ORIGINAL ARTICLE

CrossMark

Understanding Instructional Quality Through a Relational Lens

Rebekah Berlin[1] · Julie Cohen[1]

Abstract

In this paper, we analyze mathematics lessons using the Classroom Assessment Scoring System (CLASS), a standardized observation protocol that suggests that high-quality lessons are distinguished by the tenor and frequency of classroom interactions. Because the CLASS focuses on interactions, rather than the specifics of content teaching, it can be used across content areas from language arts to mathematics. While many previous studies have used CLASS as a measure of instructional quality, to date, no work has examined the affordances and constraints of using the content-agnostic CLASS to examine instructional quality in mathematics lessons. Our close qualitative analysis of three lessons highlights the importance of including practices that cut across content areas in measurement of instructional quality in mathematics classrooms. In addition, this paper is the first to highlight aspects of instruction in mathematics classrooms that are obscured by the CLASS. Discussion highlights how a relational lens foregrounds particular instructional aspects and marginalizes others.

Keywords Mathematics instruction · Classroom observation · And general versus content-specific pedagogy

1 Introduction

There is not consensus on the best way to conceptualize and measure high-quality mathematics teaching. This is evident in the recent proliferation of mathematics-specific classroom observation tools (English and Kirshner 2015). The expanding landscape of observation rubrics can make it difficult for researchers and practitioners to determine which tool best suits their purposes. One factor complicating this decision is that different protocols emphasize different dimensions of instructional quality in mathematics lessons (Kane and Staiger 2012).

Most observation protocols used in mathematics classrooms are made up of scales that measure context, content, and/or subject-specific content. Scales designed to measure behavior management, use of instructional time, and other features of the classroom environment which influence the

✉ Rebekah Berlin
rlb9t@virginia.edu

Julie Cohen
jjc7f@virginia.edu

[1] University of Virginia, Charlottesville, VA, USA

extent to which students can access content-related learning opportunities provide information about the extent to which there is a *context* that supports learning (Bell et al. 2012; Danielson Group 2017; Pianta and Hamre 2009). Scales designed to measure the teaching of academic *content* capture practices that pertain to the teaching of mathematics content but are not mathematics-specific. Content-focused scales include teacher feedback practices, questioning, and/or connections to prior academic material. Contextual and content-focused practices are undoubtedly important in mathematics classrooms, but they are also important in teaching language arts, science, and social studies.

In contrast to scales that assess practices we might expect to observe across content areas, *subject-specific* scales are designed to measure things we expect *only* to happen during mathematics instruction. These practices include the mathematical substance of teacher explanations and multiple representations of mathematical content (Charalambous and Litke 2018; Walkowiak et al. 2018). They pertain to teaching mathematics and not other subjects.

Observation protocols designed to measure instruction using only measures of content and context are content-generic, meaning they can be used in any classroom, regardless of the subject being taught (see Fig. 1). Other protocols have been designed exclusively for use during mathematics lessons. While mathematics-specific protocols can, in
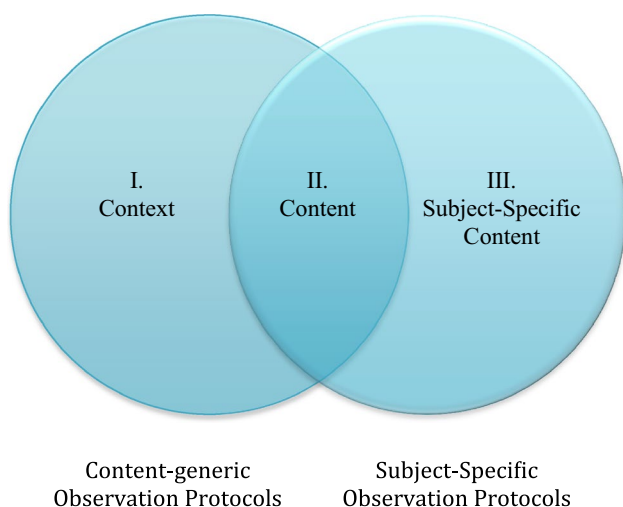
🍃 Springer

Fig. 1 Different lenses for measuring mathematics instruction

theory, include measures of context, content, and subject-specific content, most focus on practices related to the teaching of mathematics content. Some of these content-focused practices may be useful across subject areas, but many are exclusive to the teaching of mathematics.

Recent work highlights the difference between mathematics-specific and content-generic protocols. Multiple measurement studies offer evidence that content-generic and mathematics-specific observational protocols capture distinct facets of instruction in mathematics classrooms (Blazar et al. 2017; McClellan et al. 2013; Walkowiak et al. 2014) and may require different types of rater expertise (Hill et al. 2012). Theoretical work has explored the impact of using different lenses. Hill and Grossman (2013) warn that general observation rubrics miss key subject-specific aspects of instruction and argue that when districts use content generic tools, teachers are deprived of feedback on important subject-specific practices. For example, mathematics teachers might receive information on how much time students spent engaged in academic work rather than on the mathematical depth of the task in which students were engaged.

In this paper, we argue that if districts or researchers focus solely on subject-specific aspects of mathematics instruction, they too will miss vital indicators of quality that may contribute to student learning of mathematics content. While it is critical to capture the nuances of mathematics teaching and learning using subject specific tools, there are also important aspects of classrooms obscured by such tools. In particular, relational aspects of quality instruction in mathematics classrooms have been shown to support student engagement in and learning of mathematics (Hamre and Pianta 2005; Kane and Staiger 2012; Mashburn et al. 2008; Walkowiak et al. 2014). Students may be better able to learn mathematics content if teachers foster warm classroom

environments and effectively redirect off-task behavior. These kinds of practices are rarely featured in mathematics-specific observational measures.

To illustrate the importance of including scales designed to capture context and content in measures of instructional quality in mathematics classrooms we engaged in a close analysis of three upper elementary mathematics lessons. We analyzed each with a widely used subject-generic instrument, the Classroom Assessment Scoring System Upper Elementary (CLASS UE; Pianta et al. 2012). These data suggest there are compelling reasons to consider subject-generic practices in conceptualizations of high-quality mathematics instruction in mathematics classrooms.

Our approach in this paper is distinct from those in extant literature on the CLASS. The CLASS has been used to examine instructional quality in mathematics classrooms (Allen et al. 2011; Mashburn et al. 2008; Kane and Staiger 2012; Hamre and Pianta 2005; Walkowiak et al. 2014), but these studies highlight the explanatory power of the tool's domains and dimensions. Other authors have contrasted the CLASS with other frameworks to highlight areas of commonality and uniqueness (Blazar et al. 2017; McClellan et al. 2013; Walkowiak et al. 2014). To date, no studies have engaged in an in-depth analysis of what the CLASS alone reveals and obscures about instructional quality in mathematics classrooms. Ours is the first to treat the tool as the sole unit of inquiry.

2 Theoretical Underpinnings and Empirical Findings Related to CLASS UE

2.1 Tool Domains and Their Theoretical and Empirical Foundations

The CLASS UE is based on developmental theory, which suggests that the interactions children have with adults and peers drive learning and social development (Bronfenbrenner and Morris 1998). A relational lens suggests that a child's behavior in the classroom cannot be understood outside of the relationship between child-level and classroom-level processes (Slavin et al. 2003). From this perspective, proximal classroom processes, or the relationship between micro (within child) and macro (environmental) level processes, not isolated events, are the primary driver of academic and emotional development (Ford and Lerner 1992). This relational lens underpins all parts of the CLASS UE in that the measure focuses exclusively on the frequency, depth, and duration of teacher-child and child–child interactions.

The CLASS UE is divided into four domains: Emotional Support, Classroom Organization, Instructional Support, and Student Engagement (see Fig. 2). It is important to note that while early analyses provided empirical support for this

Fig. 2 The four domains and 12 dimensions of UE and Secondary. Readers interested in the indicators and behavioral markers nested under each dimension can contact Teachstone Training, LLC

Dimensions	Emotional Support	Classroom Organization	Instructional Support
	Positive Climate Teacher Sensitivity Regard for Adolescent Perspectives	Behavior Management Productivity Negative Climate	Instructional Learning Formats Content Understanding Analysis and Inquiry Quality of Feedback Instructional Dialogue
	Student Engagement		

conceptualization (Bell et al. 2012; Hafen et al. 2015; Hamre et al. 2013), findings from more complex analyses suggest alternate structures of the CLASS dimensions (Hamre et al. 2014; Kane and Staiger 2012; McCaffrey et al. 2015). Although determining the best-fitting latent structure of the tool remains an open empirical question, to be consistent with user-facing scoring, training, and rating documents, we have organized the theoretical underpinnings according to the four-domain structure. The theoretical underpinnings for each domain were drawn from an extensive literature review. While briefly outlined below, they are discussed in greater detail in the CLASS UE Manual (Pianta et al. 2012). Existing validity arguments for the CLASS also include support for these conceptual domains (see Bell et al. 2012 for an outline of the target domain, empirical evidence of the appropriateness of the scoring rules and the tool as an adequate representation of teaching quality). In each domain, there are a subset of more specific classroom-level dimensions. Each dimension is scored on a 1 (low) to 7 (high) scale. Scores of 1 and 2 are considered "low," scores of 3, 4, and 5 are at the "middle" level, and scores of 6 and 7 are considered "high." Raters score all 12 dimensions separately.

2.1.1 Emotional Support

The Emotional Support domain was drawn from research demonstrating that student success is fostered by feelings of relatedness to adults and classmates, opportunities for autonomy and choice in classroom activities, and interactions that promote a sense of competence (e.g., Allen et al. 1994, 2002; Ryan and Deci 2000). Literature documents the importance of teacher-student relationships for multiple student outcomes, including increased academic achievement, enhanced school motivation, and improved classroom behavior (Skinner et al. 1998). In particular, relationships that are characterized by a balance of challenge and support seem to promote positive student outcomes (Eccles 2004; Sandilos et al. 2017).

The broader Emotional Support domain is comprised of three specific classroom dimensions that are scored as individual practices: Positive Climate, Teacher Sensitivity, and Regard for Student Perspectives. Positive Climate measures "the enjoyment and emotional connection that teachers have with students, as well as the nature of peer interactions" (Pianta et al. 2012, p. 2). Teacher Sensitivity assesses "the level of teachers' responsiveness to the academic and social/emotional needs" of individual students (Ibid, p. 2). Regard for Student Perspectives foregrounds student choice in classroom decision-making. Within each dimension, raters are asked to score specific behavioral indicators that attend to fine-grained aspects of interactions. These include: Relationships, defined by specific behaviors such as physical proximity, peer interactions, shared positive affect, and social conversation; Positive Affect, defined by behavioral markers such as smiling, laughter, and enthusiasm; and Student Comfort, defined by behavioral indicators such as students take risks, participate freely, and seek support and guidance.

2.1.2 Classroom Organization

The Classroom Organization domain includes three dimensions: Behavior Management, Productivity, and Negative Climate. Compliant student behavior, efficient behavioral redirections, and minimal downtime and transitions characterize classrooms with strong Classroom Organization. These markers were drawn from theoretical work by developmental and ecological psychologists suggesting children develop divergent self-regulatory behaviors in different environments based on how adults manage time and behavior (Raver 2004; Kounin 1970). The authors also drew from constructivist theories on student engagement (Bowman and Stott 1994; Bruner 1996; Vygotsky 1978) as well as from empirical evidence that behavior and time management are associated with academic growth (Brophy and Evertson 1976; Good and Grouws 1977; Hoy and Weinstein 2006).

The Behavior Management dimension focuses on student behavior, the presence of specific proactive behavior management strategies, and the effectiveness and efficiency of behavioral redirections. The Productivity rubric assesses the degree to which learning time is utilized. Specifically this dimension focuses raters on classroom routines, teacher preparedness, and clarity of instructions. Negative Climate

evaluates the levels of anger, hostility, and/or disrespect in a classroom as evidenced by teacher or student behaviors such as yelling, punitive consequences, or sarcasm. While Negative Climate was originally hypothesized to load onto the Emotional Support domain (e.g., Hamre et al. 2007, 2014), more recent evidence drawn from samples with older students suggests it loads more strongly onto Classroom Organization (Hafen et al. 2015). The authors posit this may be because an increased Negative Climate could cause or be the result of classroom disruptions captured under the Behavior Management dimension.

2.1.3 Instructional Support

Based upon research that suggests the ways in which teachers represent content to children may affect student learning, the Instructional Support domain focuses on the instructional strategies teachers use to support children's cognitive and linguistic development (Taylor et al. 2003). The dimensions under this domain draw from literature on the positive association between varied instructional modalities and student engagement (Yair 2000), the positive relationship between immediate, specific, contingent feedback and student outcomes (e.g., Butler 1987; Brophy 1981; Marzano et al. 2001), and the importance of higher-order thinking skills and metacognition (e.g., Bransford et al. 2000; Davidson and Sternberg 2003; Marzano et al. 2001). In addition, research suggests that specific pedagogical strategies are instrumental in supporting student learning. These include: breaking new material into small steps (Bransford et al. 2000), connecting new knowledge to prior knowledge and real world examples (Lee 2007; Tharp and Gallimore 1988; Levin and Pressley 1981), numerous examples and opportunities to practice (Rosenshine 1995), providing students with a strong base of factual knowledge and skills that build toward "big ideas" in the larger academic discipline (Bransford et al. 2000), and highlighting similarities and differences between examples (Marzano et al. 2001).

The Instructional Support Domain includes five dimensions. Instructional Learning Formats measures how teachers facilitate learning activities to maximize student engagement. Content Understanding assesses how teachers engage students in the key ideas in an academic discipline. Analysis and Inquiry focuses on the degree to which teachers promote higher-order thinking skills such as hypothesis testing and the application of knowledge and skills in a wide array of contexts. Quality of Feedback assesses whether teacher feedback pushes students to extend their understanding of concepts and skills. Finally, Instructional Dialogue foregrounds the ways teachers engage students in rich, academic questioning and discussion. Indicators nested within the above domains include "Learning Targets/Organization", which focuses raters on behaviors such as "clear learning targets", "previews", "reorientation/summary statements" (p. 63); "Opportunity for Practice of Procedures and Skills" which directs rater attention to "supervised practice" and "independent practice" (p. 71); and "Scaffolding" where the behavioral markers are "Assistance," "Hints," and "Prompting completion and thought processes" (p. 89).

2.1.4 Student Engagement

The final domain in the CLASS UE is Student Engagement. It assesses how actively students participate in classroom activities by analyzing whether children ask questions, volunteer ideas, look at the teacher, and focus on the academic task at hand. This domain was added to the tool because of a National Research Council report (2003) that highlighted the positive association between student engagement and student outcomes.

2.2 Prior Empirical Use

2.2.1 Prior Use Across Subjects

Substantial work documents the substantial associations between the types of interactions highlighted by the CLASS and key child outcomes in preK-12 settings. Across grade levels, teachers' instructional interactions have consistently predicted student academic and language outcomes, and emotional interactions have predicted the development of students' social skills (e.g., Allen et al. 2013; Mashburn et al. 2008; Parkarinen et al. 2010). Specifically, prior work has found that struggling or "high-risk" students perform similarly to their "low-risk" peers when they are placed in classrooms with high emotional and instructional support, but significantly worse than their peers when they are placed in less supportive classrooms (Hamre and Pianta 2005). Classrooms with improved teacher–student interactions are associated with increases in student achievement across subjects (Allen et al. 2011).

There are consistent classroom trends in studies using the CLASS across a range of contexts and diverse populations of students (Downer et al. 2012). Synthesizing evidence from multiple studies, Pianta and Hamre (2009) note many preschool and elementary school classrooms have high levels of emotional support, but low levels of instructional support. They also find that many students spend a large amount of time without the opportunity to engage in any learning activity: 42% of the time in preschool to 30% of the time in fifth grade (Ibid 2009). While these trends characterize classrooms in the United States, there is ongoing research looking at the use of the CLASS in international contexts (e.g., Hu et al. 2016; Leyva et al. 2015; Pakarinen et al. 2010).

Given the developmental lens of the CLASS, there are different versions of the tool for different age groups. While

the Pre-K, K-3, Upper Elementary (UE), and Secondary tools share similar domains, the Infant and Toddler tools have different foci. The Infant tool is made up of a single domain, Responsive Caregiving, and the Toddler tool is made of two domains, Emotional and Behavioral Support and Engaged Support for Learning.

The version used in this paper, the CLASS UE, was used in the Measures of Effective Teaching Study (MET; Kane and Staiger 2012). According to the CLASS UE manual (Pianta et al. 2012), psychometric evidence from the MET show acceptable model fit for the three-factor model (RMSEA 0.11, CFI = 0.91; Acock 2013; Hair et al. 1998) and shows that each dimension loads strongly onto its associated domain (loadings range from 0.76 to 0.96). Domain-level Cronbach's alphas ranged from 0.87 to 0.92 indicating high internal consistency. Analysis of double coded videos demonstrates that raters were able to score an exact or adjacent score in 68–95% of the double coded videos, depending on the domain. Data from the MET study demonstrated a positive correlation ($r = 0.25$) between teachers' CLASS scores and value-added estimates of their effects on student achievement.[1]

2.2.2 CLASS as a Measure of Mathematics Instruction

Extant work on the CLASS focuses on the practices as outcomes for interventions (e.g., Allen et al. 2011), as a measure of instructional quality used across subjects (e.g., Mashburn et al. 2008; Kane and Staiger 2012), or as a complement to subject specific tools (Hamre and Pianta 2005; Walkowiak et al. 2014). While CLASS has been used as the sole instrument to measure instructional quality in mathematics classrooms (e.g., Bell et al. 2012), these studies focus more on measurement issues and the general quality of interactions in the *context* of mathematics classrooms rather than squarely on the *mathematical* quality of instruction.

Ours is the first study to engage in a detailed qualitative analysis of a small number of lessons to illustrate what is highlighted and what is obscured when a subject generic lens like the CLASS is applied to mathematics classrooms. To concretize and extend theoretical work detailing the limitations of using content-generic tools, we engage in a close examination of three upper elementary mathematics lessons. We ask:

1. *What do ratings from the CLASS UE make visible about instructional quality in mathematics lessons?*

2. *What do ratings from the CLASS UE obscure about instructional quality in mathematics lessons?*

3 Methods

In the present analysis, we viewed three–fourth-grade lessons from the National Center of Teaching Effectiveness video library. For more information on these lessons see Charalambous and Praetorius (2018). We watched one lesson each from Mr. Smith's, Ms. Young's, and Ms. Jones' classrooms using the CLASS UE rubrics.[2] CLASS UE requires raters to collect evidence on a range of behavioral indicators and weigh the overall composition of evidence when scoring a particular domain of a classroom. According to CLASS UE protocol, we collected evidence under the three–five behavioral indicators nested in each dimension and aggregated these into a dimension level score at the end of the lesson. See Fig. 3 for an example of a dimension face page, which provides an overview, but not the actual scoring guidance, for a dimension.

Each observer was trained and certified as a reliable CLASS rater. The CLASS UE manual specifies that video observations should be rated in 15–20 min cycles. Therefore, we divided each of the three videos into segments of equal length. For video one (total time 38 min), we rated two segments; for video two (total time 68 min), we rated four segments; for video three (total time 56 min), we rated three segments. While watching each video, each rater took notes into the CLASS UE Score Sheet, categorizing observations into the 12 dimensions under their associated behavioral indicators in real time. Following the end of each segment, raters paused the video and immediately rated the cycles. Segments were rated within a 10-min window on each of the 12 dimensions. Finally, after the last rating cycle for each video, we composited each score by averaging scores across cycles to arrive at a single score for each dimension for the observation period. Dimension scores were averaged to provide domain level scores after reverse coding Negative Climate. Finally, after each video, we created analytic memos detailing what was highlighted and obscured in using the CLASS to rate upper elementary mathematics instruction.

It is important to note that because of the number of cycles we observed, neither we, nor our readers, can make generalizations about individual teacher effectiveness. Due to the instability of ratings of single lessons, the manual explicitly states if the CLASS is being used to measure teacher quality, it must be through "multiple lessons, and ideally […] across multiple class sections" (Pianta et al. 2012, p. 8). Therefore, the results and discussion below

---

[1] Outside of the MET study, increases in CLASS scores have been shown to predict student achievement scores including a nine percentile term increase in student test scores (Allen et al. 2013) and a 0.16 standard deviation increase in student achievement (Allen et al. 2011).

[2] All teacher's names are pseudonyms.

Fig. 3 Content understanding dimension face page

are merely meant to ground our discussion of the tool in concrete examples and to provide readers a snapshot of the types of classroom evidence captured with the CLASS as compared to other observational measures.

4 Results

The three lessons varied in terms of the quality of instruction, as measured by the CLASS (see Table 1 for aggregated dimension and domain level scores). All three scored highest on the Classroom Organization domain, and two of the three lessons scored the lowest on the Instructional support domain. Ms. Young's instruction scored at the mid-level across the four domains. Ms. Jones' instruction was consistently at the mid and high level. Mr. Smith had the most varied portrait of instruction, with domain levels scores ranging from low (instructional support) to high (classroom organization).

4.1 Mr. Smith

Averaged across dimensions, across segments, and rounded to the nearest whole number, Mr. Smith's classroom received a score of 3 for Emotional Support, 6 for Classroom Organization, 2 for Instructional Support, and 3 for Student Engagement. The classroom's Emotional Support score of 3 places it in the lower end of the mid range. This score reflects that there was occasional, but inconsistent evidence of emotional support throughout the video. For

Table 1 Average dimension and domain scores

Video # Dimension	Mr. Smith		Ms. Young		Ms. Jones	
	Dimension	Domain	Dimension	Domain	Dimension	Domain
Emotional support Positive climate	3	3	3	3	6	5
Teacher sensitivity	3		4		6	
Regard for student perspectives	2		4		4	
Classroom organization Behavior management	7	6	4	5	7	7
Productivity	6		6		7	
Negative climate	1		2		1	
Instructional support Instructional learning formats	5	2	6	5	6	4
Content understanding	3		6		5	
Analysis and inquiry	1		4		2	
Quality of feedback	2		5		4	
Instructional dialogue	1		5		3	
Student engagement	3	3	4	4	6	6

Negative Climate was reverse coded before being averaged. All scores have been averaged across segments and are rounded to the nearest whole number for ease of interpretation. For some, such as Ms. Young on Emotional Support, the domain average of the averaged rounded dimension scores (3, 4, 4) is not equivalent to the domain average of the unrounded dimension level scores averaged across segments (2.75, 3.5, 3.75). The rounded numbers are presented only for ease of interpretation

example, despite a few instances of shared positive affect, such as a joke about acute angles, both Mr. Smith's and his students' affects were flat for the majority of the video. Mr. Smith occasionally connected material to common terms in students' life such as when he related acute angles to being "cute and tiny", and obtuse angles to being "obese." Though Mr. Smith sporadically appeared to scan the classroom, he spent the majority of the lesson pacing the front of the classroom and never noticed a student's raised hand or students whispering, "What are we supposed to do?" to one another. The lesson was tightly teacher controlled, and he did not provide students with authentic choices, opportunities for meaningful peer interactions, or opportunities for leadership and responsibility.

The classroom's aggregated score was 6 for Classroom Organization because little instructional time was lost due to student behavior. There were occasional instances where productivity of the classroom slowed because Mr. Smith was writing out a problem by hand or distributing materials inefficiently. There was only one instance of Negative Climate, when students laughed at another student at the board.

The classroom scored a 2 for instructional support. There was evidence of clear learning targets and multiple modalities for instruction, for example the lesson included both auditory, through the form of Mr. Smith's lecture, and kinetic, such as when students had the opportunity to circle the correct type of angle at the Smart Board, ways to engage with the lesson material. However, there was little evidence of depth, higher-order thinking, quality feedback, instructional dialogue, or opportunities for students to independently engage with the lesson material. Most tasks were

rote in nature. For example, students were asked to come to the Smart Board and use the protractor tool to open an angle to the number of degrees Mr. Smith provided or to come to the Smart Board and choose whether an angle was acute, right, or obtuse.

Finally, Student Engagement was rated as 3. There was a group of students off task for the majority of the video, whispering and laughing amongst themselves. Several students appeared compliant and on task, however, they did not seem actively engaged. Students yawned throughout the lesson and did not demonstrate active listening behaviors.

4.2 Ms. Young

Ms. Young's classroom scored a 3 for Emotional Support, 5 for Classroom Organization, 5 for Instructional Support, and 4 for Student Engagement. Though there was little evidence of teacher warmth or shared positive affect throughout the video, students demonstrated comfort with Ms. Young, approaching her to ask questions, show their work, and suggest alternate solution strategies. Ms. Young demonstrated mixed awareness of and responsiveness to students' academic needs. She circulated throughout the room and checked in with almost every student individually about their academic progress during small group work. She provided supportive feedback to some students but chastised others for not working and did not offer them instructional support. At times, she demonstrated Regard for Student Perspectives such as when she anchored abstract mathematics problems in scenarios students could relate to (equal groups became "apples in boxes"), and

allowed students to work in groups and choose their own materials to solve mathematics problems. At other times, she restricted student autonomy by telling students they were not allowed to get their own materials and not to argue with her about certain solution strategies.

Ms. Young's classroom scored a 5 for Classroom Organization because while there were clear and consistently enforced expectations for student behavior when students were on the carpet, instructional time was lost to student behavior during small group work and to a long transition from desks to carpet. There were also repeated instances of Negative Climate throughout the video. Ms. Young made comments CLASS classifies as sarcastic and derogatory such as, "Thank you for disrupting the lesson throughout the day" and "You don't have the worksheet. People are asked to do it in their journal, and they're doing it in their journal. And you're sitting down there sucking your finger." There were also several instances of mild irritability and a few of punitive control such as when Ms. Young threatened to send various students away from the group or out of the room. She eventually sent them into the hallway.

Of the three classrooms reviewed, Ms. Young's classroom scored highest for Instructional Support. Ms. Young outlined clear learning targets and the lesson was aligned to these goals. She actively facilitated student learning through a variety of modalities, strategies, and materials. Students were allowed to choose between proving the relationship between the factors in two multiplication problems through a variety of materials including graph paper, cubes, and diagrams. Lesson activities consistently focused students on independently discovering meaningful relationships between concepts and procedures, such as those between representations of multiplication and between factors. Ms. Young provided open-ended tasks and consistently pushed students to explain their cognitive processes and approaches by stating that knowing the answer to a problem was not enough, and that each student should be "justifying that your answer is true." Students received extensive practice time.

Scores on the Instructional Support domain indicated that, despite these strengths, there was substantial evidence of student confusion throughout small group time. Rather than providing encouragement, affirmation, or support for struggling students, Ms. Young often chastised students for their incorrect responses and pace. Though the tasks she presented were open-ended, often her dialogue with students limited engagement with the task so that students may have experienced tasks as close-ended. For example, there were multiple occasions where she explicitly told students the steps to complete in order to create the visual she wanted them to share on the carpet. Student Engagement was mixed throughout the video resulting in an aggregate mid-range score. Most students appeared actively engaged during the opening and closing that took place on the carpet, but many

students appeared distracted and disengaged during the group work at their desks.

4.3 Ms. Jones

Ms. Jones's classroom scored a 5 for Emotional Support, a 7 for Classroom Organization, a 4 for Instructional Support, and a 6 for Classroom Engagement. Under the Emotional Support domain there was consistent evidence of relationships, positive communication, respectful language, and student comfort throughout the lesson. Ms. Jones displayed sensitivity circulating around the room, anticipating and circumventing problems with sharing materials, group work, and lesson content. For example, when there were not enough scissors and rulers for students to use, Ms. Jones explained the system tables would use to share them to ensure every member of the group got equal access. There was little evidence, however, of authentic student autonomy or leadership, and no evidence of meaningful peer interactions, until the end of the lesson when students worked in groups cutting apart circles to represent multiplication as equal groups of fractional parts.

This classroom scored highest for Classroom Organization of the three lessons, because there was no evidence of negativity and the classroom was highly productive. Ms. Jones used behavior management strategies such as positive behavior narration, hand signals, and quick redirections. No time was lost to student misbehavior.

Ms. Jones's classroom scored a 4 for Instructional Support. The lesson had several strengths in this domain. In every segment, the lesson was aligned to the learning targets, and lesson material was presented through a variety of engaging materials. For example, students represented three ways to multiply fractions by a whole number on a three panel foldable. One of the methods involved using construction paper circles, cutting them into equal groups, and using repeated addition to find the total. Ms. Jones clearly presented lesson content, breaking down strategies for multiplying fractions into crisply delineated steps. She built on student background knowledge by connecting multiplying fractions to students' knowledge of repeated addition. She first had students represent 2×2 as $2 + 2$, 2×3 as $2 + 2 + 2$, before they represented $5 \times \frac{3}{4}$ as $\frac{3}{4} + \frac{3}{4} + \frac{3}{4} + \frac{3}{4} + \frac{3}{4}$. She also explicitly reviewed a strategy students had already learned to multiply a fraction by a whole number, before exposing them to new strategies. Additionally Ms. Jones anticipated student misunderstandings by asking questions like "can I just put R3?" so that students had to explain to her why she needed to write a remainder as a fractional part.

Despite these strengths, there was limited evidence of higher-order thinking or quality teacher-student and student–student dialogue for the first two-thirds of the lesson. Talk was heavily teacher-directed. Sometimes she engaged

in substantive feedback loops with students or provided scaffolds to those who struggled, such as when she prompted a student 1×4 is what, now 7×4 is what? At other times, however, her feedback was perfunctory; she often simply exclaimed, "Good!" and at other times she ignored incorrect responses. Most students appeared actively engaged throughout the lesson. Students were manipulating materials, asking and answering questions, and sharing ideas with the teacher. This placed the classroom in the upper range of Classroom Engagement.

5 Discussion

As is clear in the interactions described above, there are certain aspects of mathematics instruction that are foregrounded or marginalized when lessons are scored with the CLASS. Below, we argue that certain foci of the CLASS, which are missing from many mathematics-specific tools, offer essential information to those trying to understand instructional quality in mathematics. We also detail aspects of instruction in mathematics that are not captured by the CLASS.

5.1 Aspects of Instruction Highlighted by the CLASS

5.1.1 Facets of Mathematics Instruction

CLASS highlights aspects of high-quality teaching of academic content under the Instructional Support domain. For example, Content Understanding and Analysis and Inquiry focus raters in mathematics classrooms on the ways content is represented and students are able to engage with academic content. Importantly, while these aspects of instruction are relevant in mathematics classrooms, these practices are not *unique* to the teaching of mathematics.

Evidence from scales that measure the nature of instructional activities is illustrative of the way the CLASS is able to highlight meaningful differences in mathematics instruction, while only capturing practices that can be used across content areas. Within the CLASS framework, higher scoring instruction contains open-ended tasks allowing students to explore relationships between ideas. One of the reasons Mr. Smith received a low score on the Instructional Support domain is because his lesson relied on discrete questions with a single correct answer (e.g., "What type of angle is this?"). Ms. Jones, on the other hand, scored in the midrange because she posed a mix of open and close-ended tasks. Like Mr. Smith, she asked students several close-ended questions. However, in the third segment, she gave students several minutes to complete a task that allowed for student choice. She first asked students to generate equations where a fraction with a denominator of four was multiplied by a whole number. Because not all students chose the same equation

to model, there were multiple opportunities to discuss how to represent different products as both "improper fractions" and mixed numbers. Ms. Jones also capitalized on different student equations to explore how fractional pieces can be grouped to show whole numbers (eight fourths as two wholes).

A teacher can provide student choice and open-ended tasks in mathematics, language arts, science, or social studies classrooms; these practices are not limited to mathematics instruction. Nevertheless, a focus on general content practices reveals important features of mathematics instruction. While a mathematics-specific tool may have provided different insights about the mathematical quality of the instructional explanations Ms. Jones provided, the CLASS still captured important variation (e.g., a two point difference in Instructional Support) in the types of mathematical reasoning and representations students were exposed to across the two classrooms.

5.1.2 Interaction Between Content and Context

CLASS highlights the interaction between the content students are exposed to and the context in which that exposure occurs. In the CLASS framework, content is captured primarily through the Instructional Support domain. Different facets of context are measured through the Emotional Support, Classroom Organization, and Student Engagement domains. Many mathematics-focused tools do not have indicators to assess contextual factors such as student engagement or the emotional tenor of classroom interactions (Walkowiak et al. 2014) that influence the extent to which students can access these learning opportunities.

Ms. Young's classroom is particularly illustrative of the importance of capturing the relationship between content and context when assessing mathematics instruction. Of the three lessons analyzed, Ms. Young presented students with the greatest opportunity to engage with deep, rigorous mathematical tasks. There was evidence of high quality discourse about mathematical relationships, including those between 30×4 and 15×8, and broad organizing ideas such as why, when multiplying, doubling a factor doubles the product. These are reflected in a high score on the Content Understanding dimension. While, as documented in the results section, there was room for improvement in the consistency of the academic supports she provided students, analysis focused on content reveals a promising portrait of mathematics instruction.

Content without context, however, does not paint a full portrait of the interactions in her classroom. Students did not consistently take the opportunities Ms. Young provided. Several students used group work time to socialize, throw manipulatives at one another, or build patterned towers of cubes, ignoring Ms. Young's redirections. This was reflected

in lowered Behavior Management and Student Engagement scores because for the average student in the classroom, a large segment of the lesson was not spent on mathematics. Similarly, a chaotic transition from students' desks to the adjacent carpet resulted in lost instructional time and lowered the classroom's Productivity score during that segment. Put simply, the quality of the mathematical tasks Ms. Young presented may have mattered little because many students did not fully engage with them.

Along the same lines, there were multiple instances captured under the Negative Climate dimension where Ms. Young limited children's ability to engage with content. While she engaged in extended mathematical discourse with some students and asked them to share their work with the class, when other students provided incomplete or incorrect reasoning, she responded to them by saying, "No," "Don't argue," "You cannot be a part of the discussion," and "Go sit down." She sent some students out of the classroom or to the back of the classroom where she largely ignored them. In one of her only interactions with this group of students, she reminded one student that the reason he was struggling in this class was because he "refused to complete" his work on Monday. She did not offer to assist him and told him that he only had 5 min to complete it. The unequal distribution of materials, teacher time, instructional support, and warm interactions in this classroom, may have lead students to believe that mathematics is a discipline for a chosen few, not for all students in the classroom.

In classrooms like Ms. Young's, there are marked implications of excluding contextual practices that are shared across content areas from measurement of mathematics teaching. The absence of data on contextual factors may skew the conclusions researchers and practitioners draw from content-focused data. For example, were a school administrator to review only Ms. Young's scores under the Instructional Support domain, they might assume her development should focus on improving the way she responds to students' mathematical misunderstandings and errors. Using ratings from the full spectrum of CLASS dimensions however, this administrator might instead choose to focus on how to increase Ms. Young's ability to reduce the off task behavior in her classroom or how to build positive relationships with struggling students. Similarly, in research settings, classrooms like Ms. Young's may cloud the relationship between mathematics-specific teaching practices and student learning if researchers do not consider contextual factors in their measurement of mathematics teaching. While arguably a mathematics-specific tool would have picked up additional information on the content Ms. Young presented, this does not alter the fact that contextual factors in her classroom are likely impacting students' mathematical learning. Only a protocol that includes subject-generic practices such as those in the CLASS can provide this information.

5.2 Aspects of Instruction Obscured by the CLASS

5.2.1 Mathematics-Specific Content and Teaching Practices

As Hill and Grossman (2013) conjectured, the general lens of CLASS obscures nuances of mathematical instruction. More broadly, ratings from the CLASS do not indicate that mathematics was taught at all. Because of this, lesson segments can receive high scores in the Instructional Support domain, regardless of the presence or quality of the mathematics in the segment, if other general pedagogical practices are observed. For example, in Ms. Jones's video, the first 9 min did not contain any mathematics; students were constructing a foldable they were going to use throughout the lesson. She provided detailed explicit instruction about how to fold the construction paper, created a visual on the board to illustrate where she wanted students to write their name and what they should title it, and modeled the procedure with student materials. While all these constitute high quality general practices captured in the Instructional Learning Format dimension, they do not relate to mathematics. This example suggests that some scores on dimensions within the Instructional Support domain could be "inflated" by explicit instruction on myriad non-mathematical topics. This could potentially mislead users of the CLASS about the quality of mathematics instruction in a classroom.

Similarly, the Quality of Feedback and Instructional Dialogue dimensions capture general practices of classroom discourse, regardless of their mathematical substance. Thus, CLASS may classify comments of differential mathematical significance similarly. For example, one criterion of mid-range evidence of the "facilitation strategies" indicator in the Instructional Dialogue dimension is that "the teacher and/or fellow students sometimes acknowledge students' comments and repeat or extend these in ways that affirm their observations and/or recast the information in a more complex form" (Pianta et al. 2012, p. 99). Therefore, Mr. Smith's pattern of repeating student responses and adding an affirmative comment such as "Less than a right [angle]. Okay!" was counted as evidence of an equal weight to a more mathematically substantive comment from Ms. Young. When a student struggled to articulate the way he had transformed his array, Ms. Young stated, "[after cutting the original array in half] so you know you have two rectangles, and you move one of the rectangle down here to create a longer rectangle with one longer dimension and a short dimension. So now you have— this one has doubled and this side has been reduced." Ms. Young's comment used precise mathematical language to affirm a student and rephrase their resonse in academic language. Mr. Smith's "Okay!" while also affirming, did not add depth or mathematical richness to his student's understanding of angles. Ms. Young ultimately had a greater frequency of dialogue, which resulted in her having an overall higher

score, however, at the evidence level, these particular interactions were viewed identically through the lens of CLASS.

Relatedly, CLASS does not focus on precise mathematical language. Thus, statements like Ms. Jones's "four over four" instead of "four-fourths," or "I want you to have an equal sign and your final result" instead of "I want you to show your two fractions are equivalent" were not considered as evidence. It is likely that were this same lesson observed with a mathematics-*specific* lens such as the Mathematical Quality of Instruction (MQI) tool, these differences in mathematical discourse across the three lessons would be captured under the "Mathematical Language" and "Imprecision in Language and Notation" codes (see Charalambous and Litke 2018). In summary, precisely as Hill and Grossman (2013) predicted, there are some aspects of high quality mathematics instruction the CLASS will not provide users information about.

5.2.2 Teaching Mathematical Concepts and Procedures

Importantly, CLASS does not take a pedagogical stance on mathematics instruction. That is, neither procedural nor conceptual mathematics instruction is privileged. As such, the CLASS UE obscures distinctions between teaching focused on mathematical procedures and teaching focused on mathematical concepts.

Ms. Jones's classroom was characterized by exchanges focused on executing mathematical procedures, such as the one below:

> *Ms. Jones*: Very good. So I take 15 and I put inside. It becomes my dividend. And 4 becomes – what is that word that we use for the number that's outside the box? Raise your hand. What is that word that we use, Student R?
> *Student*: The divisor.
> *Ms. Jones*: Divisor. So 15 becomes my dividend, and 4 becomes my divisor, and I divide it out. Does 4 go into 1?
> *Multiple students*: No.
> *Ms. Jones*: No. So I put a zero. How many times does 4 go into 15?

Ms. Jones focuses only on the name and order of components of the process for long division. She does not explain why she is taking any of the above steps.

In contrast to the procedural exchanges highlighted in Ms. Jones's lesson, there were frequent interactions focused on mathematical concepts in Ms. Young's classroom. For example, she and a student explored why 16×6 is equivalent to $16 \times 3 + 16 \times 3$:

> *Ms. Young*: So Student C is saying that 48 plus 48 will give us 96, and that will be the same thing as 16 times

6 is 96. Yes, do you have another way of explaining it, Student C? I saw your hand up.
> *Student*: You can instead drawing *[inaudible]*, you can just draw six boxes.
> *Ms. Young*: We can draw 6 boxes showing the 3 and the 3. So if you combine all of the boxes together, 1, 2, 3, 4, 5 – so that's 16, 16, 16, 16, 16, 16.
> *Student*: And then you could just cut the middle off the one.
> *Ms. Young*: And they say like I cut the middle of this one [separates three of the boxes from the remaining 3], and that would give me my three group of 16 and three group of 16.

Though both of these exchanges focus on operations, they differ considerably in mathematical substance. Ms. Jones's focuses on the steps for dividing a two-digit number, and Ms. Young's focuses on connecting a semi-concrete representation of multiplication to an abstract numerical one. CLASS is ambivalent to this difference. These interactions both count as mid-range evidence for the "communication of concepts and procedures" indicator under the Content Understanding dimension because in both exchanges the "teacher demonstrates sufficient knowledge of the material to support student learning at a level that meets the goals of the lesson" (Pianta et al. 2012, p. 74).

Both of these interactions would also count as high-level evidence of "building on student responses" indicator under the Quality of Feedback dimension. Indeed both teachers expand "students' initial responses or action in ways that provide additional information or clarification" (Ibid p. 92). Based on similar patterns across the lessons, Ms. Young's conceptual and Ms. Jones's procedurally oriented lessons scored within one point of each other on the Instructional Support domain, though they diverged substantially in their approach to teaching mathematics.

A mathematics specific tool such as the Mathematics Scan (M-Scan), explicitly attends to these differences in language under "Depth" in its "Explanations and Justifications" dimension. As Hill and Grossman (2013) suggest, the differences between general and mathematics-specific tools have implications for providing teachers feedback. Coaches and administrators seeking to understand the volume of instructional time focused on mathematical procedures versus mathematical concepts could not gain this information from the CLASS UE.

6 Conclusion

These data suggest observation protocols that can be used across subjects, such as the CLASS, capture some, but not all, facets of instructional quality in mathematics classrooms.

For example, our analysis of Ms. Jones' classroom corroborated Hill and Grossman's (2013) conjecture that high ratings on subject-generic dimensions such as Positive Climate or Productivity do not necessarily also indicate quality mathematics instruction. Rather, they provide a context in which quality mathematical engagement is possible.

What is also clear from our analysis is that subject-generic and mathematics-specific teaching practices interact in meaningful ways. Ms. Young's lesson demonstrated that even when high-quality mathematical opportunities are available, they may be of limited impact if students do not engage with them. While multiple indicators of quality mathematics-specific instruction including mathematical discourse, meaningful mathematical choices, and student-generated mathematical justifications, were present in her classroom, student behavior reduced the extent to which these occurrences were likely to impact student learning. Because the CLASS attends to *both* the content and contextual practices, users obtain a holistic understanding of classroom practices that likely impact student experiences.

These data suggest a strong rationale for including subject-generic practices in conceptualizations of high-quality instruction in mathematics classrooms. Confining the measurement of mathematics instruction only to practices that are unique to mathematics may push out important features of classrooms in which mathematics instruction occurs. When contextual factors such as whether a classroom is a safe, productive, and engaging place are not considered, users of observation tools risk misinterpreting the relationship between mathematics-specific practices and student learning. Of course, working from a completely content-generic perspective means that while observers will assess content instruction in mathematics classrooms, it will be with broader brush strokes than a mathematics-specific tool. Therefore, there are limitations of the exclusive use of *both* subject-generic and mathematics-specific tools. These data suggest that conceptions of high-quality instruction in mathematics classrooms likely need to include both subject-specific and content-generic practices.

References

Acock, A. (2013). *Discovering structural equation modeling using stata*. College Station, Texas: Stata Press.

Allen, J. P., Hauser, S. T., Bell, K. L., & O'Connor, T. G. (1994). Longitudinal assessment of autonomy and relatedness in adolescent-family interactions as predictors of adolescent ego development and self-esteem. *Child Development, 65*(1), 179–194.

Allen, J., Marsh, P., McFarland, C., McElhaney, K., Land, D., & Jodi, K. (2002). Attachment and autonomy as predictors of the development of social skills and delinquency during midadolescence. *Journal of Consulting and Clinical Psychology, 70*(1), 56–66.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary

school instruction and student achievement. *Science, 333*(6045), 1034–1037.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system-secondary. *School Psychology Review, 42*(1), 76.

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17*(2–3), 62–87.

Blazar, D., 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.

Bowman, B., & Stott, F. (1994). Understanding development in a cultural context: The challenge for teachers. In B. L. Mallory & R. S. New (Eds.), *Diversity and developmentally appropriate practices: Challenges for early childhood education* (pp. 19–34). New York, NY: Teachers College Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn:*. In *Brain, mind, experience, and school*. Washington, DC: National Academies Press.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In *Handbook of child psychology: Theoretical models of human development* (pp. 993–1028). Hoboken, NJ: Wiley.

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51*(1), 5–32.

Brophy, J. E., & Evertson, C. M. (1976). Learning from teaching: A developmental perspective. Allyn and Bacon.

Bruner, J. S. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology, 79*(4), 474.

Charalambous, C. Y., & Litke, E. (2018). Studying instructional quality by using a content-specific lens: The case of the mathematical quality of instruction framework. *ZDM Mathematics Education*. https://doi.org/10.1007/s11858-018-0913-9.

Charalambous, C. Y., Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses: setting the ground for understanding instructional quality more comprehensively. *ZDM Mathematics Education*. https://doi.org/10.1007/s11858-018-0914-8.

Danielson Group. (2017). The framework. https://www.danielsongroup.org/framework. Accessed 1 May 2017.

Davidson, J. E., & Sternberg, R. J. (2003). *The psychology of problem solving*. Cambridge: Cambridge University Press.

Downer, J. T., López, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observations of teacher–child interactions in classrooms serving Latinos and dual language learners: Applicability of the Classroom Assessment Scoring System in diverse settings. *Early Childhood Research Quarterly, 27*(1), 21–32.

Eccles, J. S. (2004). Schools, academic motivation, and stage-environment fit. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 125–153). Hoboken, NJ: Wiley.

English, L. D., & Kirshner, D. (Eds.). (2015). *Handbook of international research in mathematics education*. New York: Routledge.

Ford, D. H., & Lerner, R. M. (1992). Developmental systems theory: An integrative approach. Thousand Oaks: Sage Publications, Inc.

Good, T. L., & Grouws, D. A. (1977). Teaching effects: A process-product study in fourth-grade mathematics classrooms. *Journal of Teacher Education, 28*(3), 49–54.

Hafen, C. A., Hamre, B. K., Allen, J. P., Bell, C. A., Gitomer, D. H., & Pianta, R. C. (2015). Teaching through interactions in secondary

school classrooms: Revisiting the factor structure and practical application of the Classroom Assessment Scoring System–Secondary. *The Journal of Early Adolescence, 35*(5–6), 651–680.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*(5), 949–967.

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). Building a science of classrooms: Application of the CLASS framework in over 4,000 US early childhood and elementary classrooms. Research paper, Foundation for Childhood Development. https://www.fcd-us.org/building-a-science-of-classrooms-application-of-the-class-framework-in-over-4000-u-s-early-childhood-and-elementary-classrooms/. Accessed 20 Apr 2017.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., … & Brackett, M. A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113*(4), 461–487.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child Development, 85*(3), 1257–1274.

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371–384.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher, 41*(2), 56–64.

Hoy, A. W., & Weinstein, C. S. (2006). Student and teacher perspectives on classroom management. In C. Everston & C. Weinstein (Eds.), *Handbook of classroom management: Research, practice and contemporary issues* (pp. 181–222). Mahwah, NJ: Lawrence Erlbaum Associates.

Hu, B. Y., Fan, X., Gu, C., & Yang, N. (2016). Applicability of the classroom assessment scoring system in Chinese preschools based on psychometric evidence. *Early Education and Development, 27*(5), 714–734.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Research paper, Bill & Melinda Gates Foundation. https://files.eric.ed.gov/fulltext/ED540960.pdf.

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Oxford: Holt, Rinehart & Winston.

Lee, C. D. (2007). *Culture, literacy, and learning*. New York: Teachers College Press.

Levin, J. R., & Pressley, M. (1981). Improving children's prose comprehension: Selected strategies that seem to succeed. In C. M. Santa & B. L. Hayes (Eds.), *Children's prose comprehension: Research and practice (pp. 44–71)*. Newark: International Reading Association.

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher–child interactions in Chile and their associations with prekindergarten outcomes. *Child Development, 86*(3), 781–799.

Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: ASCD.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732–749.

McCaffrey, D. F., Yuan, K., Savitsky, T. D., Lockwood, J. R., & Edelen, M. O. (2015). Uncovering multivariate structure in classroom observations in the presence of rater errors. *Educational Measurement: Issues and Practice, 34*(2), 34–46.

McClellan, C., Donoghue, J., & Park, Y. S. (2013). Commonality and uniqueness in teaching practice observation. Research paper, Clowder Consulting. http://www.clowderconsulting.com/wp-content/uploads/2016/01/Commonality-and-Uniqueness-in-Teaching-Practice-Observation_paper.pdf. Accessed 25 Apr 2017.

National Research Council. (2003). Engaging schools: Fostering high school students' motivation to learn. Washington: National Academies Press.

Pakarinen, E., Lerkkanen, M. K., Poikkeus, A. M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J. E. (2010). A validation of the classroom assessment scoring system in finnish kindergartens. *Early Education and Development, 21*(1), 95–124.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.

Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom Assessment Scoring System Upper Elementary Manual*. Charlottesville, VA: Teachstone.

Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75*(2), 346–353.

Rosenshine, B. (1995). Advances in research on instruction. *The Journal of Educational Research, 88*(5), 262–268.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.

Sandilos, L., Rimm-Kaufman, S., & Cohen, J. (2017). Warmth and demand: The relation between students' perceptions of the classroom environment and achievement growth. Child Development, 88(4), 1321–1337.

Skinner, E. A., Zimmer-Gembeck, M. J., Connell, J. P., Eccles, J. S., & Wellborn, J. G. (1998). Individual differences and the development of perceived control. Monographs of the Society for Research in Child Development, i-231.

Slavin, R. E., Hurley, E. A., & Chamberlain, A. (2003). Cooperative learning and achievement: Theory and research. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology* (pp. 177–198). Hoboken: Wiley.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal, 104*(1), 3–28.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge: Cambridge University Press.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children, 23*(3), 34–41.

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.

Walkowiak, T. A., Berry, R. Q., Pinter, H. H., & Jacobson, E. D. (2018). Utilizing the M-Scan to measure standards based mathematics teaching practices: affordances and limitations. *ZDM Mathematics Education*. https://doi.org/10.1007/s11858-018-0931-7.

Yair, G. (2000). Educational battlefields in America: The tug-of-war over students' engagement with instruction. *Sociology of Education, 73*(4), 247–269.