole-Class Discussion

Relationships between the setup and Academic Rigor of the Discussion. Here, we present the results from our analyses of the relationships between the quality of the setup and students' opportunities to learn in the concluding whole-class

Table 4
Cross-Tabulation of Contextual Features and Mathematical Relationships for Lessons With Tasks With Problem-Solving Scenarios (*n = 267*)

	Contextual features				
Mathematical relationships	0	1	2	3	4
4	2	1	8	3	2
3	4	8	13	12	1
2	73	53	30	9	0
1	13	12	8	0	0
0	11	2	2	0	0

Table 5
Task Potential With Setup Maintenance (*n = 460*)

	Setup maintenance			
Task potential	Maintain	Decrease	Total	% maintain
4	43	41	84	51.2%
3	67	123	190	35.3%
2	53	128	181	29.3%
1	3	2	5	60.0%
Totals	166	294	460	36.1%

Table 6
Overall Counts for Contextual Features (*CF*) *and Mathematical Relationships* (*MR*) *Scores With Percentages of Setups With Each Score in Which the Cognitive Demand Was Decreased*

Score	0		1		2		3		4	
	N	% Dec	*N*	% Dec	*N*	% Dec	*N*	% Dec	*N*	% Dec
CF	103	65.0%	76	64.4%	61	59.0%	24	50.0%	3	33.3%
MR	26	0%	59	47.5%	284	74.3%	64	67.2%	27	44.4%

discussion. We begin by discussing the results from the models with ARD as the outcome (see the full set of results in Table 7).

First, Maintenance was consistently significant and positively related to ARD while controlling for High Potential, MR, CF (where applicable), type of task, and district membership ($p < .05$).

Second, scores on MR were consistently significant and positively related to ARD. Further exploration revealed that there is some variation by model, but

Table 7
Relationships Between Aspects of the Setup and Academic Rigor of the Discussion

	(1) MR		(2) MR levels		(3) MR & CF		(4) MR & CF levels	
Academic rigor of the discussion	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio
High Potential	0.49 (0.22)	1.63**	0.51 (0.22)	1.66**	0.28 (0.29)	1.33	0.29 (0.29)	1.34
Maintenance	0.97 (0.21)	2.64***	0.81 (0.22)	2.24***	0.81 (0.27)	2.26**	0.65 (0.28)	1.92**
MR	0.56 (0.12)	1.75***			0.44 (0.17)	1.55**		
MR 0			0.28 (0.52)	1.33			0.95 (0.65)	2.59
MR 2			0.50 (0.30)	1.64			0.76 (0.40)	2.15*
MR 3			1.25 (0.38)	3.50**			1.29 (0.52)	3.63**
MR 4			2.22 (0.50)	9.25***			1.87 (0.65)	6.46**
CF					0.21 (0.13)	1.23		
CF 1							−0.07 (0.31)	0.94
CF 2							0.46 (0.35)	1.58
CF 3/4							0.67 (0.49)	1.96
Non-PSS	−0.16 (0.20)	0.85	−0.16 (0.20)	0.86				
District B	−0.38 (0.40)	0.68	−0.35 (0.39)	0.71	−0.34 (0.45)	0.71	−0.37 (0.45)	0.69
District C	−0.89 (0.43)	0.41**	−0.90 (0.43)	0.41**	−0.79 (0.51)	0.46	−0.88 (0.50)	0.42*
District D	0.25 (0.40)	1.28	0.26 (0.40)	1.29	0.22 (0.47)	1.25	0.23 (0.47)	1.26
N	460		460		267		267	

*$p < .1$. **$p < .05$. ***$p < .001$.

scores of 3 or 4 were significantly related to an increase in scores on ARD. In general, the higher the score on MR, the more likely it is for a lesson to receive a higher score on ARD. For example, as shown in Model 2 (which controls for High Potential, maintenance of the cognitive demand, type of task, and district membership), if a lesson has a MR score of 3, then the odds of scoring a 3 on ARD are 3.5 times greater than the odds of scoring a 2 on ARD, as compared to a lesson with a score of 1 on MR ($p < .05$). In other words, if there is evidence that taken-as-shared understanding of mathematical ideas is developed in the setup (corresponding to a MR score of 3), as opposed to only the teacher describing mathematical ideas to the students (MR score of 1), then the odds that the concluding whole-class discussion will be characterized by conversation in which students are providing some evidence for their reasoning (ARD score of 3) are 3.5 times the odds of the occurrence of a show-and-tell form of discussion (ARD score of 2). In addition, for a lesson with a MR score of 4, the odds of scoring a 3 on ARD are 9.25 times greater than the odds of scoring a 2 on ARD, when compared with a lesson receiving a MR score of 1 ($p < .001$). These results suggest that the quality of the attention to mathematical relationships in the setup is crucial—the greater the attention to establishing a taken-as-shared understanding of mathematical relationships in the setup, the stronger the relationship with the quality of the concluding whole-class discussion.

Third, neither of the models that include CF (see Models 3 and 4) resulted in a statistically significant relationship between CF and ARD when controlling for High Potential, Maintenance, MR, and district, even when CF was examined by score category.

Fourth, this set of models generally suggests significant differences between District A and District C with regard to ARD, which is represented by marginally significant coefficients on the dummy variable. This result suggests that there are differences between the districts that are not fully explained by differences in High Potential or our measured aspects of the setup; however, these differences are not the focus of this analysis.

Finally, results from two of these models suggest that High Potential is positively and significantly related to ARD. In other words, the level of challenge of the task is related to the quality of the academic rigor of the whole-class discussion. This result is consistent with prior research (Boston & Wolf, 2006) and suggests that the level of the challenge of the task may set the stage for high-quality discussions.

Relationships between the setup and Student Linking. We next discuss the results from the models that explored relationships between aspects of the setup and Student Linking (see the full set of results listed in Table 8). Recall that Student Linking measures the extent to which students' contributions link to and build on each other in the concluding whole-class discussion. First, we see that Maintenance was somewhat consistently significant and positively related to Student Linking in the discussion while controlling for the High Potential, MR, CF (where applicable), type of task, and district membership. Second, scores on MR were somewhat

Table 8
Relationships Between Aspects of the Setup and Student Linking

Student linking	(1) MR		(2) MR Levels		(3) MR & CF		(4) MR & CF levels	
	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio
High Potential	0.44 (0.25)	1.56*	0.45 (0.25)	1.57*	0.24 (0.32)	1.27	0.22 (0.32)	1.25
Maintenance	0.64 (0.24)	1.89**	0.49 (0.25)	1.64**	0.70 (0.29)	2.01**	0.52 (0.31)	1.68*
MR	0.41 (0.14)	1.50**			0.34 (0.18)	1.40*		
MR 0			0.44 (0.59)	1.56			1.11 (0.75)	3.05
MR 2			0.44 (0.34)	1.55			0.62 (0.43)	1.86
MR 3			0.83 (0.43)	2.28*			0.76 (0.58)	2.13
MR 4			1.85 (0.54)	6.34**			1.90 (0.68)	6.68**
CF					0.30 (0.15)	1.34*		
CF 1							0.08 (0.34)	1.09
CF 2							0.46 (0.39)	1.58
CF 3/4							1.02 (0.55)	2.78*
Non-PSS	0.26 (0.23)	1.29	0.26 (0.24)	1.30				
District B	−0.43 (0.42)	0.65	−0.41 (0.43)	0.66	−0.29 (0.42)	0.75	−0.28 (0.43)	0.75
District C	−0.73 (0.46)	0.48	−0.75 (0.46)	0.47	−0.56 (0.47)	0.57	−0.63 (0.48)	0.53
District D	0.20 (0.43)	1.22	0.23 (0.43)	1.26	0.08 (0.44)	1.08	0.13 (0.45)	1.14
N	460		460		267		267	

* $p < .1$. $^{**}p < .05$. $^{***}p < .001$.

consistently significant and positively related to Student Linking. Similar to the models with ARD as the outcome, the higher scores on MR appeared to drive the positive relationship between MR and Student Linking.

Third, both models that include scores for CF demonstrate marginally significant positive relationships between CF and Student Linking when controlling for High Potential, Maintenance, MR, and district membership. In Model 3, which included one CF score variable, the relationship is marginally statistically significant (OR = 1.34, $p < .10$). In Model 4, the model in which we investigated the differential impact of CF scores, we specifically see that high CF (i.e., CF scores of 3 or 4) is marginally significant and positively related to Student Linking when controlling for High Potential, Maintenance, MR, and district membership (OR = 2.78, $p < .10$ for CF = 3/4). In other words, if there is evidence that taken-as-shared understanding of contextual features is developed in the setup (corresponding to a CF score of 3/4), contrasted with no attention to the contextual features (CF score of 0), then the odds that students will connect their contributions to each other and show how they relate two or more times within the lesson (corresponding to a Student Linking score of 3) is nearly 3 times higher than the odds of students connecting their contributions at most one time (Student Linking score of 2). These results generally suggest that developing common language to describe key contextual features of a problem-solving scenario in the setup may help students to make connections between their solutions and those of their peers in the concluding whole-class discussion.

Relationships between the setup and Student Providing. Finally, we discuss the results from the models that explored relationships between aspects of the setup and Student Providing (see the full set of results in Table 9). Recall that Student Providing measures the extent to which students support their contributions with evidence or reasoning. First, we see that Setup Maintenance was significant and positively related to Student Providing in the discussion while controlling for High Potential, MR, CF (where applicable), type of task, and district membership. Second, scores on MR were consistently significant and positively related to Student Providing. Scores of 3 or 4 were significantly related to a score increase on Student Providing. In addition, results generally suggest that the higher the score on MR, the more likely it is for a lesson to receive a higher score on Student Providing. For example, as shown in Model 2, if there is evidence that taken-as-shared understanding of mathematical ideas is developed in the setup (corresponding to a MR score of 3), as opposed to only the teacher describing mathematical ideas to the students (MR score of 1), then the odds of scoring a 3 on Student Providing are 2.77 times the odds of scoring a 2 on Student Providing ($p < .05$). In addition, for a lesson with a MR score of 4, the odds of scoring a 3 on Student Providing are 15.65 times the odds of scoring a 2 on Student Providing, as compared to a lesson receiving a MR score of 1 ($p < .001$). These results suggest the greater the attention to establishing a taken-as-shared understanding of mathematical relationships in the setup, the stronger the relationship with the extent to

Table 9
Relationships Between Aspects of the Setup and Student Providing

Student providing	(1) MR		(2) MR Levels		(3) MR & CF		(4) MR & CF levels	
	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio	Coef (*SE*)	Odds ratio
High Potential	0.46 (0.22)	1.58**	0.45 (0.22)	1.56**	0.20 (0.30)	1.23	0.15 (0.30)	1.17
Maintenance	0.78 (0.21)	2.17***	0.55 (0.22)	1.73**	0.76 (0.28)	2.14**	0.56 (0.29)	1.74*
MR	0.58 (0.12)	1.79***			0.50 (0.18)	1.65**		
MR 0			0.42 (0.52)	1.52			1.06 (0.67)	2.87
MR 2			0.41 (0.30)	1.50			0.77 (0.41)	2.16*
MR 3			1.02 (0.37)	2.77**			1.03 (0.54)	2.81*
MR 4			2.75 (0.51)	15.65***			2.57 (0.73)	13.06***
CF					0.27 (0.14)	1.31*		
CF 1							−0.26 (0.32)	0.77
CF 2							0.21 (0.37)	1.24
CF 3/4							1.31 (0.52)	3.70**
Non-PSS	0.12 (0.21)	1.13	0.10 (0.21)	1.10				
District B	−0.52 (0.34)	0.59	−0.49 (0.34)	0.61	−0.71 (0.39)	0.49*	−0.66 (0.39)	0.52*
District C	−1.11 (0.37)	0.33**	−1.12 (0.38)	0.33**	−1.13 (0.45)	0.32**	−1.30 (0.45)	0.27**
District D	−0.09 (0.34)	0.91	−0.06 (0.34)	0.94	0.05 (0.41)	1.05	0.11 (0.41)	1.11
N	460		460		267		267	

* $p < .1$. **$p < .05$. ***$p < .001$.

which students provide reasoning or evidence for their contributions in the concluding whole-class discussion.

Third, both models that include scores for CF (Models 3 and 4) demonstrate significant positive relationships between CF and Student Providing when controlling for High Potential, Maintenance, MR, and district membership. In Model 3, which included one CF score variable, the relationship is marginally statistically significant (OR = 1.31, $p < .10$). In addition, in Model 4, which investigates the differential impact of CF scores, we see that high CF (i.e., CF scores of 3 or 4) is significant and positively related to Student Providing when controlling for High Potential, Maintenance, MR, and district membership (OR = 3.70, $p < .05$ for CF = 3/4). This odds ratio can be interpreted to suggest that if there is evidence that taken-as-shared understanding of contextual features is developed in the setup (corresponding to a CF score of 3 or 4) as opposed to no attention to the contextual features (CF score of 0), then the odds that students will provide conceptual evidence for their claims (Student Providing score of 3) are about 3.7 times the odds that students will provide at best procedural evidence in the concluding whole-class discussion (Student Providing score of 2).

More generally, our primary analyses indicate that the maintenance of the cognitive demand, attention to the mathematical relationships of the task, and attention to the contextual features of the task (particularly in taken-as-shared ways) are all positively related to Student Linking and Student Providing in the concluding whole-class discussion. Examining the outcomes of Student Linking and Student Providing reveals that attention to the contextual features during the setup may positively relate to the way students build on each other's contributions and provide conceptual explanations during the discussion.

Limitations of the Study

Before turning to a discussion of our findings, we acknowledge several limitations of this study. First, as noted earlier, our findings are descriptive in nature. Given our data and methods of analysis, we are unable to assert causality. Although we detected a relationship between the nature of activity in the setup and in the concluding whole-class discussion, based on our primary analyses, we cannot claim that what happened in the setup necessarily influenced what happened in the concluding discussion. There are likely unobserved (or unmeasured) aspects of instruction that might account for the relationships we detected between the setup and students' opportunities to learn in the concluding whole-class discussion. For example, we do not know if what was observed during the setup and concluding whole-class discussion phases of the observed lessons were purely the results of norms the teacher had established over the course of the school year, rather than anything particular to the setup phase of instruction. Given our relatively small number of observations per teacher, we were unable to robustly examine a model with teacher fixed effects to rule out the possibility that other characteristics of the participating teachers contributed to our significant findings. In the future, similar, larger scale studies should use teacher fixed effects to help rule out such alternate hypotheses and strengthen the causality argument.

Second, the questions we asked of our data, and therefore our findings, generally assumed a three-phase lesson structure. Although investigating variability within lesson structures is beyond the scope of this study, it is important to note that although teachers were asked to include a whole-class discussion, not all did. Also, teachers spent varied amounts of time on the setup versus other phases of instruction. Although we collected data on time spent on various phases of instruction, we did not analyze it for the purpose of this study.

Third, we did not investigate the relationship between the setup and the second phase of instruction (when students work on solving the task). In fact, the evidence of whether a setup is productive in terms of supporting students' participation is probably best determined by analyzing what happens when students start to solve the task. However, we were unable to systematically examine students' activity in phase two of the lessons because of the nature of the video data. At the start of the research project, it was assumed that the video recordings would be coded by the standard IQA, which attends to phase two in a global manner (e.g., whether the teacher and students reduce the cognitive demand of the task as students work to solve it). The teacher wore a lapel microphone, and the videographers were instructed to place a paddle microphone by one or two groups of students to capture the nature of their talk during phase two. In order to reliably compare activity in the setup to student talk and activity during phase two, we would have needed access to all of the students' talk and activity during that phase of instruction.

Fourth, given the nature of the data collection, we were not privy to what had happened before (or after) the 2 days of instruction. It is possible that a teacher may have established key understanding of the context or mathematical ideas in a prior lesson, and therefore did not need to spend time setting those up in the lesson that was video recorded. In fact, we asked that coders indicate whether there was evidence that aspects of the context might have been developed in prior lessons; coders only indicated this in 29 of the 267 lessons with tasks with PSS. However, this was not checked for reliability, and hence we did not account for this in the study. We also did not attend to the nature of the mathematical topic of the specific lessons that were recorded. Instead, we only asked that coders identify whether a task had a problem-solving scenario. It could be that what is useful to setup for a task varies by mathematical topic.

Discussion and Conclusion

Stein and colleagues (Stein et al., 1996, Stein & Lane, 1996) identified that the setup phase of instruction could influence the extent to which students participate in high-cognitive-demand activity over the course of a lesson. This study contributes to the mathematics education research community's understanding of the nature of teacher–student interactions during the setup phase of instruction that might support more students to substantially participate in and learn through cognitively demanding activity in the classroom, especially in the concluding whole-class discussion.

Our findings suggest that the quality of the setup appears to be related to students' opportunities to learn in the concluding whole-class discussion. First, we found that the quality of the attention to mathematical relationships in the setup was generally positively related to the quality of the concluding whole-class discussion. In particular, engaging in activity aimed at establishing taken-as-shared understanding of mathematical relationships in the setup was more strongly (and significantly) related to the quality of the discussion than low-quality attention to the mathematical relationships in the setup. Furthermore, establishing taken-as-shared understanding of mathematical relationships in a consistent manner was more frequently related to the quality of the discussion than doing so less consistently. This suggests the importance of not only attending to the mathematical relationships in the setup, but doing so in ways that support students to build upon one another's ideas. In other words, it is not enough to have students share their isolated ideas about mathematical relationships in the setup. It appears that some orchestration of discussion (Stein et al., 2008) of those key mathematical relationships is also desirable in the setup.

Second, we found positive relationships between the quality of the setup and the quality of the concluding whole-class discussion regardless of whether the task included a problem-solving scenario. In other words, even in cases where the task did not include a scenario, we detected positive associations between attending to the mathematical relationships of the task in taken-as-shared ways in the setup and the quality of the concluding discussion.

Some of our initial interest in investigating the setup was prompted by prior research that has suggested equity concerns in terms of the cultural suppositions of situations used in problem-solving scenarios (Ball et al., 2005; Boaler, 2002; Silver et al., 1995; Tate, 1995), as well as our experiences observing instruction in which it was evident that not all students were familiar with the contextual features of the scenario. A key question is: When a lesson includes a task with a problem-solving scenario, how is attention to the contextual features in the setup related to students' opportunities to learn in the concluding whole-class discussion? We did not detect a statistically significant relationship between contextual features and the Academic Rigor of the Discussion. However, we did detect statistically significant relationships between high-quality attention to contextual features and the nature of student contributions (i.e., Student Linking, Student Providing) in the concluding whole-class discussion, when controlling for Mathematical Relationships, High Potential, Setup Maintenance, and district membership. In other words, our findings suggest that students were more likely to make connections to one another's ideas and to provide conceptual evidence for their reasoning in the whole-class discussion when taken-as-shared understanding of the contextual features of the problem-solving scenario was established in the setup.

An important, related finding of our study was that in lessons with problem-solving scenario tasks, teachers were more likely to attend to the mathematical relationships than the contextual features, and in higher quality ways. There was no attention to the contextual features in 39% of lessons with problem-solving scenario tasks, while only 6% of those lessons lacked attention to mathematical relationships. Additionally, we found statistically significant differences in the distribution of quality of attention to the two features, with a higher mean for Mathematical Relationships. Even with these differences, high-quality attention to either aspect was relatively rare. Given that we found that high-quality attention to contextual features is positively related to students' contributions in the whole-class discussion, it seems important to research this phenomenon further. For example, how do teachers conceive of the contextual features of a problem-solving scenario? How does their thinking about the contextual features relate to their mathematical goals for the lesson?

In general, when teachers and students attended to contextual features and mathematical relationships in taken-as-shared ways and maintained the high cognitive demand of the task in the setup, discussions were of higher quality. A key question to ask is: How likely is it that setups meet these conditions? We found that the percentage of lessons (about 6.7%) in which these conditions were met was low. Although it was rare, a noteworthy finding of our study is that teachers can develop taken-as-shared understanding of contextual features and mathematical relationships in the setup and maintain the cognitive demand of an activity. We view it is as valuable to have found evidence that it is possible to do both.

A related finding is that in our sample of 165 teachers, it was very common for the cognitive demand of a task to be lowered during the setup phase of instruction (in 294 of 460 lessons, or 64%). This is concerning given that the cognitive demand of an activity is a significant predictor of students' opportunities to learn (Stein & Lane, 1996). Stein et al. (1996) suggested three reasons why teachers might lower the cognitive demand of a high-level task in the setup: their goals for instruction, their subject matter knowledge, or their knowledge of their students. We suggest that further research is necessary to elaborate and specify the conditions not only in which teachers lower the cognitive demand, but also the conditions in which they maintain the cognitive demand of an activity in the setup. What does that teaching look like? How do teachers who maintain the cognitive demand conceive of their mathematical goals in the setup? What role does mathematical knowledge for teaching (Hill et al., 2004) play in the extent to which teachers maintain the cognitive demand of classroom activity? How might teachers' views of their students' mathematics capabilities impact how they set up tasks and the extent to which they maintain the cognitive demand of activity over the course of the lesson (Garrison, 2011)? Answers to questions like these would go some way toward informing the design of professional development with the potential to support key shifts in teachers' practice aimed at maximizing the extent to which all students are engaged in rigorous mathematical work.

A motivation for this study was to contribute to the specification and elaboration of ambitious mathematics teaching. Based on findings from this study, we are suggesting that setting up tasks is a high-leverage practice. However, in order to support teachers in developing this form of practice, it would be necessary to fully describe what setting up tasks in high-quality ways includes. Specifically, it requires the decomposition of the practice of setting up tasks, which involves making "visible the grammar of practice to novices and may require a specific technical language for describing the implicit grammar and for naming the parts" (Grossman et al., 2009, p. 2069). We view our work on delineating four key aspects of setting up tasks as making significant headway toward the decomposition of this practice. However, more research is needed to describe the "grammar of practice" in detail sufficient to support teachers in both analyzing and enacting the practice. Part of this work involves providing images, or representations, of the practice of setting up tasks. We provided one such image in our description of Mr. Smith's setup. Based on our viewing of numerous setups, we are fairly confident that there are multiple ways in which one might set up a task that reflect the key aspects we identified in our study. However, research would be necessary to more systematically develop a set of images that could serve as a foundation for "describing the implicit grammar" of setting up tasks.

It is worth reminding the reader that our study was particular to middle-grades mathematics teachers working in districts that had adopted ambitious goals for students' learning and provided teachers with reform-oriented curricula, as well as other supports. It is likely that the setup is particularly important when using curricular materials aimed at supporting students to engage in cognitively demanding activity. In other words, it may not be necessary to focus as intently on the setup if the goals of instruction are only to support students in developing procedural understanding of mathematics. With the recent adoption of the Common Core State Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) across most states, which emphasize conceptual understanding and procedural fluency in a range of domains, and the development of more cognitively demanding state mathematics assessments, we anticipate that the setup phase of instruction will be of importance if teachers are to engage students in commensurate activity in the classroom.

In closing, we view what happens in the setup phase of instruction as important for students' opportunities to engage in rigorous mathematical activity. Our data suggest that this is a phase of instruction that is not often carefully attended to in middle-grades mathematics teaching, yet appears to be related to the extent to which students are able to participate in concluding whole-class discussions in high-quality ways. More generally, we view this type of analysis—one that is generated from observations of teachers' practice and aimed at making visible forms of practice that teachers can develop to support all students' participation—as necessary to specify how to accomplish ambitious teaching in classrooms.

References

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

Ball, D. L., & Bass, H. (2000). Making believe: The collective construction of public mathematical knowledge in the elementary classroom. In D. C. Phillips (Ed.), *Constructivism in education: Opinions and second opinions on controversial issues* (pp. 193–224). Chicago, IL: The University of Chicago Press.

Ball, D. L., Goffney, I. M., & Bass, H. (2005). The role of mathematics instruction in building a socially just and diverse democracy. *The Mathematics Educator, 15*(1), 2–6.

Ball, D. L., Sleep, L., Boerst, T., & Bass, H. (2009). Combining the development of practice and the practice of development in teacher education. *Elementary School Journal, 109*(5), 458–474. doi:10.1086/596996

Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education, 33*(4), 239–258. doi:10.2307/749740

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. (2012). Assessing the quality of mathematics instruction. *Elementary School Journal, 113*(1), 76–104.

Boston, M. D., & Wolf, M. K. (2006). *Assessing academic rigor in mathematics instruction: The development of Instructional Quality Assessment Toolkit* (Report No. 672). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2003). *Classroom discussions: Using math talk to help students learn*. Sausalito, CA: Math Solutions.

Cobb, P., & Jackson, K. (2011). Towards an empirically grounded theory of action for improving the quality of mathematics teaching at scale. *Mathematics Teacher Education and Development, 13*(1), 6–33.

Cobb, P., & Smith, T. (2008). The challenge of scale: Designing schools and districts as learning organizations for instructional improvement in mathematics. In T. Wood, B. Jaworski, K. Krainer, P. Sullivan, & D. Tirosh (Eds.), *International handbook of mathematics teacher education* (Vol. 3, pp. 231–254). Rotterdam, the Netherlands: Sense.

Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in classroom mathematical practices. *The Journal of the Learning Sciences, 10*(1/2), 113–163. doi:10.1207/S15327809JLS10-1-2_6

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. *American Educational Research Journal, 29*(3), 573–604.

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education, 23*(1), 2–33. doi:10.2307/749161

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Darling-Hammond, L. (2007). The flat earth and education: How America's commitment to equity will determine our future. *Educational Researcher, 36*(6), 318–334. doi:10.3102/0013189X07308253

Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist, 23*(2), 167–180. doi:10.1207/s15326985ep2302_6

Franke, M. L. (2006). Unpacking practice [Web page]. Retrieved from http://www.cfkeep.org/html/stitch.php?s=11093407507142

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225–256). Greenwich, CT: Information Age.

Garrison, A. (2011, September). *The cognitive demand of mathematical tasks: Investigating links to teacher characteristics and contextual factors*. Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics, 39*(1−3), 111–129. doi:10.1023/A:1003749919816

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record, 111*(9), 2055–2100.

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

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11–30. doi:10.1086/428763

Jackson, K., & Cobb, P. (2010, April). *Refining a vision of ambitious mathematics instruction to address issues of equity*. Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Jackson, K., Shahan, E., Gibbons, L., & Cobb, P. (2012). Launching complex tasks. *Mathematics Teaching in the Middle School, 18*(1), 24–29. doi:10.5951/mathteacmiddscho.18.1.0024

Kane, M. T. (2006). Validation. In R. E. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council of Education and Praeger Publishers.

Kazemi, E., Franke, M. L., & Lampert, M. (2009). Developing pedagogies in teacher education to support novice teachers' ability to enact ambitious instruction. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing Divides: Proceedings of the 32nd Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 12–30). Palmerston North, NZ: MERGA.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

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

Lampert, M., Beasley, H., Ghousseini, H., Kazemi, E., & Franke, M. L. (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. doi:10.1007/978-1-4419-0594-9_9

Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. *The Elementary School Journal, 109*(5), 491–509. doi:10.1086/596998

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. D. (2009). *Moving straight ahead: Linear relationships* (Connected Mathematics Project 2, Grade 7). Upper Saddle River, NJ: Prentice Hall.

Lubienski, S. T. (2000). Problem solving as a means toward mathematics for all: An exploratory look through a class lens. *Journal for Research in Mathematics Education, 31*(4), 454–482. doi:10.2307/749653

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale." *Educational Assessment, 13*(4), 267–300. doi:10.1080/10627190802602541

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M. D., Steele, M., & Resnick, L. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

McClain, K. (2002). Teacher's and students' understanding: The role of tool use in communication. *Journal of the Learning Sciences, 11*(2/3), 217–249.

McClain, K., & Cobb, P. (1998). The role of imagery and discourse in supporting students' mathematical development. In M. Lampert & M. Blunk (Eds.), *Talking mathematics in school: Studies of teaching and learning* (pp. 56–81). New York, NY: Cambridge University Press.

Moschkovich, J. (1999). Supporting the participation of English language learners in mathematical discussion. *For the Learning of Mathematics, 19*(1), 11–19.

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

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Retrieved from http://www.corestandards.org

Silver, E. A., Smith, M. S., & Nelson, B. S. (1995). The QUASAR project: Equity concerns meet mathematics reforms in the middle school. In W. G. Secada, E. Fennema, & L. B. Adajian (Eds.), *New directions in equity in mathematics education* (pp. 9–56). New York, NY: Cambridge University Press.

Smith, M. S., Bill, V., & Hughes, E. K. (2008). Thinking through a lesson: Successfully implementing high-level tasks. *Mathematics Teaching in the Middle School, 14*(3), 132–138.

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. doi:10.1080/10986060802229675

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. doi:10.1080/1380361960020103

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

Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning trajectory based instruction: Toward a theory of teaching. *Educational Researcher, 41*(5), 147–156. doi:10.3102/0013189X12442801

Tate, W. F. (1995). Returning to the root: A culturally relevant approach to mathematics pedagogy. *Theory into Practice, 34*(3), 166–173. doi:10.1080/00405849509543676

Thompson, A. G., Philipp, R. A., Thompson, P. W., & Boyd, B. (1994). Calculational and conceptual orientations in teaching mathematics. In A. Coxford (Ed.), *Professional development for teachers of mathematics*, 1994 yearbook of the National Council of Teachers of Mathematics (pp. 79–92). Reston, VA: NCTM.

Thompson, P. W. (1996). Imagery and the development of mathematical reasoning. In L. P. Steffe, P. Nesher, P. Cobb, G. Goldin, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 267–283). Hillsdale, NJ: Erlbaum.

Van de Walle, J. A., Folk, S., Karp, K. S., & Bay-Williams, J. M. (2010). *Elementary and middle school mathematics: Teaching developmentally*. Upper Saddle River, NJ: Pearson Education.

Authors

Kara Jackson, Department of Integrated Studies in Education, McGill University, 3700 McTavish Street, Montreal, QC H3A 1Y2, Canada; kara.jackson@mcgill.ca

Anne Garrison, Department of Teaching and Learning, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203-5721; a.garrison@vanderbilt.edu

Jonee Wilson, Department of Teaching and Learning, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203-5721; jonee.wilson@vanderbilt.edu

Lynsey Gibbons, Division of Curriculum and Instruction, University of Washington, Box 353600, Seattle, WA 98195-3600; lgibbons@uw.edu

Emily Shahan, Department of Teaching and Learning, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203-5721; emily.shahan@vanderbilt.edu

Accepted May 29, 2012