ZDM (2018) 50:445–460
https://doi.org/10.1007/s11858-018-0913-9

ORIGINAL ARTICLE

Studying instructional quality by using a content-specific lens: the case of the Mathematical Quality of Instruction framework

Charalambos Y. Charalambous[1] · Erica Litke[2]

Accepted: 9 January 2018 / Published online: 18 January 2018

Abstract
In this study, we use Mathematical Quality of Instruction (MQI), a content-specific observation framework, to examine the mathematical quality of instruction of three focal lessons in order to examine the instructional aspects illuminated by this framework as well as discuss those aspects not captured by MQI. While prior work provides evidence on the validity and reliability of the MQI measures, no prior work systematically explores the strengths and limitations of MQI in capturing instructional quality. Our analysis points to the affordances of MQI for highlighting differences within lessons across instructional dimensions related to the mathematics of the lesson, as well as for comparing across lessons with respect to the depth and quality of the mathematics instruction provided to students. We discuss how the depth of information provided by MQI may guide instructional improvement efforts. In addition, we explore three categories of instructional aspects not highlighted when examining instruction through the lens of MQI, addressing areas in which MQI in particular, and observation instruments in general, might be limited in their capacity to support teachers in instructional improvement efforts.

Keywords Content-specific framework · Elementary grades · Instructional quality · Mathematical quality · Mathematics teaching · Observational instruments

1 Introduction

One way to answer the question, "What constitutes good teaching in mathematics?" is to identify teaching that promotes student learning. Yet how exactly teachers are able to do this has traditionally been difficult to determine. Scholars interested in this question have pursued at least two different paths. Historically, researchers have first examined what aspects of teaching *in general* contribute to student learning (e.g., Brophy and Good 1986; Creemers and Kyriakides 2008; Muijs and Reynolds 2000), which resulted in the gradual genesis of an evidence-based list of generic teaching factors (e.g., maximizing student learning time, presenting information in structured and coherent ways, asking both process and product questions, providing students with timely and descriptive feedback; see more in Muijs et al. 2014). More recently, however, attuned to Shulman's (1986) plea to attend to both subject matter and the requirements demanded of teachers in teaching particular content, other researchers have attempted to understand the content-specific characteristics of instruction in mathematics that make the teaching of mathematics distinct from teaching other subjects (e.g., Boston 2012; Learning Mathematics for Teaching (LMT) 2011; Walkowiak et al. 2014). This has resulted in the development of mathematics-specific frameworks focused on teaching factors such as the use of precise mathematical language or linking between representations as a means to help students grasp abstract mathematical ideas.

One such mathematics-specific framework, Mathematical Quality of Instruction (MQI), reflects scholarly attempts to examine content-focused aspects of teaching mathematics. Its developers highlight the importance of directly examining the *mathematical* quality of instruction (LMT 2011). The MQI has been used by researchers to describe instruction, better understand the relationship between teacher knowledge and instructional practice, and explore the relationship between instructional characteristics and student outcomes. Any observation protocol, however, has a particular lens through which it views instruction and there are likely

✉ Charalambos Y. Charalambous
cycharal@ucy.ac.cy

1 Department of Education, University of Cyprus, Theophanides Building, Rm. 508, 11-13 Dramas Str., 1077 Nicosia, Cyprus

2 University of Delaware, Newark, DE, USA

important elements of instruction that are not captured by MQI. By analyzing a small sample of lessons using MQI, we are able to contextualize what characteristics of instruction MQI makes salient. Additionally, by looking both within and across lessons, we also illuminate aspects of instruction that MQI does not bring to light. In doing so, we explore both the affordances and limitations of capturing instructional quality through this particular lens.

2 Theoretical underpinnings and empirical findings related to MQI

2.1 Mathematical Quality of Instruction: conceptualization and operationalization

In developing the MQI framework and its associated observation protocol, researchers followed both a top-down and a bottom-up approach. In addition to focusing on instructional features that contribute to student learning in mathematics (e.g., Borko et al. 1992; Ma 1999; National Council of Teachers of Mathematics 2000; Stigler and Hiebert 1999; Thompson and Thompson 1994), researchers drew on a close analysis of a large sample of video-recorded elementary mathematics lessons, as well as their own experiences in teaching and studying mathematics instruction. The development of the MQI framework followed an iterative process involving cycles of fine-grained observations of video-recorded lessons and theoretical discussions of relevant literature on teaching mathematics and on the knowledge needed to teach this subject matter (see LMT 2011 for a full description of this process). Guided by Cohen et al.'s (2003) conceptualization of the instructional triangle,[1] the developers conceptualized instruction as comprised of dynamic interactions among the teacher, the content, and students situated in educational settings. This conceptualization, along with psychometric analyses to refine constructs (see LMT 2011 for discussion), provided a scheme that guided the organization of MQI into dimensions with each dimension comprising a number of items (Hill 2010; Hill et al. 2018; see also; Lynch et al. 2017 for a discussion of the theoretical underpinnings of each dimension). Since its development, the MQI framework has gone through different iterations. In its current form, it includes four main dimensions with twenty items (see Table 1 for a brief description of MQI and its scoring criteria). Two of the MQI dimensions—*Mathematical Richness* and *Mathematical Errors and Imprecisions*—are situated in the relationship between teacher and content, while *Common Core Aligned Student Practices* focuses on the interaction between the students and the content. Finally, *Working with Students and Mathematics* attends to the relationship between the teacher and the students by focusing on how the teacher facilitates students' interactions with the mathematics; as such, it focuses on the ways in which the teacher hears and responds to students' thinking and contributions in the context of the lesson.

[1] Other scholars (e.g., Hawkins 2002) have also proposed a triadic relationship between the teacher, the student and the subject-matter, but the instructional triangle in the form of interactions that take place in and are shaped by particular contexts has been popularized by Cohen et al. (2003).

2.2 Empirical support for MQI

2.2.1 Measuring teaching and teacher quality with MQI: validity studies

MQI has been used by researchers in several studies to examine the relationship between instructional quality and other characteristics typically associated with teacher and teaching effectiveness. For example, multiple studies (e.g., Hill et al. 2008, 2012c, 2015) have explored associations between teachers' quality of instruction and their performance on assessments of mathematical knowledge for teaching (MKT)—a distinct type of knowledge "needed to perform the recurrent tasks of teaching mathematics to students" (Ball et al. 2008, p. 399). In two studies, one focused on ten elementary school teachers (Hill et al. 2008) and the other focused on 24 middle school teachers (Hill et al. 2012c), correlations between MKT and the MQI dimensions ranged from 0.30 to 0.80 across both studies. These associations were stronger and more evident for teachers in the upper and lower tails of the MKT distribution. In a more recent study with a larger sample of 272 fourth- and fifth-grade teachers (Hill et al. 2015), associations between MQI and teacher knowledge were somewhat lower (ranging from 0.30 to 0.40). This may be due in part to the fact that the teacher knowledge assessment included both MKT and mathematics content knowledge items (i.e., items capturing teachers' knowledge of the content instead of their knowledge to teach this content). Of all teacher background characteristics included in the study (e.g., certification route, coursework, degree, efficacy beliefs), teacher knowledge was the only predictor significantly associated with teachers' MQI scores—thus providing concurrent validity evidence supporting that MKT can aid teachers in offering quality instruction, as captured by MQI. Equally important, teacher knowledge did not relate to content-generic measures of instructional quality. There was no statistically significant association between teacher knowledge and scores on instructional dimensions measured by the Classroom Assessment Scoring System (CLASS), a content-generic

Table 1 Description of the four main dimensions of the Mathematical Quality of Instruction and scoring criteria

Dimension	Description	Codes
1. Richness of the Mathematics	Captures the depth of the mathematics offered to students. Rich mathematics can be established by either focusing on the meaning of facts and procedures (e.g., through linking representations or providing explanations) or by focusing on key mathematical practices (e.g., working on multiple solutions, developing generalizations, using appropriate mathematical language).	a. Linking and connections: explores if explicit connections are drawn between representations of mathematical concepts b. Explanations: examines whether mathematical explanations are offered by the teacher and/or the students c. Mathematical meaning and sense-making: captures the extent to which instruction is geared toward supporting students to make meaning of mathematical ideas d. Multiple procedures or solution methods: investigates whether teacher/students use(s) and discuss(es) multiple approaches to a given problem e. Patterns and generalizations: examines the extent to which teacher/students identify/ies patterns and use(s) them to develop generalizations f. Mathematical language: captures the depth and density of the mathematical language used during instruction g. Overall code: a holistic code that captures the overall depth of the mathematics offered to students
2. Errors and Imprecision	Explores the degree to which instruction is mathematically incorrect.	a. Mathematical content errors: captures instances in which instruction includes major mathematical errors b. Imprecision in language and notation: includes instances of teachers imprecise mathematical language or notation that interfere with the presentation of the content c. Lack of clarity in presentation of mathematical content: captures instances when the presentation of mathematical ideas is unclear or confusing, thus obscuring the mathematical content d. Overall code: a holistic code that captures the overall presence of teacher errors in doing and talking about mathematics
3. Working with Students and Mathematics	Refers to teachers' ability to appropriately interpret and respond to students' mathematical ideas and errors, as well as their capability to capitalize on these ideas and errors during instruction.	a. Remediation of student errors and difficulties: examines the degree to which the teacher responds to student errors or misunderstandings and offers procedural/conceptual remediation b. Teacher uses mathematical contributions: captures whether and how the teacher responds to and builds on students' mathematical productions to steer the lesson toward a mathematical goal c. Overall code: a holistic code that provides an evaluation of the teacher–student interactions around the content

Table 1 (continued)

Dimension	Description	Codes
4. Common Core Aligned Student Practices	Focuses on student involvement with the mathematical content through cognitively demanding activities.	a. Students provide explanations: assesses students' explanations of ideas, procedures or solutions
		b. Student mathematical questioning and reasoning: explores instances of students engaging in meaning-oriented mathematical practices, such as posing mathematically motivating questions, making hypotheses, offering examples and counter-examples, etc.
		c. Students communicate about the mathematics of the segment: captures the extent to which students contribute mathematically to the lesson
		d. Task cognitive demand: examines student engagement in tasks in which they think deeply and reason about mathematics; this code refers to the *enactment* of the task, regardless of the initial demand of the curriculum/textbook task or how the teacher sets up the task for students.
		e. Students work with contextualized problems: captures the extent to which students work on problems situated in real-world contexts
		f. Overall code: a holistic code capturing evidence of students' involvement in the mathematics of the lesson and the extent to which students participate in and contribute to meaning-making and reasoning

Score points: 1 (element not present), 2-Low (brief or pro forma enactment of a given element), 3-Mid (more than brief instance(s) of a given element but without element-specific features that would indicate a score of high) 4-High (substantive and detailed enactment of a given element) instrument. This provides divergent validity evidence as we might expect these constructs to be unrelated.

Other studies (e.g., Hill et al. 2012a, 2015) have shown that correlations between MKT and MQI vary by both district context and curriculum, pointing to two potential moderators of the relationship. This moderating effect of curriculum was also corroborated in a series of multiple-case studies (Charalambous and Hill 2012). A validation study conducted in Ireland (Delaney 2012) also pointed to the importance of greater alignment between the mathematics content examined in MKT and the topics taught in the lessons observed by MQI. Qualitative studies have further examined the relationship between MKT and MQI, finding that teachers with higher MKT scores provided richer mathematical explanations, more appropriately used and connected representations, and highlighted key mathematical ideas; unsurprisingly, these teachers also committed fewer and more trivial mathematical errors or imprecisions (e.g., Hill et al. 2008).

Both large-scale studies (e.g., Kane and Staiger 2012) and smaller-scale studies (e.g., Blazar and Kraft 2017; Blazar et al. 2016) have pointed to a positive trend between MQI scores and student learning, although with somewhat mixed results. The Measures of Effective Teaching (MET) project, the largest of these studies, found positive significant associations (ranging from $r=0.12$ to $r=0.16$) between teachers' MQI scores and students' scores on both state tests and a more cognitively demanding project-administered test (Kane and Staiger 2012); although in the expected direction, the magnitude of these correlations was low, something that could (at least partly) be attributed to the use of a truncated version of MQI in the MET project. Stronger associations were found in smaller-scale studies. For example, Hill et al. (2011) drew on a sample of 222 middle-school teachers and their students to estimate teacher value-added scores and then correlated these scores with the MQI performance of 24 teachers. They found a positive relationship between teachers' value added scores and their scores on MQI. These correlations varied from $r_{\text{rho}}=0.30$ in a school-fixed-effects model to $r_{\text{rho}}=0.56$ in a simple value-added model that adjusted for student prior achievement and also included teacher random effects. Another study (Blazar 2015) that accounted for the non-random sorting of students to 111 teachers in three districts found that ambitious instruction, as measured by the *Richness* and *Common Core Aligned Student Practices* dimensions, positively predicted fourth- and fifth-grade students' scores on a low-stakes mathematics test($b=0.114$, $SE=0.044$, $p<0.05$) ; in contrast, other, more generic instructional dimensions such as classroom organization and classroom emotional support did not significantly predict student learning. A recent study using data from 310 fourth- and fifth-grade teachers and their 10,575 students found relationships between teachers' MQI scores and both cognitive and non-cognitive aspects (i.e., self-efficacy and happiness) of student learning (Blazar and Kraft 2017). For example, the extent to which teachers commit mathematical errors during instruction was found to be negatively associated with students' performance on a high-stakes test ($b=-0.024$, $SE=0.013$, $p<0.10$), students' self-efficacy ($b=-0.094$, $SE=0.033$, $p<0.01$), and happiness in the class ($b=-0.181$, $SE=0.081$, $p<0.05$).

Researchers also investigated the construct validity of MQI by assessing the degree to which observations scored with MQI differed from observations scored with a more generic framework (CLASS). Using data from more than 2000 video-recorded lessons scored with both instruments, Blazar et al. (2017) found the MQI items formed factors separate from those formed by the CLASS items. These results held regardless of the analyses (simple and bi-factor models) run. In each case, the *Errors and Imprecision* items consistently formed a distinct factor, while the remaining MQI items were grouped under an overarching factor representing what many scholars have described as "ambitious instruction" (cf. Cohen 2011).

Other research has provided convergent and divergent validity evidence. Researchers from the MET study compared teachers' performance on generic observation rubrics to their scores on content-specific observation instruments (Kane and Staiger 2012). As expected, MQI related more strongly with another instrument that also incorporated mathematics-specific elements, the UTeach Observation Protocol, UTOP ($r_{\text{MQI-UTOP}}=0.85$) than with two instruments that attended only to generic aspects of instruction ($r_{\text{MQI-CLASS}}=0.69$ and $r_{\text{MQI-Framework for Teaching}}=0.67$). That these instruments were found to have strong correlations might partly be due to the fact that all have been designed to capture the same underlying construct—instructional quality. However, the relatively small differences in the magnitude of the correlations could also suggest that these instruments differ at least slightly in the lens (i.e., the specific instructional aspects) through which they capture this construct.

2.2.2 Reliability studies

Using generalizability theory, scholars explored the scoring design needed to obtain sufficiently reliable estimates of instructional quality with respect to each of the MQI dimensions (Hill et al. 2012a, b). Results suggested that a combination of three lessons coded independently by two raters was sufficient for obtaining reliable estimates of teaching quality at a level of $\rho=0.70$ for three of the MQI dimensions; the only exception was *Working with Students and Mathematics* for which a combination of either four lessons and three raters or three lessons and four raters was needed to obtain sufficiently reliable estimates. Collectively, these results challenged a typical approach in teacher evaluation requiring a single rater to observe and score only a couple of lessons per teacher. Further explorations showed that only *Common Core Aligned Student Practices* was sensitive to the length of observation (Hill et al. 2012b): in this dimension, watching only the first 30 min of a lesson (as opposed to watching the entire lesson) yielded notably lower estimated reliabilities; for the other dimensions, reliabilities largely remained unchanged. Additional analyses suggested a positive impact of rater selectivity and training on the reliability of the MQI ratings and no impact of the lesson content (arithmetic/algebra vs. geometry/measurement) on reliability (Hill et al. 2012a).

2.3 Studies capitalizing on MQI as a tool for professional development

Though created as a tool for researchers to measure and analyze instructional quality, MQI has more recently been used for professional development purposes. For example, it has been piloted as a framework for web-based video coaching cycles in which teachers analyze and reflect upon their instruction. Preliminary findings from an experimental study suggest that teachers who received coaching with MQI improved significantly in all dimensions with the exception of Errors and Imprecision (Kraft and Hill 2017). In addition, the students reported that their teachers asked more substantive questions, required more use of mathematical vocabulary, and provided more opportunities for mathematical discussions (Hill et al. 2016). Other researchers introduced pre-service teachers to MQI as a tool with which to analyze video of their own and their classmates' lessons (Mitchell and Marin 2015). While these novice teachers initially focused on issues of classroom climate and management, after using MQI as a lens, they became more attentive to classroom features related to the mathematical quality of their instruction.

3 Research aim and questions

As described above, previous research has provided empirical support concerning the validity and the reliability of MQI, while also providing some positive, although preliminary, evidence about its potential to support teacher learning. Past research, however, has not systematically and explicitly examined the affordances and limitations of MQI in capturing instructional quality. Recognizing that any given observation instrument functions as a magnifying lens that can highlight certain aspects of the phenomenon under exploration but leave others less visible, in this manuscript we use the MQI to examine three elementary-school lessons and ask:

1. *What aspects of instruction does MQI highlight and what is the quality of these aspects?*
2. *What instructional aspects are not highlighted by examining instruction using MQI?*

By adopting a reflective stance, we aim to offer insight into both the strengths and limitations of MQI—something that aligns with its dynamic and evolving character over the past decade.

4 Methods

4.1 Lessons analyzed

The three lessons examined in this study were all taught by fourth-grade teachers in the US and are part of the video library of the National Center of Teacher Effectiveness at the Harvard Graduate School of Education. The first lesson in the sample was a geometry lesson taught by Mr. Smith, the second was a lesson on strategies for multiplication taught by Ms. Young, and the last, taught by Ms. Jones, focused on multiplying a fraction by a whole number (all teachers' names are pseudonyms; for a complete description of the lessons see Charalambous and Praetorius 2018).

4.2 Data analysis approach

Each lesson was scored independently by each author, both of whom are trained, certified MQI raters.[2] Like Hill et al. (2008, 2018), we scored each lesson in 7.5 min segments, pausing after each segment to assign a score for each of the items listed in Table 1. In addition, each rater assigned an overall score for each MQI dimension to each segment (e.g., Overall Richness, Overall Errors, etc.). This latter score is a holistic appraisal of the segment with respect to each dimension and does not represent an average of the scores assigned to the codes of the dimension at hand. Segments were scored on a Likert-scale from 1 (not present) to 4 (high). In addition to scoring each segment on all elements, both raters assigned a holistic score to the lesson as a whole, assessing its overall quality (*Overall MQI)* using a Likert scale ranging from 1 (poor and problematic instruction) to 5 (high quality instruction in terms of combinations of characteristics from different dimensions). Like the holistic scores assigned to each segment, this lesson score does not represent an average

[2] Certified MQI raters first undergo training on how to score with the instrument and then undergo a certification process. Raters are asked to score a set of selected lesson segments. Raters whose scores coincide with MQI master scores are certified to independently score lessons using MQI.

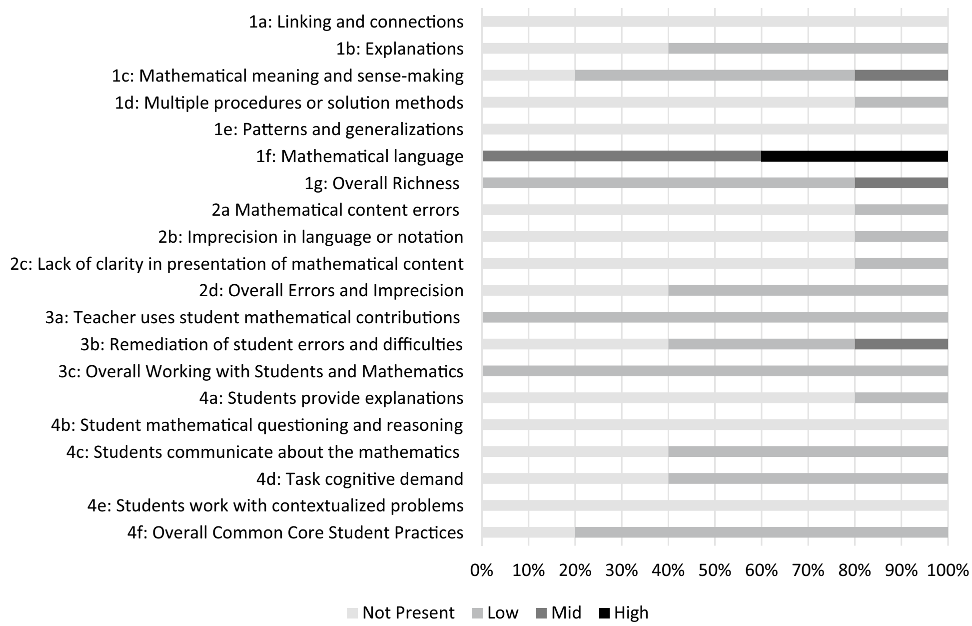

Fig. 1 Percentage of segments of Mr. Smith's lesson scoring each of the four levels on each MQI item

of the segment or dimension scores; rather it gives raters the opportunity to evaluate the lesson instructional quality holistically, following a set of benchmarks corresponding to different combinations of dimensional characteristics. After watching and scoring each lesson, each author also wrote a lesson summary that included a narrative description of the lesson, its mathematical strengths and weaknesses, and other salient features of the lesson not captured by MQI.

Having scored each lesson with MQI and generated a lesson summary, we then reconciled our scores. We discussed divergent ratings to generate a consensus view of lesson quality. We note that for the purposes of this analysis, we were not aiming to measure teachers' instructional quality, but rather understand the affordances and limitations of viewing instruction with MQI. Hence, we discourage any generalizations regarding these teachers' instruction from simply viewing a single lesson. We scored lessons to illuminate themes rather than to assign definitive scores and used discussions of where ratings diverged as opportunities to make sense of instruction. Notes from these discussions and the lesson summaries served as analytic memos (Patton 2002). Using the constant comparative method (Maykut and Morehouse 1994), we coded these analytic memos, first assigning labels to instructional aspects either by using the existing MQI codes or by developing new descriptors for instructional aspects not captured by MQI. We then developed categories to organize these instructional aspects. We discussed and refined these categories, focusing on those with potential to support students' mathematical understanding.

5 Findings

We first summarize the characteristics of each lesson as measured by MQI. Next, we discuss instructional quality across lessons, highlighting the features described by the MQI framework. Finally, we present salient aspects of the focal lessons not captured by MQI.

5.1 Mr. Smith's lesson

Mr. Smith's lesson was characterized by low to mid levels of mathematical richness, low to moderate take up of students' mathematical ideas, little student engagement in *Common Core Aligned Student Practices*, and few errors. This description can be seen in Fig. 1, in which we show the percentage of segments scoring at each level for all MQI items. In addition, both authors rated this lesson a mid on the *Overall MQI*, as this lesson was largely error free and contained good use of precise and accurate mathematical language and some brief instances of mathematical richness. However, there was little substantive student involvement in cognitively demanding work and many elements of richness were absent.

Three items in the *Richness* dimension were notable in Mr. Smith's geometry lesson, although the remaining items were largely absent. The lesson featured strong use of mathematical language, with all segments scoring either a mid or high on this code. Mr. Smith's instruction included not only consistent use of mathematical language, but an explicit focus on definitions and on pressing students to use appropriate mathematical terms. In addition, the lesson featured some work around mathematical meaning and sense-making, although this was not the focus of instruction. While merely stating geometric definitions would not have been sufficient to score above not present on this code, Mr. Smith did work to make meaning of the definitions and geometric properties, both briefly and at times more than briefly. For example, at one point he presented students with a picture of an obtuse angle and asked them to reason about its measurement. He discussed with students how they would use their understanding of what a 90° angle looked like to recognize that the measure under consideration was less than 90° and to reason from there to an estimate. Mr. Smith also gave some brief explanations, such as why particular angle measures are obtuse or acute. However, these explanations were not fully developed, sustained, nor a focus of the mathematical work.

When working with students around the mathematics, Mr. Smith briefly remediated a few student errors and engaged with student mathematical ideas in largely perfunctory ways, as evidenced by the low scores across segments. For example, in discussing the different classifications of angles, he introduced reflex angles by drawing an example on the board and clarified:

Mr. Smith	I'm going past 180… Anybody have any idea what you think the name of that angle is?... You probably haven't seen it before… This is what we call a reflex angle… So, if my right angle is 90 degrees, acute angle is less than 90, obtuse is between 90 and 180, a straight is 180. How do we describe this one?
Multiple Students	[Multiple comments.]
Mr. Smith	Well, it's going to be greater than what?
Multiple Students	One hundred and eighty.
Mr. Smith	It's going to be greater than 180. What does it have to be less than, though?
Multiple Students	[Multiple comments.]
Mr. Smith	Why 360?
Multiple Students	Because that's the whole rotation.
Mr. Smith	Okay. That's the full rotation.

In this example, Mr. Smith acknowledged correct responses, but did not use them to move the mathematics of the lesson along. Indeed, throughout the lesson, he rarely built upon students' productions.

While the instruction in this lesson was largely clear and error-free, there were some brief mathematical issues. For example, at one point, Mr. Smith led the class in classifying angles as acute, obtuse, reflex, or right by sliding them across a SmartBoard screen into the appropriate category. In the process of doing so, Mr. Smith placed a picture of a reflex angle into the category marked obtuse and placed the angle measurement of 156° into the category marked reflex. While this error was captured in the *Mathematical Content Errors* code, it arguably did not obscure the mathematical point of the segment in the context of the lesson. Similarly, we noted instances in which Mr. Smith's language was briefly mathematically imprecise (drawing rays but calling them "line segments" in passing) or where his presentation of the mathematical content was briefly unclear. Overall, however, the mathematical elements of his instruction were accurate and clear. We note these errors not because we wish to make claims about Mr. Smith' mathematical content knowledge, but because they are indicative of the types of issues in the presentation of mathematical content that MQI is able to capture.

Finally, this lesson was also characterized by minimal student engagement in mathematical ideas, as measured by the *Common Core Aligned Student Practices* dimension. Students contributed to the mathematics of the lesson, but they did so primarily with brief answers or the occasional offer of a solution. There were very few student mathematical explanations, and those that occurred were brief (e.g. "Because that's the whole rotation" as an explanation for why a reflex angle is less than 360°). More typical were short responses to bounded questions as indicated in the example above. In addition, there were no instances of students asking mathematically motivated questions, making conjectures, providing counterclaims or any similar behaviors, as evidenced by all segments scoring not present on *Student Mathematical Questioning and Reasoning*. While students completed the tasks asked for by the teacher, they did not engage at high levels with the mathematics and the mathematical work asked of them primarily required recall and classification.

5.2 Ms. Young's lesson

Ms. Young's lesson was characterized by mid to high levels of mathematical richness, consistent take-up of students' mathematical contributions in the development of the lesson, as well as some opportunities for students to engage in *Common Core Aligned Practices*. This was reflected in an overall MQI score of mid/high, but also can be seen through scores across each dimension. Figure 2 displays the percentage of segments in Ms. Young's lesson that scored at each of the four levels of MQI by item.

Ms. Young's instruction focused on making meaning of multiplication in important ways. Across segments, Ms. Young's instruction was driven by a sustained focus on meaning-making. Over half of the segments scored high on *Mathematical Meaning and Sense-Making*, reflecting a consistent focus on the meaning of the multiplication strategies discussed in the lesson. For example, in the beginning of the lesson, she asked students to examine 16×3 and 16×6 and determine how the two problems are related. Rather

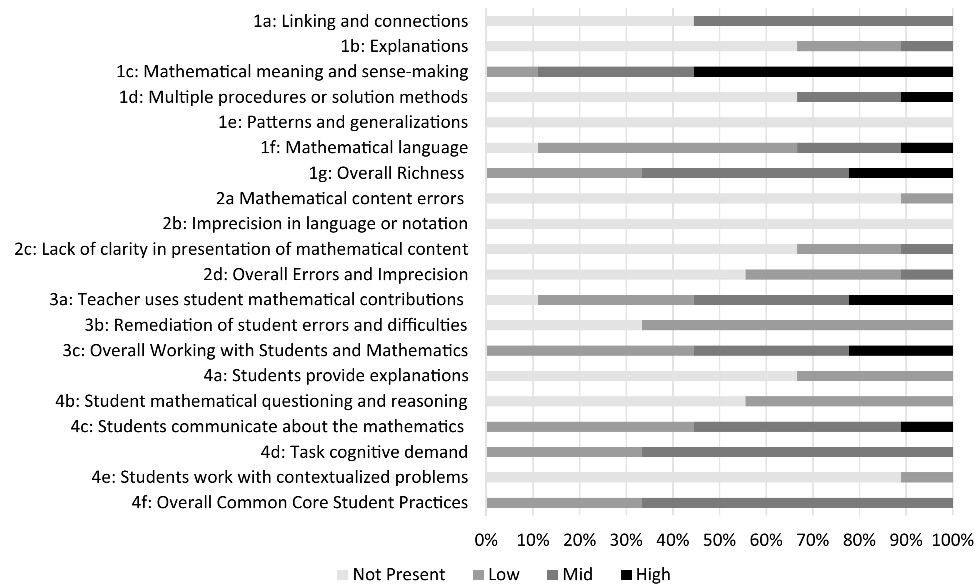

Fig. 2 Percentage of segments of Ms. Young's lesson scoring each of the four levels on each MQI item

than simply focusing on the multiplication algorithm, Ms. Young's instruction encouraged students to examine the relationship between the solutions and the respective factors. She asked students how the two expressions were related and a student responded:

Student	…3 plus 3 gets you 6, so the answer plus the answer again will give the next one down there.
Ms. Young	The answer of what?…
Student	The answer of 16 times 3 and plus another 16 times 3 will give you the answer down here.
Ms. Young	Did anybody hear what he said?
Student	Yes.
Ms. Young	Can you repeat it in a different way? Yes?
Student	I think he said that if you get the answer of 16 times 3 and you double the answer of 16 times 3, you get the answer of 16 times [6].
Ms. Young	So is that true?
Multiple Students	Yes.
Ms. Young	So why would it be true? Yes, Student Z?
Student	'Cause if you're doing an answer it'll get—you put into a *[inaudible]*. Sometimes if you double the bigger number, it's gonna stay the same.
Ms. Young	I want us to use [the] specific example that we're talking about, before we make a general statement of what is going to be true for all numbers or some class of numbers at least… He said that if you doubled the answer of 16 by 3, it would give us the answer of 16 by 6. Can we explain that using either a story context or drawing or an array any way?

Here, Ms. Young focused the students not on the solutions to the multiplication problems but on the relationships inherent in multiplication as an operation (that doubling a factor will double the solution). She pressed the students to use the examples to think in general terms and encouraged them to further make sense of the idea using multiple representations.

Throughout the lesson, Ms. Young utilized multiple representations (e.g., symbolic, array model) and engaged in careful linking across these representations. For example, after formalizing the idea described above, she used an array model that a student had produced of 16×6 to show the two copies of 16×3. She highlighted the 16×6 array and said:

> He has two rectangles [inside the array of 16×6]. So this one would be the 3 by 16, and this one is the 3 by 16. And when you combine the two 3 s you get 6. So Student Y knows how to… show that there are two 3 by 16 inside 6 by 16.

Ms. Young also drew frequent and explicit, though brief, links highlighting the connections between the models for multiplication and their corresponding number sentences. In addition, in several segments she focused on connecting multiple solution methods for the multiplication problems at hand. For example, she engaged in a brief discussion of when the doubling and halving method would be efficient, highlighting the utility of having multiple approaches for multiplication problems. Finally, some of the segments in

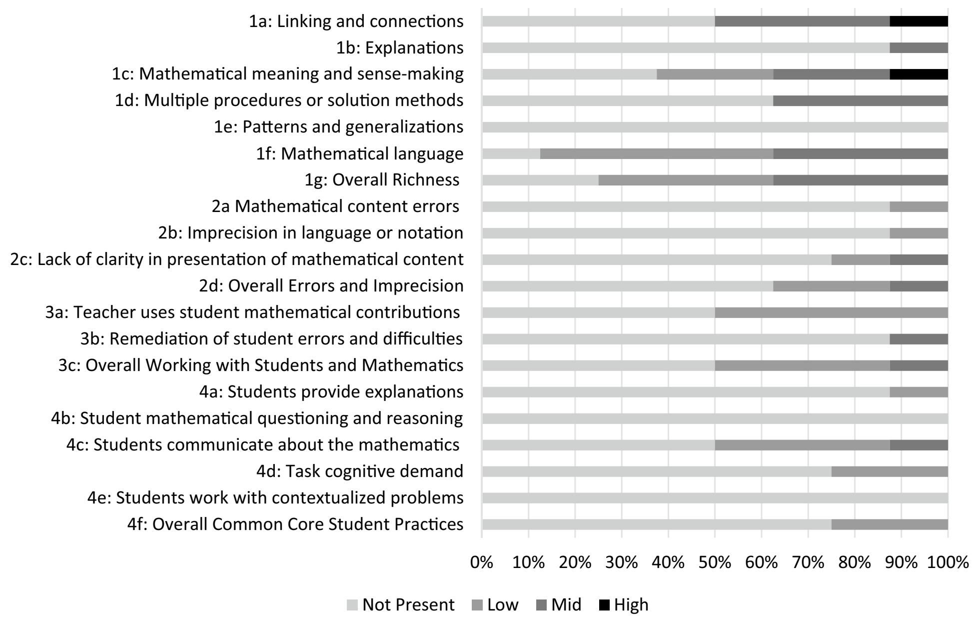

Fig. 3 Percentage of segments of Ms. Jones' lesson scoring each of the four levels on each MQI item

the lesson contained brief mathematical explanations, such as why the product of 16×6 is double the product of 16×3. With the exception of minor instances that did not detract from the mathematics, the lesson was clear of content errors and imprecisions.[3]

In this lesson, Ms. Young often capitalized on student contributions to help develop mathematical understanding. Her instruction included not only take up of student ideas, but the interweaving of these ideas into the development of the mathematics in the lesson, as reflected in multiple segments scoring mid or high on *Teacher Uses Students Mathematical Contributions*. In the vignette described above, she elicited the idea of doubling from a student and then asked other students to repeat the contribution to clarify the mathematical idea. Throughout the lesson, she asked multiple students to share their solutions to various problems, made explicit connections between students' solutions, and highlighted key features of students' mathematical work for the class.

Importantly, and in sharp contrast to the other two lessons under study, this lesson included significant opportunities for students to engage in cognitively demanding work. For example, Ms. Young introduced the doubling and halving strategy for multiplication using 30×4 and 15×8. Rather than teach students the procedure, Ms. Young focused on the ways in which the various representations of the two expressions 30×4 and 15×8 were equivalent, allowing students to unearth and make meaning of the relationships between them. Some students were asked to build the two expressions with unit cubes, others were asked to draw an array, while others were asked to develop a story context. As she circulated the room while students worked, she asked questions that pushed students to develop their ideas about the meaning of the relationship between the expressions. Yet, at times Ms. Young heavily scaffolded students' thinking, resulting in variable levels of cognitive demand. Still, in many lesson segments students were given some opportunity to make meaning of mathematical ideas themselves.

5.3 Ms. Jones' lesson

Ms. Jones' lesson was characterized by occasional elements of mathematical richness, however it was notable for how few opportunities students had to engage in mathematical reasoning, explanations, or other similar practices. Instead, Ms. Jones—rather than the students—did the majority of the cognitive work of the lesson and did little to solicit or take up students' mathematical contributions. This was reflected in some mid and high scores in the *Richness* dimension, but multiple scores of not present across the dimensions focused on student engagement with the content. This is evident in Fig. 3, which shows the percentage of segments at each score point for each MQI item. In addition, although the lesson was relatively free of mathematical errors and imprecisions, Ms. Jones' presentation of the content lacked clarity—especially when she was trying to contextualize abstract mathematical ideas—in ways that may have obscured the mathematical point for students.

Across segments, there were moments in which elements of *Richness* figured prominently. In particular, Ms.

[3] We acknowledge that in some instances, the class was not following the multiplication convention $N\times M$ as repeating M-size groups N times. Given the objectives of this lesson and ambiguity around whether not following this convention is mathematically problematic, we did not score for it.

Jones engaged in linking between representations in half of the lesson segments at a mid or high level. For example, she modeled $5 \times ¾$ by drawing a visual representation and linking the ways in which five copies of the quantity of ¾ appeared in the diagram. She also worked to make sense of the multiplication of a fractional quantity by a whole number, developing students' understanding of the operation of multiplication as repeated addition. Drawing five sets of four circles, and shading three circles in each, she addressed the class, saying:

Ms. Jones [M]ultiplication is repeated addition. Right? ... Didn't your teacher tell you that in second and third grade? That another way to say multiplication is repeated addition. … 2 times 3 is like saying 2 plus 2 plus 2. I add 2 together 3 times. Yes?

Student Yes.

Ms. Jones So when you were doing multiplication, it's still just repeated addition, except this time instead of adding together 2 plus 2 plus 2, you're adding together three-fourths plus three-fourths plus three-fourths.

Here, Ms. Jones engaged the students in thinking about multiplication with a fraction using the same conceptual definition of multiplication as they had with whole numbers, helping to give mathematical meaning to the procedure. Other elements of the *Richness* dimension were present to some degree as well. A main goal of Ms. Jones' lesson was to consider three ways to multiply a number by a fraction. While in approximately one-third of segments, multiple methods were present and briefly compared, no segments scored high on this code, indicating that an explicit comparison of the methods was not taken up in the lesson.

One notable feature of instruction in this lesson was that Ms. Jones did the majority of the mathematical work, and in fact most of the talking. Students rarely contributed to the mathematics and Ms. Jones did little to weave their ideas into the development of the mathematics. Reflecting this, the majority of segments scored not present across codes for *Working with Students and Mathematics*. We noted brief instances of remediation while Ms. Jones circulated among students working in groups, however, this type of interaction was rare and the lesson was characterized by high amounts of teacher talk and little student contribution. In addition, Ms. Jones engaged in the majority of the mathematical thinking in the lesson. She frequently modeled or demonstrated procedures and students were simply asked to repeat them. For example, Ms. Jones demonstrated the procedure for multiplying a whole number by a fraction and then asked students to create their own, similar problem to solve. While this activity had potential for students to engage with the mathematics in meaningful ways, Ms. Jones heavily scaffolded the work, for example directing students that their whole number had to be between one and four and their fraction had to be in fourths. Interventions such as this were common throughout the lesson. Notably, there was not a single instance of student mathematical questioning and reasoning and only one instance of a student providing a brief mathematical explanation. Otherwise, student participation consisted of executing previously modeled procedures, at times with step-by-step directions from Ms. Jones, and contributing one- or few-word responses to bounded questions. As a result, the lesson scored low across dimensions in *Common Core Aligned Student Practices*.

Finally, the lesson also included two segments in which the presentation of the mathematical content was muddled. Ms. Jones made an effort to connect the mathematical ideas to realistic contexts, but in doing so presented situations that were disconnected from the mathematics and unclear. She told a story of burping her daughter when she was an infant and tried to explain that the burps were analogous to repeated addition. While there was some logic to her analogy, the connection to the mathematics was not made explicit for students. Ms. Jones next talked about her toddler daughter handing her blocks from a pile and again attempted to connect this to the idea of repeated addition. The connections both between the two stories and to the mathematics were difficult to follow. Although this segment of instruction was scored as containing *Lack of Clarity*, it raised a tension for us as to whether the result of the instruction was a somewhat disconnected set of stories or unclear and confusing presentation of the mathematical ideas.

5.4 Comparing across lessons

Looking across lessons, we note that while all three lessons contained strong elements, none of the lessons scored at the highest level of quality on MQI. Both Mr. Smith's and Ms. Jones' lessons scored a mid for *Overall MQI*, while Ms. Young's lesson scored a mid/high. These similar holistic ratings were arrived at for different reasons. Both Mr. Smith and Ms. Jones' lessons were characterized by low to mid levels of *Richness*, few errors, and little engagement in *Common Core Aligned Student Practices*. However, upon closer inspection of individual codes, the quality of the instruction in these two lessons differed in meaningful ways. For example, while Ms. Jones occasionally drew connections between different representations, in Mr. Smith's lessons such connections were totally absent. To some degree, this may be related to the content of the instruction (e.g., Ms. Jones used visual and symbolic representations of fractional quantities while Mr. Smith's lesson focused on defining geometric terms and included geometric classifications). However, other codes were less related to content. *Mathematical*

Explanations occurred more frequently in Mr. Smith's lesson, whereas *Multiple Solution Methods* was more evident in Ms. Jones' lesson. This comparison allows us to think differently about improving instruction in both Mr. Smith and Ms. Jones' classroom. Such comparisons, in fact, hold not only within dimensions, but also across dimensions. For instance, a different lesson might similarly score mid on *Overall MQI* quality, but may include fewer elements of *Richness* but more emphasis on student cognitive demand and mathematical contributions. Looking at each dimension for the explanation for these overall scores is a strength of MQI and provides insight into how those interested in instructional improvement might proceed in working with these (or similar) teachers.

5.5 Instructional areas not highlighted by MQI

Through our analysis, we noted salient aspects of instructional quality that were not as readily captured by the MQI dimensions. These elements fell into three categories: generic instructional aspects, mathematics-specific instructional aspects, and instructional aspects that may be outside the scope of observational instruments more broadly. We discuss each category below.

5.5.1 Generic instructional aspects not captured by MQI

Unsurprisingly, MQI did not surface aspects of the lessons that were related to general pedagogical characteristics. Prior research has shown such aspects to be pivotal for student learning (cf. Muijs et al. 2014) and some have made their way into generic observational instruments (see Berlin and Cohen 2018; Kyriakides et al. 2018). For example, our description above does not surface variation in classroom management and organization. Ms. Jones' lesson, for example, included clear procedures and routines to ensure smooth transitions between activities and maximize student learning time. She employed a variety of techniques and signals and students appeared to easily follow these routines, suggesting that such routines may be well-established in her classroom. In contrast, we noted that transitions from one activity to another in Ms. Young's lesson were not particularly efficient; for example, after working independently on proving that 30×4 is equivalent to 15×8, it took students approximately 3 min to move from their desks to the rug in order to share their work. Another general instructional aspect that we noted in all three classrooms related to how the teachers structured and presented information to support student learning. Ms. Young, for instance, started her lesson by clearly outlining the lesson goal—thus orienting students to what was to follow. Similarly, Mr. Smith and Ms. Jones reminded students of what had been learned in prior lessons and connected the content of the day's lesson to students' prior knowledge. These are considered strong instructional practices, yet MQI is not designed to note their presence (or absence). Finally, there was variability across lessons related to the degree of general student engagement with classroom activities, such as their willingness to participate or remain on task. For example, in some lessons students appeared more engaged than in others and within lessons, we noted that some students worked hard on the assigned tasks while others were off task or disengaged.

5.5.2 Mathematics-specific aspects not captured by MQI

Although MQI was designed to focus on mathematics-specific aspects, no observation protocol can attend to everything occurring in the classroom. These lessons surfaced instructional aspects not captured by MQI, but related to the mathematics of the lesson. For example, one key feature of Ms. Jones' lesson was the clear presentation of mathematical procedures. She outlined the steps for multiplying a fraction by a whole number both verbally and in writing. She also engaged in a meta-narrative, explaining how these steps applied to other, similar problems.

Another mathematical aspect that emerged in this analysis pertained to the use of mathematical tools. Mr. Smith's lesson, in particular, included careful work teaching students how to use a protractor to measure different types of angles. In addition to demonstrating for students how to line up the protractor in order to get the correct angle measure, he oriented students to its correct use and its structure, discussing, for example, why protractors have two sets of measurements and when to use the top versus the bottom set of measurements. The appropriate use of tools is one of the *Common Core Standards for Mathematical Practice* and may be particularly salient for lessons focused on geometry and measurement, such as this one. Hence, although the fourth MQI dimension has evolved to involve certain practices related to the Common Core Standards for Mathematical Practice, this analysis pointed to at least an additional practice that can enrich this dimension.

A third mathematics-specific instructional aspect relates to issues of teaching mathematics equitably. We focus specifically on two aspects of equitable instruction: equitable participation and explicit presentation of mathematical content. In both Mr. Smith and Ms. Young's lessons, certain students were selected to contribute their mathematical ideas publicly. While Mr. Smith asked some students to come to the board and share their ideas, not all students were invited to do so. Similarly, during group work, Ms. Young approached and supported certain groups of students but gave much less of her time to other groups. During the whole-class discussion, particular students were selected to share their solution methods—while others were not. These instructional decisions may have been entirely appropriate, for example informed by students' progress or difficulties, however they raise questions of whether all students have equal opportunities to participate in and learn the mathematics of these lessons. For example, equity concerns would surface if we saw systematic differences in opportunities for participation across lessons that are related to students' gender, race, or ethnicity. However, because for this study, we have access only to the video-recorded lessons (and just one per teacher) and not to student demographic or achievement data, we are unable to analyze patterns of teachers' decisions around participation. Furthermore, in its current form, MQI does not attend to teacher decision-making or measure equitable participation. Those interested in such a measure would need to combine MQI with additional information. In addition to participation, teaching mathematics equitably also pertains to the issue of access, including the explicitness with which the teacher presents the content and launches the tasks so that *all* students can productively engage with the content (cf. Shaughnessy et al. 2015). We noticed different degrees of this explicitness even within the same lesson. For example, Ms. Jones was at times very explicit in outlining the steps for the fraction multiplication; at other times, however, her presentation of the content was muddled, thus apparently causing more difficulties for students struggling with the content.

5.5.3 Instructional aspects difficult to capture through classroom observation instruments

Finally, while classroom observation instruments can shed light on many aspects of instructional quality, they are limited in part by format in their capacity to illuminate certain aspects that are simply difficult to observe. We noted aspects of instruction in the three focal lessons that were not captured by MQI, but would arguably be difficult to capture with any observation instrument.

For example, in two of the lessons, students worked in groups while the teacher circulated the room. Capturing and distilling instructional quality in these moments is incredibly challenging. One limitation of video recorded lessons is that raters often cannot see what students are actually doing, nor can they hear what each student is saying. As such, raters may miss important interactions or other instructional moments. Even when the camera is able to zoom in on particular groups of students, other students are left outside the frame, impeding observers' ability to develop a full picture of classroom interactions. When more than one camera is being used—which was the case for all these three lessons—considering how different groups of students engage with an assigned task still remains a scoring challenge. Which group should drive the scoring of a given segment? This challenge is lessened to some degree in the context of a whole-class discussion or when students share their solution approaches; even then, however, we can only observe (and score) based on the student(s) chosen to present.

We also noted that our scoring provided insight about the *teaching* of mathematics in these lessons, but did not provide solid information on the degree to which students learned the mathematical ideas of a given lesson. Observational instruments by design focus on observable teacher and student behaviors. While such behaviors may serve as a proxy for student learning, they do not directly measure whether the specific teaching has been successful and for whom. The degree to which students learn and engage with the mathematical ideas of a lesson is particularly difficult to capture when students are not afforded opportunities to present their ideas or thinking. MQI, like other instruments, attends to teachers' and students' contributions (and teachers' responses to student contributions) to gauge the *opportunities* afforded to students to make mathematical meaning. However, the degree to which students make use of these opportunities, and the extent to which these opportunities translate into the learning goal of the lesson remains difficult to ascertain. Both Ms. Jones' and Mr. Smith's lessons highlight this issue—while students engaged in several potentially fruitful mathematical tasks, the learning residue from the students' engagement with them was uncertain, especially given such little opportunity for students to share their thinking.

6 Discussion and conclusions

Like in previous studies (e.g., Hill et al. 2008, 2012c, 2018), viewing lessons through the MQI lens afforded insight into content-related aspects of instructional quality. In particular, we noted the extent to which the three lessons afforded mathematically rich environments for their students; the degree to which instruction included mathematical errors or imprecisions; the extent to which students had significant opportunities to productively engage with the content; and the degree to which the teachers effectively facilitated students' interaction with the content. We were not interested in using the scores on these dimensions to rank order teachers (and indeed this would be inappropriate given that we viewed only a single lesson per teacher); rather we believe that scoring these lessons using MQI provides rich information regarding instructional practice along these dimensions. Simultaneously, we acknowledge that there are aspects of instructional quality not highlighted through the choice of this particular framework. These insights are particularly important as we consider how we might use this framework (or any observational instrument) as a tool for working with in-service teachers to improve instruction or for mentoring and supporting pre-service teachers to critically analyze and reflect upon their early teaching experiences.

The information provided by scoring lessons with MQI, particularly the specificity of the delineations of the individual codes, can serve formative evaluation purposes and has potential to scaffold instructional improvement. Looking at the lessons in this study, for example, we find that while Mr. Smith's and Ms. Jones' lessons were both scored mid overall on MQI, there were notable differences between the lessons both across the four MQI dimensions and within given dimensions. The information provided by MQI suggests avenues for coaching these particular teachers toward deepening their instruction. A strength of the MQI is its focus on the mathematics of the lesson. In addition, the specificity of the elements of MQI allows administrators or district personnel to prioritize those aspects of instruction they deem most important for improving student learning in their context, as well as to meet teachers where they are and focus improvement on particular instructional aspects related to the mathematics. In a school or district that valued student-centered classrooms, for example, Ms. Jones might be encouraged to reflect upon the opportunities afforded to her students to productively engage with the content and reason mathematically—with the intention of helping shift the cognitive demand of the lesson onto the students. Another option for improving Ms. Jones' instruction in a context in which teacher-centered instruction is more the norm, might be to focus on deepening the elements of mathematical richness. In considering how to work with Mr. Smith to improve instruction, an administrator might focus on deepening the mathematical richness, but might also consider working with him to both encourage and build more substantively on students' contributions. The choice of focus would depend on the priorities of those engaged in the improvement work. For a lesson such as Ms. Young's, which scored higher overall, the next level of work might have a different focus—administrators might push her to maintain the cognitive challenge in her lesson so as to deepen students' mathematical engagement and increase productive struggle with the content. However, the reader is reminded that a single lesson alone cannot (and should not) inform such decisions (cf. Hill et al. 2012a, b). Additionally, as recent professional development work with teachers has suggested, simply scoring practice with MQI may not be sufficient for improving instruction; in contrast, it is through guided reflection upon their practice and cycles of coaching that teachers begin to notice and improve on particular instructional aspects in their work (Hill et al. 2016; Mitchell and Marin 2015).

Like any lens, MQI brought to the surface some aspects of instructional quality, while leaving others unexplored. The recognition that MQI did not capture more generic aspects of instruction is not surprising. Other studies examining the dimensionality of MQI together with a content-generic instrument found they measure different constructs (Blazar et al. 2017). This suggests that MQI can function complementarily with more content-generic instruments—and in doing so capture and explain more variation in teachers' instruction compared to that captured by each instrument in isolation. Whether to use a content-specific framework or a generic framework (or both) in practice thus depends on what those engaged in the work prioritize and wish to measure. Of more interest, perhaps, are the mathematics-specific aspects of instruction that were salient in these lessons, but were not highlighted by MQI. Ms. Jones' careful work on teaching procedures points to a tension on how to assess the teaching of procedures given the contemporary emphasis on conceptual understanding in mathematics. MQI is focused on conceptually-oriented aspects of mathematics instruction, and yet procedures are still taught. At the high school level, researchers have argued for the importance of procedural knowledge for the development of conceptual understanding; hence, high quality instruction on procedures may be worthy of study (Litke 2015; Star 2005). It is possible that the MQI has gradually evolved into a framework that measures a particular type of instruction (what is often called "ambitious" instruction), and hence places less emphasis on other, potentially important, aspects of instruction less aligned to these ideas.

Mr. Smith's geometry lesson surfaced the importance of considering teachers' use of mathematical tools. Although this feature of instruction may be particularly pertinent to geometry/measurement lessons in which such tools are employed, we suspect that such a focus also be helpful for tapping aspects of instructional quality in other mathematical topics in which students use manipulatives, calculators, or even statistical packages. Given that using appropriate mathematical tools strategically comprises one of the *Common Core Standards for Mathematical Practice* (National Governors Association Center for Best Practices & Council of Chief State School Officers 2010), such a feature of instruction might be considered alongside those in which the teacher facilitates students' interaction with mathematics.

By design, observation instruments must make inferences about students' actual learning of and engagement with the content based on observable behaviors. As such they are ill-suited to measure that which cannot be seen, or seen easily. This presents a measurement challenge, but also a challenge to the inferences that can be drawn from the use of such frameworks. Thus, regardless of purpose, classroom observations need to be complemented with other approaches that gauge students' understanding of the mathematics, such as assessments, student work, or post-lesson interviews with students. Furthermore, our analysis suggested that MQI (like many other observational instruments) attends more to the teacher, focusing on teachers' behaviors and attempts to craft opportunities for student learning and—with the exception of the *Common Core Aligned Practices*—is less attuned to the extent to which students make use of these opportunities.

This is an area in which future development work is warranted so that attention is given to both capturing the opportunities created *and* the extent to which they are capitalized upon. Doing so efficiently represents an open challenge awaiting more innovative ways of capturing instructional quality and examining its effects on student learning.

Like any study, this one is not without limitations. The number of the lessons analyzed, the content and the mathematical topics of these lessons, as well as the single grade level considered may have impacted our analysis and thus the conclusions drawn. Sampling more lessons from different grades, different mathematics topics, and from teachers exhibiting more variation in their instruction might have led to somewhat different conclusions. Our aim in this study, however, was not to provide a comprehensive report of the strengths and limitations of MQI. Rather, in line with the practice-based approach followed in developing MQI, this study illustrates how the lenses adopted for analyzing instruction can be sharpened through a dynamic and iterative process that involves putting these lenses into actual practice while concurrently adopting a critical stance toward them.

References

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education, 59*(5), 389–407. https://doi.org/10.1177/0022487108324554.

Berlin, R., & Cohen, J. (2018). Understanding instructional quality through a relational lens. *ZDM Mathematics Education, 50*(3). (**this issue**).

Blazar, D. (2015). Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review, 48*, 16–29. https://doi.org/10.1016/j.econedurev.2015.05.005.

Blazar, D., Braslow, D., Charalambous, C. Y., & Hill, H. C. (2017). Attending to general and mathematics-specific dimensions of teaching: Exploring factors across two observation instruments. *Educational Assessment, 22*(2), 71–94. https://doi.org/10.1080/10627197.2017.1309274.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis, 39*(1), 146–170. https://doi.org/10.3102/0162373716670260.

Blazar, D., Litke, E., & Barmore, J. (2016). What does it mean to be ranked a "high" or "low" value-added teacher? Observing differences in instructional quality across districts. *American Educational Research Journal, 53*(2), 324–359. https://doi.org/10.3102/0002831216630407.

Borko, H., Eisenhart, M., Brown, C., Underhill, R., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education, 23*(3), 194–222. https://doi.org/10.2307/749118.

Boston, M. D. (2012). Assessing the quality of mathematics instruction. *Elementary School Journal, 113*(1), 76–104. https://doi.org/10.1086/666387.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd edn., pp. 328–375). New York, NY: Macmillan.

Charalambous, C. Y., & Hill, H. C. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Unpacking a complex relationship. *Journal of Curriculum Studies, 44*(4), 443–466. https://doi.org/10.1080/00220272.2011.650215.

Charalambous, C. Y., & Praetorius, A.-K. (2018). Studying instructional quality in mathematics through different lenses: In search of common ground. *ZDM Mathematics Education, 50*(3). (**this issue**).

Cohen, D., Raudenbush, S., & Ball, D. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 1–24. https://doi.org/10.3102/01623737025002119.

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

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.

Delaney, S. (2012). A validation study of the use of mathematical knowledge for teaching measures in Ireland. *ZDM, 44*(3), 427–441. https://doi.org/10.1007/s11858-012-0382-5.

Hawkins, D. (2002). *The informed vision: Essays on learning and human nature*. New York: Algora Publishing. (**originally published 1967**).

Hill, C. K., Umland, K., Litke, E., & Kapitula, L. R. (2012c). Teacher quality and quality teaching: Examining the relationship of a teacher assessment to practice. *American Journal of Education, 118*(4), 489–519. https://doi.org/10.1086/666380.

Hill, H. C. (2010). *The Mathematical Quality of Instruction: Learning Mathematics for Teaching*. Paper presented at the 2010 annual meeting of the American Educational Research Association, Denver, CO.

Hill, H. C., Blazar, D., & Lynch, K. (2015). Resources for teaching. *AERA Open, 1*(4), 1–23. https://doi.org/10.1177/2332858415617703.

Hill, H. C., Blunk, M., Charalambous, C. Y., Lewis, J., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511. https://doi.org/10.1080/07370000802177235.

Hill, H. C., Charalambous, C. Y., Blazar, D., McGinn, D., Kraft, M., Beisiegel, M., & Lynch, K. (2012a). Validating arguments for observational instruments: Attending to multiple sources of variation. *Educational Assessment, 17*(2–3), 1–19. https://doi.org/10.1080/10627197.2012.715019.

Hill, H. C., Charalambous, C. Y., & Kraft, M. (2012b). When rater reliability is not enough: Teacher observation systems and a case for the G-study. *Educational Researcher, 41*(2), 56–64. https://doi.org/10.3102/0013189X12437203.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*(3), 794–831. https://doi.org/10.3102/0002831210387916.

Hill, H. C., Kraft, M., & Herlihy, C. (2016). Developing common core classrooms through rubric-based coaching: Early findings report. http://cepr.harvard.edu/files/cepr/files/mqi-coaching-research-findings.pdf. Accessed 10 Feb 2017.

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

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle: Bill & Melinda Gates Foundation. http://www.metproject.org/reports.php. Accessed 30 May 2013.

Kraft, M. A., & Hill, H. C. (2017). *Developing ambitious mathematics instruction through web-based coaching: An experimental trial*. Harvard University Working Paper.

Kyriakides, L., Creemers, B. P. M., & Panayiotou, A. (2018). Using educational effectiveness research to promote quality of teaching: The contribution of the dynamic model. *ZDM Mathematics Education, 50*(3). (**this issue**).

Learning Mathematics for Teaching Project. (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education, 14*(1), 25–47. https://doi.org/10.1007/s10857-010-9140-1

Litke, E. (2015). *The state of the gate: A description of instructional practice in algebra in five urban districts*. Unpublished Doctoral dissertation, Harvard Graduate School of Education.

Lynch, K., Chin, M., & Blazar, D. (2017). Relationships between observations of elementary mathematics instruction and student achievement: Exploring variability across districts. *American Journal of Education, 123*(4), 615–646. https://doi.org/10.1086/692662.

Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, NJ: Erlbaum Associates.

Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophical and practical guide*. London: The Falmer Press.

Mitchell, R. N., & Marin, K. A. (2015). Examining the use of a structured analysis framework to support prospective teacher noticing. *Journal of Mathematics Teacher Education, 18*(6), 551–575. https://doi.org/10.1007/s10857-014-9294-3.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timplerley, H., & Earl, L. (2014). State of the art—teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256. https://doi.org/10.1080/09243453.2014.885451.

Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: Some preliminary findings from the evaluation of the mathematics enhancement program (primary). *School Effectiveness and School Improvement, 11*(3), 273–303. https://doi.org/10.1076/0924-3453(200009)11:3;1-G;FT273

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

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

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd edn.). Thousand Oaks, CA: Sage Publications.

Shaughnessy, M., Ball, D. L., Mann, L., & Garcia, N. (2015). (How) can explicitness about mathematical practices support equitable instruction? Paper presented at the NCSM Annual Conference. Boston, MA. https://static1.squarespace.com/static/577fc4e2440243084a67dc49/t/578d1e8759cc687748192f60/1468866184299/041315_NCSM.pdf. Accessed 10 Sept 2016.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X015002004.

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education, 36*(5), 404–411.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Thompson, P. W., & Thompson, A. G. (1994). Talking about rates conceptually, Part I: A teacher's struggle. *Journal for Research in Mathematics Education, 25*(3), 279–303. https://doi.org/10.2307/749339.

Walkowiak, T. A., Berry, R. Q., Meyer, J. P., Rimm-Kaufman, S. E., & Ottmar, E. R. (2014). Introducing an observational measure of standards-based mathematics teaching practices: Evidence of validity and score reliability. *Educational Studies in Mathematics, 85*(1), 109–128. https://doi.org/10.1007/s10649-013-9499-x.