The Quality of Mathematics Instruction in Kindergarten:

Associations with Students' Achievement and Motivation

Panayota Mantzicopoulos

Purdue University

Brian F. French

Washington State University

Helen Patrick

Purdue University

Washington State University

Mantziopoulos, P. Y., French, B. F., & Patrick, H. (2019). The quality of mathematics instruction in kindergarten: Associations with students' achievement and motivation. *Elementary School Journal, 119,* 651-676.

Acknowledgement

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A140664 to Helen Patrick, Panayota Mantzicopoulos, and Brian F. French. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Abstract

We examined associations between the quality of kindergarten teachers' mathematics instruction and their students' achievement and motivation in mathematics. Using a sample of 20 kindergarten teachers and their 285 students, we rated five video-recorded mathematics lessons per teacher throughout the spring semester with the Mathematical Quality of Instruction (MQI). We collected information about students' achievement from teacher-ratings of student performance relative to state standards and an individually-administered standardized measure of mathematical reasoning. We obtained data about students' mathematics motivation (self-competence beliefs, interest, effort expenditure, and need for support) through individual interviews with the children and via teacher ratings. Multi-level modeling analyses indicated that scores on the MQI's Ambitious Mathematics Instruction scale and the Whole Lesson Scale predicted students' end-of-year progress on kindergarten mathematics standards but not their standardized test scores. The Whole Lesson scores were associated with teacher-rated students' motivation for mathematics (interest and need for support).

The Quality of Mathematics Instruction in Kindergarten:

Associations with Students' Achievement and Motivation

There is a pressing need to document the effectiveness of teachers' mathematics instruction in the early grades because early competencies are fundamental to the development of proficiency in both mathematics and related subjects. Students who have not mastered basic mathematics skills during the early years of school "can expect problems throughout their schooling and later" (National Research Council, 2001, p. 18). Early mathematics skills: (a) develop cumulatively; (b) develop at a faster rate in kindergarten and first grade than in later grades (Shanley, 2016); and (c) are consistent predictors of mathematics achievement, both in the short term (e.g., from kindergarten to first grade; Aunio & Niemivirta, 2010), as well as in subsequent grades (Bodovski & Farkas, 2007; Watts et al., 2015). Motivation to learn mathematics is equally as important for student success as mathematics skills (Patrick, Mantzicopoulos, & Sears, 2012). In the early years of school young children build their understandings of the meaning and value of mathematics and develop beliefs about their skills as mathematics learners (National Council of Teachers of Mathematics, 2000). Both mathematics learning and motivation are influenced by teachers' practices (e.g., Lerkkanen et al., 2012).

Even though teachers' classroom practices affect student outcomes, both in general (Nye, Konstantopoulos & Hedges, 2004) and for mathematics specifically (National Research Council, 2001), few studies have focused on mathematics instruction at the start of school. Notable among them is research from the Early Childhood Longitudinal Study–K (ECLS-K), which began in 1998 when students were in kindergarten. In these studies, teacher reports of the content covered and the time spent on mathematics predicted student achievement (e.g., Bodovski & Farkas, 2007; Bottia, Moller, Mickelson, & Stearns, 2014; Guarino, Dieterle, Bargagliotti, & Mason, 2013). However, less is known about how kindergarten teachers' observed mathematics practices are related to students' mathematics outcomes. Therefore, this is the focus of our study. We use a mathematics-specific observation measure of instructional quality – the Mathematical Quality of Instruction (MQI; Learning Mathematics for Teaching Project [LMTP], 2011) – to evaluate 100 mathematics lessons taught by 20 public school kindergarten teachers. We then examine how well teachers' instructional practices (reflected by MQI scores) predict kindergarteners' end-of-year mathematics achievement and motivation.

Documenting Mathematics Instruction with Math-Specific Observation Measures

In the last decade, there has been increased attention to the use of classroom observations as a meaningful way to document instruction (Cohen & Goldhaber, 2016). Fueled by national mandates intended to promote teacher effectiveness, observations now play a central role in evaluating instruction. Scores are thought to provide objective, clear, transparent, and specific information about teachers' strategies (Goldring, et al., 2015). Thus, observational measures of instruction, which have been used for research purposes since at least the 1970s (e.g., Brophy, & Good, 1970), now have important implications for policy and practice. Expectations that observational measures capture instruction that is associated with key student outcomes, however, underscore the need for research to provide substantiating evidence. We address this need by focusing on early mathematics instruction.

Most observation measures used currently to document teachers' instruction are generic, or content-general, guided by the view that good instruction is independent of subject matter (Danielson, 2013). With this approach, mathematics instruction is evaluated without particular attention to the mathematics being taught. In contrast, mathematics-specific measures are built on the premise that effective mathematics teaching is based on discipline-specific and pedagogical content knowledge and skills (e.g., Ball, Thames, & Phelps, 2008). These skills are reflected in the teacher's "specialized fluency with mathematical language, with what counts as a mathematical explanation, and with how to use symbols" (Ball, Hill, & Bass, 2005, p. 21). Considering our interest in mathematics, we opted for documenting instruction with a mathematics-specific, rather than with a content-general, assessment.

Of the observation measures that target mathematics instruction (Charalambous & Praetorius, 2018), a relatively small number are appropriate for the early elementary grades. Most prominent are the Reformed Teacher Observation Protocol (RTOP; Piburn et al., 2000), the Inside the Classroom Observation and Analytic Protocol (Horizon Research, 2003), the U-Teach Observation Protocol (UTOP; The UTeach Institute, 2014), and the MQI (Hill, Charalambous, & Kraft, 2012). These measures purport to be appropriate for documenting teachers' mathematics practices across grade levels, from early elementary through at least the middle school grades. However, we could not find studies that have used these measures to examine associations between early mathematics instruction and student outcomes.

We decided to use the MQI, a measure intended for documenting instruction in grades K-8 (Hill, 2011; Kane & Staiger, 2012), for several reasons. First, the MQI is built on the perspective that instruction involves dynamic interactions among teachers, students, and the mathematical content of a lesson. Within this context, the MQI centers on the *quality* of mathematics instruction rather than on teaching as a set of generic instructional strategies, such as classroom organization, routines, and general communication (LMTP, 2011).

Second, the MQI has an explicit and direct focus on "the nature of the mathematical content available to students during instruction" (LMTP, 2011, p. 30). This focus is aligned with the National Research Council's (2001) view that "the quality of instruction is a function of teachers' knowledge and use of mathematical content, teachers' attention to and handling of students, and students' engagement in and use of mathematical tasks" (p. 315). The MQI affords explicit attention to the mathematical quality of a lesson, regardless of the teacher's instructional orientation (e.g., didactic, child centered, inquiry-driven). This is in contrast to other math-appropriate measures (e.g., ROTP; Piburn et al., 2000; Inside the Classroom Observation and Analytic Protocol; Horizon Research, 2003) that stem from an interest in documenting reform-oriented teaching. Such measures, therefore, privilege some strategies (e.g., collaborative group work, hands-on learning) over others.

Third, researchers have produced validity evidence with MQI scores derived from upper elementary samples (LMTP, 2011). Factorial analyses support the measure's two-factor structure (Blazar, Braslow, Charalambous, & Hill, 2017) while additional validation evidence includes the association between MQI scores and teachers' mathematics content knowledge (e.g., Hill, et al., 2008) and student outcomes (e.g., Blazar & Kraft, 2017).

Fourth, the MQI was included in the recently concluded Measures of Effective Teaching (MET) Project (Bill & Melinda Gates Foundation, 2013) in grades 4-8. Like other popular observational measures that are currently being used across the nation to document teaching (Center on Great Teachers and Leaders, 2013), the MQI has the potential to inform future research efforts as well as promote effective teaching practices.

Despite the MQI's many strengths, there are no studies, to our knowledge, which have used it, or another mathematics-specific observation measure, to (a) document the quality of mathematics instruction in kindergarten, and (b) examine its links with students' mathematics outcomes. We respond to the critical need for such research with evidence from a sample of public-school kindergarten classrooms.

The Quality of Mathematics Instruction and Students' Mathematics Outcomes

Mathematics Achievement. Students' achievement is unequivocally an important outcome of instruction. Of note, the federal definition regards student achievement as tantamount to effective teaching: an effective teacher is one "whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth" (U.S. Department of Education [USDOE], 2009, p. 12). Consistent with this perspective, teachers' mathematics practices, as reflected in MQI scores, are associated significantly with upper elementary (Blazar, 2015; Blazar & Kraft, 2017) and middle school students' mathematics performance (Kane & Staiger, 2012), although the associations are modest (approximately 1/5 of a standard deviation) (Blazar & Kraft, 2017). Moreover, the students of teachers whose instruction lacks mathematical precision and clarity or includes errors tend to have low mathematics achievement (Blazar, 2015). There is a dearth of evidence, however, on the relationship between teachers' mathematics instruction (including content errors) in the early grades and young students' mathematics achievement.

We add to this literature by using two complementary types of mathematics achievement measures: (a) scores from a widely-used standardized mathematics achievement measure; and (b) teacher ratings of students' mathematics skills relative to state kindergarten standards. The standardized mathematics achievement measure is consistent with states' use of standardized test scores to gauge the quality of teachers' practices (USDOE, 2009), albeit not yet in kindergarten. Standardized achievement tests are distal to the classroom context, are intended to assess broad levels of mathematics knowledge and skills, and include items that are not linked to particular curricula. Therefore, scores may "afford a preclusion of bias towards a particular curriculum" (Hickey, Zuiker, Taasoobshirazi, Schafer, & Michael, 2006, p. 183). As such, they are considered appropriate for use across different districts or states.

Unlike standardized test scores, teachers' ratings of their students' progress on grade-level mathematics standards is a measure that is proximal to the classrooms' mathematics contexts; it documents learning that is aligned with the state's expectations for students' mathematics competencies by the end of kindergarten. This measure reflects the content that kindergarten teachers are accountable for in their mathematics teaching and are therefore likely to focus on when they evaluate and support the development of their students' mathematics skills. Scores from this assessment will provide an additional lens through which to gauge the relationship between teachers' mathematics strategies and students' mathematics knowledge.

Motivation for Mathematics. In addition to achievement, students' motivation for learning is a crucial outcome for consideration. Motivation helps direct students' attention and facilitates their engagement during instruction, which are associated with concurrent and future learning and achievement (Miele & Wigfield, 2014). Students' motivation for learning mathematics is relatively stable from elementary school through the end of high school (Gottfried, Fleming, & Gottfried, 2001s), and thus it is vital to consider its development in the early school years. Moreover, low motivation is an important precursor of disengagement and poor achievement, even when students have the competencies for success (Skinner, 2016).

Just as achievement is influenced by teachers' practices, so is student motivation (Patrick, et al., 2012; Wigfield et al., 2015). Thus, it is important to identify instructional strategies that support and enhance motivation and to consider these strategies alongside practices that promote achievement. Part of addressing this issue involves examining the extent to which scores reflecting teachers' instructional practices are related to students' motivation.

Research with young children in subjects like reading (Nolen, 2001; Wigfield, Guthrie, Tonks, & Perencevich, 2004) and science (Patrick & Mantzicopoulos, 2015) suggests that subject-specific motivation emerges as children interact with subject-specific content. However, other than evidence that teachers' broad practices (i.e., child-centered vs. didactic) are related to children's liking of mathematics (Lerkkanen et al., 2012), there is a dearth of research on the quality of mathematics-specific instruction and the motivation of students in the early grades.

Data with older students, albeit limited, indicate that the quality of teachers' observed mathematics practices is related to students' motivation. Comparisons between teachers with MQI scores in the top and bottom quartiles indicate small to moderate differences in their students' school liking and effort expenditure (Kane & Staiger, 2012). More recently, Blazar and Kraft (2017) examined math-specific motivation in relation to 4th- and 5th-grade teachers' MQI ratings. Teachers' scores on the MQI's Errors and Imprecision scale had small associations with students' reports of liking mathematics and their self-efficacy for mathematics. Specifically, a one-SD change in teachers' errors was associated with a 0.18 SD decrease in math liking and a 0.09 SD decrease in math self-efficacy. Contrary to expectation, students' motivation, was not predicted by the MQI's main scale (i.e., Ambitious Mathematics Instruction), which comprises strategies that support students' mathematical reasoning through the use of multiple solutions, mathematical language, and mathematics-related questioning and explanations.

We expect, based on the research outlined here, that our focus on the quality of mathematics instruction will highlight connections between teachers' practices and young children's mathematics motivation (e.g., whether they find mathematics interesting, how good they think they are at mathematics, and whether they want to learn it). We use information from both teachers and students to provide evidence on students' key mathematics-related motivational beliefs and behavioral indicators of motivation, premised on social-cognitive theories of motivation (Wigfield et al., 2015). Specifically, we examine students' reports of their competence in mathematics and their enjoyment or liking of math, both of which have been identified in samples of young children across academic subjects, including mathematics (Fredricks & Eccles, 2002). Additionally, we include teacher ratings of behavioral indicators of students' enjoyment of mathematics, as well as effort, persistence, and need for teacher support during mathematics activities (Schunk, Meece, & Pintrich, 2014; Wentzel & Brophy, 2014).

Summary of Research Aims

We use a multi-method, multi-informant approach to document associations between the quality of mathematics instruction in kindergarten and students' math achievement and motivation at the end of the school year. Specifically, we aggregate each teacher's MQI scores across five mathematics lessons from the spring of kindergarten. We examine hypotheses about the contribution of mathematics instruction on the following outcomes: (a) students' mathematics reasoning, reflected by standardized test scores; (b) student-reported motivation for mathematics; (c) teacher-reported student progress on kindergarten mathematics standards; (d) teacher-rated student interest; and (e) students' need for support in mathematics (teacher-reported). We test each hypothesis using a multi-level-modeling framework in order to address the nesting of students in classrooms. In our analyses we control for the effects of demographic characteristics (i.e., students' sex and socioeconomic status) as well as students' Fall mathematics knowledge and Fall need for support for learning mathematics.

Method

Participants

Teachers and Schools. The participants were 20 kindergarten teachers (19 female, 1 male; 18 White, 2 Hispanic) from six public schools within one midwestern state in the United States. Teachers' experience ranged from 1 to 33 years (*M* = 16 years). Kindergarten teacher participation rates in the schools ranged from 88% to 100%. Of the 22 participating teachers, 2 recorded fewer lessons than required; they were not included in this study.

We used data from the state's Department of Education to select schools that differed on a range of characteristics. Specifically, the six schools varied in terms of locale (rural, small town, urban fringe of large city), state-issued report card grade (A - C), students' ethnic composition (3% - 46% Hispanic; 0.2% - 46% Black), and students' family socioeconomic status (28% - 74% free & reduced-cost lunch).

Students. There were 285 kindergarteners in this study. We received informed consent for 324 students to participate (i.e., 79.4% of students), however there were not complete fall-to-spring data for 39 students. Of those, 30 moved out of the classroom, 5 were absent during testing, and 4 could not be tested due to special needs.

There were 155 males (54%) and 130 females. According to school records, 63.2% of the children were White; 22.8% were Hispanic, 9.1% were Black, and 5.0% were Multiracial or Other. Approximately half (152, 53.3%) of students received free or reduced-cost lunch. We used this information as an indicator of socioeconomic (SES) status for each child (0 = free-or reduced lunch status; 1 = paid lunch status).

Lessons

Teachers used iPads to video-record their mathematics lessons approximately once a week, for 10 weeks during the spring semester. Recorded lessons were spread evenly across the semester; they represented teachers' typical mathematics instruction, rather than being standardized across teachers or scripted by the research team. Of 211 mathematics lessons, we randomly selected 5 per teacher. On average, the lessons lasted 24 minutes and covered different topics; all lessons targeted concepts and skills listed in the state's standards for kindergarten mathematics. Specifically, 40 lessons addressed number sense standards (counting, writing numbers comparing values of two numbers, working with the number line to count), and 31 lessons involved computation and algebraic skills (e.g., addition and subtraction problems, composing and decomposing numbers). The remaining 29 lessons targeted data analysis skills (e.g., graphing and sorting activities), measurement, and geometry (2- and 3-dimensional shapes). We applied the MQI, which we describe next, to these 100 lessons.

Teacher Measures and Procedure

Mathematical Quality of Instruction (Hill, 2014). We used the most recent (2014) version of the MQI, which contains two sets of scales, each rated in a separate phase with its own procedure and format. We outline these two phases next.

Lesson segment scales. Raters first divide the lesson into segments of 7½ minutes. After viewing each segment, they stop to rate "whether the focus is on *mathematical* c*ontent* during half or more of the segment," using a dichotomously scored item (1 = yes; 0 = no) (Hill, 2014, p. 3). Items on the 2014 version are comparable to the earlier version (Hill, 2011). However, the 2014 segment-level items, used in this study, are scored on a 4-point scale (0 = not present, 1= low, 2 = mid, 3 = high). Ratings of 1 ("low") reflect a rote or procedural approach, "pro forma" responses, or brief instances of the strategy reflected in a particular item. Ratings of 2 ("mid") reflect instances of variable or mixed mathematical quality, where both high and low quality strategies are present. Finally, ratings of 3 ("high") are given when the strategies and/or activities described by the item are consistently substantive, detailed, and/or mathematically meaningful.

The MQI's items are grouped into four domains (Hill, 2014). *Richness of Mathematics* items (n = 6) assess the extent to which instruction focuses on mathematical language, multiple solutions, developing generalizations from particular cases, and making mathematics facts and procedures meaningful. *Working with Students and Mathematics* items (n = 2) document teachers' responses to students' contributions as well as remediation of students' difficulties. *Errors and Imprecision* items (n = 3) reflect teachers' content errors, problems in the use of mathematics language, and lack of clarity. *Common Core Aligned Student Practices* items (n = 5) assess students' participation and contributions to the mathematics tasks by, for example, explaining, questioning, or reasoning about mathematics. Item scores are averaged across each 7½ minute segment; the item means within each of the four domains are then averaged to create *domain scores*.

Each domain also includes one holistically scored item that serves as a single indicator of the lesson's quality in that domain. We did not use these single items, given that single indicators of performance in a particular domain are likely to be less reliable than the group of items comprising the domain (Nunnally & Bernstein, 1994). Moreover, each holistic item was analogous to its corresponding multi-item scale. Correlations between each holistic item and its associated scale were 0.98 (Richness of Mathematics), 0.91 (Working with Students), 0.97 (Errors and Imprecision), and 0.95 (Common Core Aligned Practices).

Factor analysis of the items (excluding the holistically-rated items) from 4th and 5th grades supports two factors (Blazar et al., 2017). One factor comprises items from the *Errors and Imprecision* scale, whereas a second factor (*Ambitious Mathematics Instruction*) comprises items from the remaining 3 scales (Richness of Mathematics, Working with Students, and Common Core Aligned Practices). We used this solution with our data and created a 3-item Errors and Imprecision scale (α = 0.55) and a 13-item Ambitious Mathematics Instruction scale (α = 0.88).

Whole Lesson scale. At the end of each lesson (i.e., after raters have viewed and scored all 7½ minute segments), the entire lesson is scored with the *Whole Lesson* scale. This scale is new to the 2014 version. It is comprised of 9 items, each scored on a 5-point scale (1 = not at all true of this lesson, 5 = very true of this lesson). The item content is generally consistent with the MQI's 4 domains (Richness, Errors, Working with Students, Common Core Aligned Practices), but also includes items not directly targeted by these domains (density of mathematics, student engagement, efficient use of time). Specifically, items document: (a) the quality of the lesson and the tasks embedded in it (e.g., mathematical density, richness, mathematics-focused tasks, clarity and precision); and (b) the teacher's actions (e.g., use of student ideas, remediation of student difficulties, efficient use of time, student involvement, engagement, common core aligned student practices) (National Center for Teacher Effectiveness [NCTE], n.d.). In our study, the internal consistency reliability of this scale was high (α = 0.91), confirming that items are highly interrelated and likely distributed along a unidimensional construct of mathematics quality. We therefore averaged scores on the 9 Whole Lesson items to create a score for analysis purposes.

The Whole Lesson scale also includes a 10th holistic item, serving as a single indicator of the quality of mathematics instruction across the entire lesson. This single item correlates very highly (r = 0.97) with scores on the 9-item Whole Lesson scale, and, as a single item, it is likely less reliable than the Whole Lesson scale (Nunnally & Bernstein, 1994).

Rater training and post-training agreement. Lessons were scored by three raters, each of whom had passed the MQI certification test, after completing the MQI's on-line training (National Center for Teacher Effectiveness [NCTE], n.d.). Certification involves rating four, 20-minute-long videos in agreement with master-coder scores. Agreement is computed by a distance of less than ".20 absolute deviations from the master score" (Hill et al., 2012, p. 58). However, specific guidelines for establishing rater calibration and agreement (for monitoring raters after certification) are not provided.

To calibrate our group of raters and to document rater agreement in our study, we used a system comparable to that reported for the MET project (Bell et al., 2014). Specifically, before scoring the videos for the present study, each rater watched and scored 10 kindergarten mathematics lessons that were not part of this study. After each lesson, we calculated exact agreement for ratings of segment-level domain items (scored 0-3 every 7½ minutes), and the set of 9 Whole Lesson items (scored 1-5 at the end of the lesson). The average exact agreement across pairs of raters was 72% (segment-level domains) and 58% (Whole Lesson). These levels of agreement compare well with those of the MET project's MQI raters (i.e., 53.4% to 76.6%), based on exact agreement during post-certification rater calibration activities (Bell et al., 2014).

Procedures for observing and scoring lessons. All three raters, who were blind to each-other's scores, independently scored lessons from each of the 20 teachers in accordance with the MQI's protocol. Lessons were not viewed sequentially or grouped by teacher but were randomly assigned to each rater.

Rater agreement of scored lessons. The three raters' agreement, estimated with the intra-class correlation coefficient (ICC), was high (> 0.81) for three domain scores, however the ICC was 0.43 for the Errors and Imprecision scale. As we note in the Results section, the latter estimate most likely reflects the lack of variability in teachers' Errors and Imprecision, rather than significant rater discrepancies. Generalizability analyses with this data set confirm that variance due to rater differences was at low levels (Mantzicopoulos, French, & Patrick, 2018).

Student Measures and Procedure

Overview. We collected information from both teachers and students. In both fall and spring, teachers rated each student's progress in meeting math standards, in addition to their motivation for learning mathematics. Data from students were collected during individual interviews at the end of the school year (April and May).

Standards-based Mathematics Achievement. We used the state's kindergarten mathematics standards to create an assessment of students' progress on standards-based mathematics skills. This measure (7 items) was based on teacher reports of the extent to which each student had mastered skills on standards that addressed number sense, ability to solve real world problems using numbers, understanding concepts of time, measurement, geometry, and data analysis skills. Items were rated on a scale from 1 ("does not demonstrate yet") through 5 ("independent mastery"). Teachers rated the items for each student in both the fall and spring semesters. The internal consistency reliability estimates were 0.92 (fall) and 0.95 (spring).

Standardized Mathematics Achievement (Mathematics Reasoning). We also measured students' achievement at the end of the year with two standardized math subtests (Applied Problems and Quantitative Concepts) from the Woodcock-Johnson Tests of Achievement III (WJ-III; McGrew & Woodcock, 2001). Items assess students' mathematics knowledge (number knowledge, counting, identifying shapes, telling time) and quantitative reasoning (using simple addition and subtraction problems presented in pictures). Items require verbal responses from the students and increase in difficulty. Median internal consistency reliabilities (based on split-half procedures) range between .88 and .93 with samples of 5- and 6-year old children (McGrew & Woodcock, 2001). We averaged the scores from the Applied Problems and Quantitative Concepts subtests to create a measure of *Mathematics Reasoning*, as outlined by the WJ-III technical manual (McGrew & Woodcock, 2001).

Teacher Rating Scale of Children's Mathematics Motivation. We assessed students' motivation for mathematics with two teacher-rated subscales, created by adapting the Teacher Rating Scale of Children's Motivation for Science (Mantzicopoulos, Patrick, &

Samarapungavan, 2013). This involved changing the word “science” to “math.” Teachers rated (from 1 – 5) each child on key behavioral indicators of interest in math and perceived math competence (Schunk, et al., 2014; Wentzel & Brophy, 2014). The *Interest in Math* scale (4 items; $\alpha = .94$) asks about students’ effort, enthusiasm, and interest in learning math. A sample item is “How excited or enthusiastic is s/he about doing math?” The *Need for Math Support* scale (4 items; $\alpha = .92$) refers to indicators of low perceived self-competence and interest, such as frustration, giving up when work is hard, and needing encouragement to engage in math. A sample item is “How likely is s/he to give up when math is hard?”

Motivation for Mathematics. We also assessed students’ motivation by individually administering the Motivation for Mathematics scale during the last two months of school. This scale measures children’s perceived math self-competence (analogous to expectancy and efficacy beliefs; Wigfield et al., 2015) and interest in mathematics. We created this scale by adapting the Puppet Interview Scale of Competence in and Enjoyment of Science, which has produced reliable and valid scores with kindergarteners (Mantzicopoulos, Patrick, & Samarapungavan, 2008). For some items, adaption involved changing the word “science” to “math”; for other items we exchanged topics within the science standards to match topics from the math standards.

Items addressed students’: (a) self-competence beliefs for mathematics (8 items; e.g., “I am good at adding numbers together”, “I am good at answering questions about shapes”) and (b) interest in mathematics (5 items; e.g., “I like figuring out questions with numbers”, “I like math”). Students responded on a 3-point scale (0 = no; 1 = a little; 2 = a lot). Factor analysis indicated that the items formed one factor, therefore we averaged scores on the 13 items ($\alpha = .78$) to create one scale of motivation for mathematics.

Analysis Plan

We created composite scores using the items comprising the Ambitious Mathematics Instruction scale and the Whole Lesson scale by averaging each teacher's ratings across their five lessons and the three raters. We used this procedure to minimize the effects of imprecise estimates resulting from MQI scores that are based on single observations and/or single raters (Hill et al., 2012; Kane & Staiger, 2012; Whitehurst, Chingos, & Lindquist, 2014). Although we provide descriptive data on all scales, we did not include the Errors and Imprecision score in our prediction models because, as we note in the Results section, this scale had almost no variance.

After examining the descriptive statistics and correlations of all scores we conducted a series of multilevel models (MLM; students nested within teachers) to evaluate the associations of MQI scores with students' achievement and motivation outcomes. We included students' sex, SES, Fall standards-based mathematics achievement, and Fall need for math support as covariates in each model. Model assumptions were evaluated and no violations were indicated.

We estimated four models for each of the five mathematics outcomes (i.e., mathematics reasoning, child-reported motivation, and teacher rated standards-based achievement, interest, and need for support). Model 1, the null model, included only the dependent variable. It allowed us to estimate the ICC, or percentage of variance in each outcome accounted for by differences among teachers (i.e., level 2) and within students (i.e., level 1).

Next, we estimated conditional models to investigate the associations between students' outcomes and teachers' scores on each MQI scale of interest (i.e., Ambitious Mathematics Instruction and Whole Lesson scales), while controlling for student characteristics. Specifically, in Model 2 we entered the student-level covariates and then added MQI scores in Models 3 and 4. Scores on the Ambitious Mathematics Instruction and the Whole Lesson scales were highly correlated (r = .85), therefore we examined the contribution of each in separate models. In Model 3, we added the Whole Lesson score to the set of covariates, whereas in Model 4 we substituted the Ambitious Mathematics Instruction score for the Whole Lesson score.

We entered the MQI scores as fixed effects; covariates were also entered as fixed effects at the student level. Because our focus was on the level 2 predictors (i.e., MQI scores), all variables were grand-mean centered (Enders & Tofighi, 2007). We used restricted maximum likelihood estimation (REML) to obtain the parameter estimates and employed maximum likelihood estimation to obtain the deviance estimates for model comparison of the fixed effects components (Snijders & Bosker, 2012). In presenting the results, we focus on the null model and, based on model fit results, the selected best conditional model.

To identify the best fitting model, we judged model fit across the following criteria: (a) change in within and between variance estimates; (b) deviance statistics; (c) R^2 values for both within and between variance for the models, as defined by Raudenbush and Bryk (2002); and (d) the significance of our predictors. Note that the interpretation of the R^2 values is different from what is common in multiple regression. In the two-level model, R^2 is a proportional reduction of the variance statistic, and is used to compare one model to another to understand how within and between variance is reduced through the model building process. We used the combination of indices to inform a comprehensive model evaluation process.

Results

In this section we present: (a) descriptive evidence on the MQI scales; and (b) findings from the multilevel analyses conducted to test the hypothesis that the quality of mathematics instruction, as reflected in teachers' MQI scores, is related to students' mathematics achievement and motivation outcomes.

Descriptive Statistics and Correlations

***Teachers' Mathematics Practices*.** In their lessons, teachers covered content directly connected to mathematics. This was indicated by raters' scores on the dichotomously-rated items that assessed the extent to which mathematics-relevant content was included in each segment of the lesson (M = .82; SD = .13). Ambitious Mathematics Instruction scores correlated strongly and positively with scores on its constituent scales—Richness of Mathematics, Working with Students, and Common Core Aligned Practices scales (rs ranged from .86 to .93). These high correlations: (a) are consistent with the factor analytic evidence provided by Blazar et al. (2017) that the items in these three scales form one factor; and (b) support our decision to examine the Ambitious Mathematics Instruction scale, rather than focus on each of its three constituent individual scales.

Ambitious Mathematics Instruction. The average segment-level score for the Ambitious Mathematics Instruction scale was on the low end of the 4-point rating system (M = 0.59; SD = 0.15), midway between "not present" and "low." To further unpack this score, we examined the ratings given on each of the 4 points of the scale, across all 7½ minute segments for each of the 13 items comprising the Ambitious Mathematics Instruction composite. More than 87% of the segment-level ratings did not exceed 1, approximately 10% were rated 2, and only 1% of the segments were rated 3 – the highest score on the 4-point scale.

We also examined descriptive differences when grouping the segment-level scores by the items that made up each subscale. Of the Richness items, 12% were rated 2, 27% were rated 1, and 60% were rated 0. The proportions of Working with Students items rated 2, 1, and 0 were 11%, 48%, and 40%, respectively. Of the Common Core Aligned Practices items, 8% were rated 2, 31% were rated 1, and 60% were rated 0.

Whole Lesson Scale. In comparison to the scores given in 7½ minute segments, the Whole Lesson scale scores indicated that the overall quality of instruction was in the mid-range (M = 3.09; SD = 0.32). When examined at the item level across all raters and lessons, the distribution of ratings (5 = very true of this lesson; 1 = not at all true of this lesson) was as follows: 5 (11%), 4 (17%), 3 (46%), 2 (21%), and 1 (5%). Thus, 74% of the ratings were at, or above, a rating of 3. Despite the difference in the distribution of scores between the Ambitious Mathematics Instruction and the Whole Lesson scales, the correlation between the two was strong (r = 0.86), suggesting that the two scales assessed overlapping constructs.

Errors and Imprecision. As we noted earlier, teachers' scores on this scale lacked variability. Their scores, averaged across their 5 lessons and the three raters, ranged from 0.0 to 0.08 (M = 0.02). Across the 7½ minute segments, 99% of the items were rated 0, providing a consistent picture of instruction that is free of mathematical errors and inaccuracies. When Errors and Imprecision scores were averaged at the lesson level across the scale's 3 items, less than 4% of the scores were rated above 0; the range was from 0.20 to 1.40. Only three lessons had an average Errors and Imprecision score of 1.0 or greater (1 = "low" on the MQI). Of these, two lessons received an average score of 1.0 and one received an average of 1.4, indicating that the few errors made by teachers represented brief instances of imprecision or lack of clarity.

Students' Achievement and Motivation. Table 1 presents the correlations between the student-level predictors and outcomes. The correlations were in the expected direction and strength. No correlations were strong enough to raise multicollinearity concerns.

Predicting End-of-year Mathematics Outcomes from MQI Scores

To examine the hypothesis that teachers' practices are associated with students' mathematics achievement and motivation, we focused on two MQI scales: Ambitious

Mathematics Instruction (items rated 0-3 every 7½ minutes) and Whole Lesson (rated 1-5 at the end of the lesson). Because students were nested within teachers we used multilevel modeling, entering the two MQI scale scores separately, given the strong correlation between them. This allowed us to evaluate whether the Whole Lesson score accounted for the variance between teachers as well as or better than the average of the lesson segment domain items.

Null models. We first estimated Model 1, with no predictors, to determine the degree of similarity between students with the same teacher and decompose the outcome variance between levels. This provided an estimate of the percentage of variance in each mathematics outcome that could be accounted for by the nesting of students within teachers. Because our predictors of interest, reflecting the quality of mathematics instruction, were at the teacher level, it was especially important to establish the extent to which differences between teachers accounted for variance in students' achievement and student motivation. If we found, for example, that there was little between-teacher (level 2) variance in mathematics outcomes, then we would have little variance to explain with the teacher-level predictors (i.e., MQI scores) used in this study.

Based on Model 1 estimates, we found that there was much more variability between teachers in terms of ratings of their students, compared to student-reported motivation and scores on the Math Reasoning composite. Specifically, the ICCs, estimated from the null model and reported in Tables 2 through 6, were largest for the teacher-reported outcomes: 38.7% for students' standards-based math achievement, 28.4% for students' interest in mathematics, and 23.8% for students' need for math support. The differences between teachers were smallest for the student-reported outcomes: 6.3% and 3.5% of the total variance in student's standardized test scores on mathematics reasoning and student-reported motivation respectively. Given the small ICCs for the latter set of outcomes, differences in teachers' mathematics practices are expected to make very small, if any, contributions to our explanatory models.

In Models 2-4, reported next, we unpack the available teacher and student variance reflected in the ICCs. That is, for each outcome we examine the extent to which: (a) the set of student covariates account for the available student variance; and (b) teachers' mathematical strategies, as reflected in the MQI scales, explain the available teacher-level variance.

Multi-level models (Models 2 – 4). Results of the multi-level analyses for each outcome are shown in Tables 2-6, each of which show a series of four models. Model 1 is the null model; the covariates (SES, sex, and students' teacher-rated fall standards-based math achievement and need for math support) are added in Model 2; the Whole Lesson score is added to the covariates in Model 3; and the Ambitious Mathematics Instruction score, without the Whole Lesson score, is added to the covariates in Model 4.

Students' mathematics reasoning. Only student-level predictors – the level 1 student covariates – were significant, explaining 39.6% of the available student-level variance in mathematics reasoning (Table 2). Students receiving free or reduced-cost lunch tended to score lower on mathematics reasoning than did their peers who were on paid lunch status. Females tended to have lower scores compared to their male counterparts. Of interest, students who entered kindergarten with lower mathematics skills and who required more teacher support for learning math were likely to score lower on this standardized math reasoning composite at the end of the year. As seen in Table 2, Model 2 (with student-level variables only) was the best fitting model. The addition of either MQI score (Models 3 and 4) did not explain additional variance in student's math reasoning.

Student-reported math motivation. No student or teacher level variable explained a statistically significant amount of variability in student-reported motivation for math (Table 3).

Neither the level 1 (within students) nor the level 2 (between teachers) variance estimates changed from the null model across Models 2-4 (i.e., with student covariates and MQI scores). Recall, however, that only 3.5% of the variability in student-reported motivation was between teachers, and the teacher-level variance estimates were not significant. Additional predictors at the individual level would be needed to explain the remaining significant student-level variance.

Teacher-reported math standards-based achievement and motivation. We identified a series of consistent results in the analyses examining the teacher-rated student outcomes of standards-based math achievement, interest in math, and need for math support (Tables 4-6). Across all outcomes and models, neither child SES nor sex was a significant predictor. In contrast, across all outcomes and models, teachers' fall ratings of students' need for math support and standards-based math achievement were significant predictors of their spring ratings. These two predictors (entered before the MQI scores) accounted for approximately 37.8% to 49.7% of the available student-level variance.

Across all three outcomes, the Whole Lesson scale (Model 3) explained more of the between-class (i.e., teacher) variability than did the Ambitious Mathematics Instruction score (Model 4). That is, approximately 21% to 43% of the teacher-level variability in Model 2 was accounted for after adding the Whole Lesson scores. In contrast, even though the Ambitious Mathematics Instruction score was statistically significant for two of the three student outcomes (standards-based math achievement and Need for Support in Mathematics), it accounted for only 7% to 19% of the between-teacher variance. For all three outcomes, significant unexplained variance remained, both at the student and teacher level, as indicated by the significant student and teacher residual parameters in the final models.

To aid interpretation of Model 3 and the magnitude of the Whole Lesson scale, we compare plausible values for the outcomes (teachers' ratings of students) with plausible values from Model 2. Plausible values are interpreted as confidence intervals around the parameter of interest, in this case the mean of the outcome variable (Raudenbush & Bryk, 2002).

When predicting students' standards-based mathematics achievement, plausible values for Model 2 and Model 3 ranged from 3.28 to 5.25 and 3.44 to 5.09, respectively. The tighter range for Model 3 reflects the addition of the MQI Whole Lesson scale and the associated greater proportion of between-teacher variance that is accounted for, as seen in Table 4. This suggests students' expected teacher-rated achievement would be about 48% higher in the highest- vs. lowest-rated classroom, reflecting the positive parameter estimate for the Whole Lesson scale.

For student interest in mathematics, the plausible values of Model 2 and Model 3 were 3.23 to 5.19 and 3.33 to 5.08, respectively. The magnitude is similar to the achievement of mathematics standards, where a student's expected rating of math interest would be 53% higher in the highest-rated, compared to the lowest rated, classroom.

Finally, for students' need for support in mathematics*,* Model 2 and Model 3 plausible values were 1.13 to 2.60 and 1.31 to 2.42, respectively. Again, the tighter range for Model 3 reflects the addition of the Whole Lesson scale in the model and that 43.3% of the between-teacher variance was explained. The negative parameter estimate suggests that a higher teacher Whole Lesson score is related to a lower rating for a student indicating less need for support.

Discussion

Assessing mathematics instruction with observation measures has the potential to inform those interested in mathematics education about the links between instruction and important student outcomes, such as mathematics achievement and motivation for learning math. However, despite consensus on the long-term consequences of early mathematics learning for student success (National Research Council, 2001), there is a dearth of evidence from mathematics classrooms in the early school grades. To address this void, we used the MQI in a range of diverse kindergarten classrooms to examine associations between the quality of mathematics teaching and young students' mathematics learning and motivation.

Our findings add to the literature on effective mathematics instruction in the following ways. First, we provide descriptive data about the quality of mathematics instruction in kindergarten, using the 2014 version of the MQI, including its new Whole Lesson scale, which is rated only at the end of the lesson and has fewer items than the original scale (Hill, 2014). To our knowledge this is the first published study with the Whole Lesson scale. Second, we identify that the observed quality of mathematics instruction that kindergarteners receive throughout the spring semester: (a) predicts their achievement and motivation to engage with mathematics, as rated by their teachers; but (b) is not related to their performance on a standardized test of mathematical reasoning or to their self-reported motivation for math.

Measuring Kindergarten Teachers' Mathematics Practices

Ambitious mathematics instruction. Ambitious mathematics instruction involves practices such as using mathematical language, identifying multiple solutions, and eliciting and responding to students' explanations and reasoning. Teachers in our study were generally rated as engaging in low levels of these practices; the average score fell midway between "low" and "not present" and scores were skewed towards the low end. Interestingly, despite teachers' low ratings, their lessons were consistent with the content of the state's kindergarten mathematics standards. In addition to the specific content and concepts to be taught, the standards emphasize the importance of students developing conceptual understanding and the ability to synthesize and apply mathematics. However, the standards do not stipulate how specific mathematics content should be taught. Thus, for meaningful mathematics instruction to occur, more guidance for teachers may be necessary than identifying curricular content to be mastered.

Kindergarten teachers' use of ambitious mathematics practices are comparable to those reported with upper elementary (Blazar, 2015) and middle school (Bell et al., 2014) teachers. These studies used earlier versions of the MQI, where items were scored on a 3-point scale, and not all scales were comparable to the 2014 version we used. However, upper elementary teachers' lessons were also rated below the mid-point on Ambitious Mathematics Instruction (Blazar & Kraft, 2017), as were middle school teachers' lessons on some of its constituent domains (e.g., Richness of Mathematics, Working with Students and Mathematics; Bell et al., 2014; Kane & Staiger, 2012). Moreover, middle school teachers' scores were skewed to the low end (Kane & Staiger, 2012), in a similar fashion to the scores of the teachers' in our study.

Errors and imprecision. Errors and Imprecision was the only domain in which the kindergarten teachers were uniformly rated positively. That is, the teachers in our study made few, if any, mathematical errors. The mathematical content of each lesson was generally clear and without ambiguities in mathematical concepts, procedures, or language. Alternatively, the Errors and Imprecision scale may not be sensitive enough to document errors and ambiguities characteristic of early mathematics instruction.

For our sample of kindergarten teachers, there was insufficient variability in their ratings of errors and imprecision for the scale to be included in our analyses. Given our relatively small sample of teachers, however, future research is needed to examine whether this pattern of results is found in other kindergarten classrooms, as well as other early elementary grades.

Like with the Ambitious Mathematics Instruction scale, our kindergarten teachers' ratings for Errors and Imprecision were consistent with those of upper elementary and middle school teachers. Evidence from the MET project, based on middle school classrooms, found that of the 93.4% of lessons containing classroom work connected to mathematics, “very few … were wrought with mathematical errors and imprecision” (Kane & Staiger, 2012, p. 24). Similarly, data from upper elementary classrooms (Blazar & Kraft, 2017) indicated that that, on average, teachers did not commit major mathematical errors (M = 1.12 on a 1-3 scale). However, as we discuss in the next section, teachers’ errors predicted student outcomes.

Whole lesson scale. As we have noted, the Whole Lesson scale is a new addition to the MQI. Items are scored at the end of the lesson on a 5-point scale. Teachers’ Whole Lesson ratings were strongly correlated with scores on Ambitious Mathematics Instruction. At the same time, average ratings for the Whole Lesson scale were considerably more positive than the average ratings on the scales comprising Ambitious Mathematics Instruction. Specifically, Whole Lesson scores were in the mid-range of the continuum, which, according to the MQI training document (NCTE, n.d.), reflect typical instruction. Descriptively, mid-range ratings of Whole Lesson items represent lessons in which the teacher: (a) covers a reasonable amount of content, although without adequate evidence that topics are interconnected or move toward big ideas; (b) proceeds relatively smoothly (i.e., with few distractions) through topics; (c) includes some aspects of rich mathematics (e.g., representations, student explanations, multiple solutions); (d) attends to and remediates student difficulties, albeit briefly; (e) acknowledges student ideas without further extending them; and (f) engages students with the mathematics content but only occasionally in substantive ways (NCTE, n.d.). At this time, it is not clear why mean ratings from the Ambitious Mathematics Instruction scale convey a picture of teachers’ mathematics practices that is qualitatively lower than that reflected in the Whole Lesson codes.

Prediction of Mathematics Outcomes

Mathematics achievement. Teachers’ ratings of students’ end-of-year progress toward meeting kindergarten mathematics standards was predicted by scores on the quality of mathematics instruction throughout the spring semester, particularly as measured by the Whole Lesson scale. This finding is not surprising, considering that teachers are keenly aware of (and accountable for) the content that their students need to master in order to meet state-specific, grade-level standards. Their practices are thus intended to directly target the development of standards-specific skills in their students. We found that how teachers address grade-relevant mathematics content is related to their students' progress. Although this inference is based on teacher reports rather than direct assessments of students, it is noteworthy given evidence that teachers are accurate judges of students' progress and skills across different content areas, including mathematics (Bassok & Latham, 2017; Südkamp, Kaiser, & Möller, 2012).

In contrast, students' general mathematical reasoning – assessed by a standardized achievement measure not linked to specific kindergarten standards – was unrelated to instructional quality, at least as measured by the MQI. This is noteworthy, given that standardized test scores are typically used to make comparisons of students' progress across districts, states, and nations. Our findings suggest that, at least in kindergarten, students' performance on mathematics-specific, yet broadly-focused, achievement tests that are distal to the instructional context may not reflect the contributions that early instructional practices make to student learning. Alternatively, as we note later in this section, the MQI may not be sensitive to those mathematics practices in the early grades (e.g., drill and practice activities) that impact students' achievement on standardized tests, at least in the short term.

Motivation for mathematics. Teachers' end-of-year assessments of their students' need for support and encouragement to learn mathematics was associated negatively with the quality of their mathematics instruction. However, the Whole Lesson scale explained more than twice the variance between teachers in students' need for support than the Ambitious Mathematics Instruction scale did. Of note, students with low teacher ratings on their entry-level mathematics skills and motivation (i.e., rated high in need for teacher support) were likely to be rated as needing high levels of support for learning math. We interpret these findings in light of evidence that teachers of low achieving and high needs students tend to: (a) be less qualified than teachers of high achieving students (e.g., Kalogredes & Loeb, 2013); (b) focus on remediating deficits through basic skills instruction, using drill and practice approaches (e.g., Means & Knapp, 1991). This, in turn, leaves less instructional time for cognitively demanding practices that call, for example, for students' reasoning about multiple solutions, explanations, or generalizations.

Students' interest in mathematics, as rated by their teachers at the end of the year, was also predicted by their teacher's Whole Lesson score, but not the aggregated segment scores comprising Ambitious Mathematics Instruction. In contrast, students' own reports of their motivation for mathematics were not related to scores on either of the MQI's instructional scales.

For developmental reasons, which include young children's generally overoptimistic evaluation of their competence (Stipek & Mac Iver, 1989), measuring motivational beliefs in the early years of school is challenging. However, we are cautious about attributing the nonsignificant findings associated with students' self-reported motivation to these measurement difficulties. There is consistent evidence that children's subject-specific motivational conceptions are grounded in the social processes that support children's engaged participation with the content (Patrick & Mantzicopoulos, 2015). Our confidence in the student motivation scale is also supported by results from a recent study that employed the science version of this measure; kindergarteners' self-reported motivation for science was significantly related to the quality of teachers' science practices (Mantzicopoulos, Patrick, Strati, & Watson, 2018). Science instructional quality, however, was assessed with a measure different from the mathematics-specific MQI used in the present study. Perhaps the MQI, with its specific focus on mathematics and limited attention to the social environment, is not sufficiently sensitive to classroom social processes and norms that foster the development of students' motivation to learn mathematics.

Motivation has important consequences for students' school success (Wigfield et al., 2015) and is gaining attention in light of the recent Every Student Succeeds Act (USDOE, 2016). Therefore, it is critical for researchers to document associations between teachers' use of specific practices during mathematics lessons and students' motivation for learning the content taught.

Comparison with other studies. In addition to their similarity in descriptive statistics, discussed earlier in this section, some of our findings about the associations between MQI scores and student outcomes parallel those from research that used an earlier version of the MQI in 4th and 5th grade classrooms (Blazar & Kraft, 2017). In that study, as with ours, Ambitious Mathematics Instruction did not predict achievement on a standardized test, or student-rated motivation. However, upper elementary, but not kindergarten, teachers' mathematical errors – which were of extremely low incidence in our kindergarten classrooms – did predict scores on those outcomes. Of interest, 4th and 5th grade teachers who made more mathematical errors were more likely to have students with lower self-efficacy for mathematics, be less happy in class, and perform less well on a standardized mathematics test, although the latter relationship was "marginally significant" (Blazar & Kraft, 2017, p. 158).

Perhaps the salience of specific dimensions of teaching varies significantly by grade level or is affected by the mathematics curricula teachers use. Support for this assertion is gleaned from recent studies that examined early mathematics achievement as a function of specific curricula (Agodini & Harris, 2016) or teacher-reported mathematics practices in kindergarten

(Bottia et al., 2014) and longitudinally (i.e., from kindergarten to 1st grade; Guarino, et al., 2013). Data from a national study of kindergarten teachers' self-reported mathematics practices show that frequent use of drill and practice is an effective way to increase the mathematics achievement of both White and Latino students (Bottia et al., 2014). Indeed, the acquisition of early number skills through "traditional approaches" to teaching has been noted as an appropriate strategy in other studies (e.g., Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Guarino et al., 2013).

Converging evidence was found in a randomized control trial of 1st and 2nd grade mathematics curricula (Agodini & Harris, 2016). Common to curricula identified as most effective in the early grades was the provision of daily, repeated opportunities for young students to: (a) routinely engage "with concepts, facts, and procedures" (p. 233); (b) develop procedural fluency through drill and practice activities; and (c) engage with other students on the mathematical content of a lesson (Agodini & Harris, 2016).

As a grade-level independent assessment, the MQI does not attend to the development of number and procedural fluency, even though young children "need a great deal of practice doing a task, even after they can do it correctly" (National Research Council, 2009, p. 128). Perhaps, emphasis on practicing numbers and operations may take time away from instruction that focuses on mathematical meaning-making and explanation—practices that are explicitly documented in the MQI. On a post-hoc basis, this may explain why the practices of kindergarten teachers in our study were rated on the low end of the Ambitious Mathematics Instruction scale. Of note, research also indicates that although "traditional" drill-and-practice instruction has a positive impact on the development of students' numeracy skills in kindergarten, other practices (e.g., student explanations) become significant in first grade by contributing to students' mathematics competencies (Guarino et al., 2013). Although this conclusion is drawn from teacher reports on the ECLS-K, it suggests that practices comprising the domain of Ambitious Mathematics Instruction (i.e., students' offering mathematical explanations, reasoning abstractly, critiquing the work of other students) may be differentially relevant to student learning at different grade levels. This issue merits attention in future research.

Conclusions and Limitations

Despite the attractiveness of the MQI as a measure of sound mathematics instructional practices, our research offers limited support for its use to document effective mathematics teaching in kindergarten, given its small associations with student achievement and motivation. However, our findings should be interpreted in light of potential limitations of our study. The small number of teachers may limit the generalizability of our results. It also precluded us from conducting psychometric analyses to confirm the dimensionality of the MQI's segment-rated items, as well as the Whole Lesson scale; empirical evidence on this issue is needed. Additionally, research is needed to examine if the MQI predicts kindergarteners' mathematics outcomes better with attention to: (a) the developmental significance of mathematics skills from kindergarten to first grade; and (b) the particular curricula, and related practices, that address these skills. Similar research is also warranted for first and second grades.

In closing, our findings call for continued research on the observation protocols that document the quality of mathematics practices. Because these measures highlight the types and range of effective instruction, it is critical that their use is informed by rich and robust evidence that (a) describes teachers' practices across content areas and grade levels, and (b) documents the learning and motivational consequences of these practices for students through lenses that are proximal as well as distal to the instructional context.

References

Agodini, R., & Harris, B. (2016). How teacher and classroom characteristics moderate the effects of four elementary math curricula. *The Elementary School Journal*, *117*, 216-236.

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences, 20*, 427-435.

Aunola, K., Leskinen, E., Lerkkanen, M. & Nurmi, J. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96,* 699-713.

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator, 29*(1), 20-22. 43-46.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*, 389-407.

Bassok, D., & Latham, S. (2017). Kids today: The rise in children's academic skills at kindergarten entry. *Educational Researcher, 46*, 7-20.

Bell, C. A., Qi, Y., Croft, A. J., Leusner, D., McCaffrey, D. F., Gitomer, D. H., & Pianta, R. C. (2014). Improving observational score quality: Challenges in observer thinking. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 50-97). San Francisco, CA: Jossey Bass.

Bill & Melinda Gates Foundation. (2013). *Measures of effective teaching*. http://k12education.gatesfoundation.org

Blazar, D. (2015). Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review, 48*, 16-29.

Blazar, D., Braslow, D., Charalambous, C. Y., & Hill, H. C. (2017). Attending to general and content-specific dimensions of teaching: Exploring factors across two observation instruments. *Educational Assessment, 22*, 71-94.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis*, *39*, 146-170.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and Instruction. *The Elementary School Journal*, *108*, 115-130.

Bottia, M. C., Moller, S., Mickelson, R. A., & Stearns, E. (2014). Foundations of mathematics achievement: Instructional practices and diverse kindergarten students. *The Elementary School Journal, 115*, 124-150.

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61*, 365-374.

Center on Great Teachers and Leaders (2013). *Measures of teacher performance* [Databases on state and principal evaluation policies]. Retrieved from http://resource.tqsource.org/stateevaldb/Compare50States.aspx

Charalambous, C. Y., & Praetorious, A. (2018). Studying mathematics instruction through different lenses: Setting the ground for understanding instructional quality more comprehensively. *ZDM: Mathematics Education, 50*, 355-366.

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher, 45*, 378-387.

Danielson, C. (2013). *The Framework for Teaching evaluation instrument: 2013 edition*. Princeton, NJ: The Danielson Group.

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121-138.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*, 519-533.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher, 44,* 96-104.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93*, 3-13.

Guarino, C., Dieterle, S. G., Bargagliotti, A. D., & Mason, W. M. (2013). What can we learn about effective early mathematics teaching? A framework for estimating causal effects using longitudinal survey data. *Journal of Research on Educational Effectiveness, 6*, 164-198.

Hickey, D. T., Zuiker, S. J., Taasoobshirazi, G., Schafer, N. J., & Michael, M A. (2006). *Studies in Educational Evaluation, 32*, 180-601.

Hill, H. (2011). *Mathematical Quality of Instruction*. Paper presented at the National Center for Teacher Effectiveness Conference. Retrieved from https://cepr.harvard.edu/ncte-conference-2011

Hill, H. (2014). *Mathematical Quality of Instruction* (MQI: 4-point version). Retrieved from: https://cepr.harvard.edu/mqi

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking "pedagogical content knowledge": Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for*

Research in Mathematics Education, 39, 372-400.

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*, 56-64.

Horizon Research (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States*. Retrieved from: http://www.horizon-research.com/insidetheclassroom/reports/looking/complete.pdf

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching*. Seattle WA: Bill & Melinda Gates Foundation.

Kalogredes, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher, 42*, 304-316.

Learning Mathematics for Teaching Project. (2011). Measuring the Mathematical Quality of Instruction. *Journal of Mathematics Teacher Education, 14*, 25-47.

Lerkkanen, M., Kiuru, N., Pakarinen, E., Viljaranta, J., Poikkeus, A., Rasku-Puttonen, H., . . . Nurmi, J. (2012). The role of teaching practices in the development of children's interest in reading and mathematics in kindergarten. *Contemporary Educational Psychology, 17*, 266-279.

Mantzicopoulos, P., French, B. F., & Patrick, H. (2018). The Mathematical Quality of Instruction (MQI) in kindergarten: An evaluation of the stability of the MQI using Generalizability Theory. *Early Education and Development, 29*, 893-908.

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2008). Young children's motivational beliefs about learning science. *Early Childhood Research Quarterly, 23*, 378-394.

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2013). Science literacy in school and home contexts: Kindergarteners' science achievement and motivation. *Cognition and Instruction, 31,* 62-119.

Mantzicopoulos, P. Y., Patrick, H., Strati, A., & Watson, J. S. (2018). Predicting kindergarteners' achievement and motivation from observational measures of teaching effectiveness. *Journal of Experimental Education, 86,* 214-232.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical Manual. Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Means, B., & Knapp, M. S. (1991). Cognitive approaches to teaching advanced skills to educationally disadvantaged students. *Phi Delta Kappan, 73*, 282-289.

Miele, D. B., & Wigfield, A. (2014). Quantitative and qualitative relations between motivation and critical-analytic thinking. *Educational Psychology Review, 26*, 519-541.

National Center for Teacher Effectiveness (n.d.). Whole lesson codes [*Training module for the Mathematical Quality of Instruction*]. Unpublished module. Retrieved from http://isites.harvard.edu/icb/icb.do?keyword=mqi_training

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

National Research Council (2001). *Adding it up: Helping children learn mathematics*. Washington DC: The National Academies Press.

National Research Council (2009). *Mathematics learning in early childhood: Paths towards excellence and equity*. Washington, DC: The National Academies Press.

Nolen, S. B. (2001). Constructing literacy in the kindergarten: Task structure, collaboration, and motivation. *Cognition and Instruction, 18*, 95-142.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. (3rd Ed). New York: McGraw-Hill.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*, 237-257.

Patrick, H., & Mantzicopoulos, P. (2015). The role of meaning systems in the development of motivation. In C. M. Rubie-Davies, J. M. Stephens, & P. Watson (Eds.), *The Routledge international handbook of social psychology of the classroom* (pp. 67-79). New York: Routledge.

Patrick, H., Mantzicopoulos, P., & Sears, D. (2012). Effective classrooms. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 2, pp. 443-469). Washington, DC: American Psychological Association.

Piburn, M., Sawada, D., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). *Reformed Teaching Observation Protocol (RTOP: Reference manual).* Retrieved from http://files.eric.ed.gov/fulltext/ED447205.pdf

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Schunk, D. H., Meece, J. L., & Pintrich, P. R. (2014). *Motivation in education: Theory, research, and applications* (4th ed.). Upper Saddle River, NJ: Pearson.

Shanley, L. (2016). Evaluating longitudinal mathematics achievement growth: Modeling and measurement considerations for assessing academic progress. *Educational Researcher, 45*, 347-357.

Skinner, E. A. (2016). Engagement and disaffection as central to processes of motivational resilience and development. In K. R. Wentzel & D. R. Miele (Eds). *Handbook of motivation at school* (2nd edition) (pp. 145-168). New York, NY: Routledge.

Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and applied* *multilevel analysis.* (2nd Ed.). Thousand Oaks, CA: Sage.

Stipek, D., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development, 60*, 521-538.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*, 743-762.

The UTeach Institute. (2014). *UTeach observation protocol for mathematics and science*. Retrieved from https://utop.uteach.utexas.edu/node/3

U.S. Department of Education. (2009). *Race to the Top program: Executive summary*. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education. (2016). *Every Student Succeeds Act*.

Watts, T. W., Duncan, G. J., Chen, M., Claessens, A., Davis-Kean, P. E., Duckworth, K., . . . Susperreguy, M. I. (2015). The role of mediators in the development of longitudinal mathematics achievement associations. *Child Development, 86*, 1892-1907.

Wentzel, K. R., & Brophy, J. E. (2014). *Motivating students to learn* (4th ed.). New York: Routledge.

Whitehurst, G. J., Chingos, M. M., & Lindquist, K. (2014). *Evaluating teachers with classroom observations. Lessons learned in four districts*. Retrieved from: http://www.brookings.edu/~/media/research/files/reports/2014/05/13-teacher-evaluation/evaluating-teachers-with-classroom-observations.pdf

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science, Vol. 3: Socioemotional processes* (7th ed., pp. 657–700). Hoboken, NJ: John Wiley.

Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's motivation for reading: Domain specificity and instructional influences. *The Journal of Educational Research, 97*, 299-309.

Table 1

Descriptive Statistics and Correlations between Student Measures

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Sex[a]									
2. SES[b]	.05								
Fall									
3. Standards-based Math Achievement[c]	.04	.11*							
4. Need for Support in Math[c]	-.08	-.22**	-.67**						
Spring									
5. Math Reasoning[d]	-.11*	.23**	.44**	-.50**					
6. Standards-based Math Achievement[c]	.01	.22**	.54**	-.55**	.53**				
7. Interest in Math[c]	.03	.17**	.39**	-.51**	.42**	.63**			
8. Need for Support in Math[c]	-.05	-.21**	-.58**	.69**	-.53**	-.73**	-.71**		
9. Motivation for Math[e]	.01	.00	-.10*	-.08	.08	-.01	.02	.06	
Mean	1.46	0.47	3.12	2.23	16.27	4.29	4.28	1.83	1.55
SD	0.50	0.50	0.88	.09	2.51	0.85	0.85	0.97	0.38

Note. [a]Scored 1 = boys, 2 = girls; [b]Scored 1 = self-paid lunch, 0 = free or reduced-cost lunch; [c]Score based on teacher ratings; [d]Score on Woodcock-Johnson III Math Reasoning Composite; [e]Score based on child motivation scale.
*p < .05; **p <.01.

Table 2

Student Background and Teacher MQI Scores Predicting Student Mathematics Reasoning

Variable	Model 1		Model 2		Model 3		Model 4	
	β	SE	β	SE	β	SE	β	SE
Fixed Parameters								
Intercept	16.270*	.204	17.265*	.420	17.260*	.415	17.271*	.421
Student Variables								
Sex			-.636*	.230	-.642*	.230	-.642*	.230
SES			.502*	.243	.494*	.243	.502*	.243
SBAch-Math[a]: Fall			1.053*	.211	1.079*	.211	1.078*	.212
Math Support: Fall			-.949*	.200	-.938*	.199	-.942	.200
Teacher Variables: MQI								
Whole Lesson					1.241	.805		
Ambitious Math Instruction							1.777	1.810
Random Parameters								
Student (within)	5.941*	.517	3.590*	.315	3.590*	.314	3.586*	.314
Teacher (between)	.401	.278	1.035*	.443	.959*	.424	1.058	.457
Model Fit								
-2LL (deviance)	1328.890		**1200.026**		1197.506		1198.981	
ICC	6.32%							
R^2*within*			39.57%[b]		--		—	
R^2*between*			--		7.34%[c]		--	

Note. [a] SBAch-Math = standards-based mathematics achievement.
[b]R^2 values represent the variance explained in comparison to the variance in Model 1.
[c]R^2 represents the variance explained in comparison to the variance in model 2. R^2 values are not presented when variances did not change or increased. Bold deviance value indicates the preferred final model given the results.
* $p < .05$.

Table 3

Student Background and Teacher MQI Scores Predicting Student-reported Motivation for Mathematics

Variable	Model 1		Model 2		Model 3		Model 4	
	β	**SE**	**β**	**SE**	**β**	**SE**	**β**	**SE**
Fixed Parameters								
Intercept	1.546*	.028	1.540*	.071	1.539*	.071	1.541*	.071
Student Variables								
Sex			.005	.045	.004	.045	.004	.045
SES			-.016	.046	-.017	.046	-.016	.046
SBAch-Math [a]: Fall			.032	.037	.038	.037	.034	.037
Math Support: Fall			-.023	.036	-.021	.036	-.023	.036
Teacher Variables: MQI								
Whole Lesson					.096	.096		
Ambitious Math Instruction							.082	.209
Random Parameters								
Student (within)	.138*	.012	.137*	.012	.137*	.012	.137*	.012
Teacher (between)	.005	.006	.007	.006	.007	.006	.008	.007
Model Fit								
-2LL (deviance)	251.048		**247.649**		246.478		247.447	
ICC	3.50%							
R^2within	--		0.72%[b]		--		--	
R^2between	--		--		--		--	

Note. [a] SBAch-Math = standards-based mathematics achievement.
[b]R^2 values represent the variance explained in comparison to the variance in Model 1.
[c]R^2 represents the variance explained in comparison to the variance in model 2. R^2 values are not presented when variances did not change or increased. Bold deviance value indicates the preferred final model given the results.
* $p < .05$.

Table 4

Student Background and Teacher MQI Scores Predicting Students' Standards-based Mathematics Achievement

Variable	Model 1		Model 2		Model 3		Model 4	
	β	SE	β	SE	β	SE	β	SE
Fixed Parameters								
Intercept	4.203*	.129	4.267*	.146	4.268*	.132	4.271*	.138
Student Variables								
Sex			-.025	.061	-.027	.061	-.028	.061
SES			.092	.065	.092	.064	.095	.064
SBAch-Math [a]: Fall			.414*	.058	.422*	.058	.420*	.058
Math Support: Fall			-.261*	.054	-.259*	.054	-.260*	.054
Teacher Variables (MQI)								
Whole Lesson					.926*	.318		
Ambitious Math Instruction							1.630*	.750
Random Parameters								
Student (within)	.468*	.041	.248*	.022	.248*	.022	.248*	.022
Teacher (between)	.296*	.110	.252*	.089	.177*	.066	.209*	.077
Model Fit								
-2LL (deviance)	634.643		458.832		**451.011**		454.134	
ICC	38.74%							
R^2within			47.01%[b]		--		--	
R^2between					29.76%[c]		17.06%[c]	

Note. [a] SBAch-Math = standards-based mathematics achievement.
[b] R^2 values represent the variance explained in comparison to the variance in Model 1.
[c] R^2 represents the variance explained in comparison to the variance in model 2. R^2 values are not presented when variances did not change or increased. Bold deviance value indicates the preferred final model given the results.
* $p < .05$.

Table 5

Student Background and Teacher MQI Scores Predicting Teacher-Rated Student Interest in Mathematics

Variable	**Model 1**		**Model 2**		**Model 3**		**Model 4**	
	β	**SE**	**β**	**SE**	**β**	**SE**	**β**	**SE**
Fixed Parameters								
Intercept	4.240*	.113	4.214*	.156	4.214*	.147	4.217*	.153
Student Variables								
Sex			.034	.070	.031	.070	.031	.070
SES			.087	.075	.084	.075	.088	.075
SBAch-Math [a] -Fall			.269*	.067	.275*	.067	.275*	.067
Math Support-Fall			-.359*	.063	-.357*	.062	-.357*	.062
Teacher Variables: MQI								
Whole Lesson					.816*	.342		
Ambitious Math Instruction							1.280	.799
Random Parameters								
Student (within)	.540*	.047	.336*	.029	.335*	.029	.335*	.029
Teacher (between)	.214*	.084	.251*	.092	.199*	.075	.233*	.087
Model Fit								
-2LL (deviance)	669.392		540.925		**535.291**		538.226	
ICC	28.38%							
R^2within			37.78%[b]		0.30%[c]		0.30%[c]	
R^2between			--		20.72%[c]		7.17%[c]	

Note. [a] SBAch-Math = standards-based mathematics achievement.
[b] R^2 values represent the variance explained in comparison to the variance in Model 1.
[c] R^2 represents the variance explained in comparison to the variance in model 2. R^2 values are not presented when variances did not change or increased. Bold deviance value indicates the preferred final model given the results.
* $p < .05$.

Table 6

Student Background and Teacher MQI Scores Predicting Teacher-Rated Student Need for Support in Mathematics

Variable	**Model 1**		**Model 2**		**Model 3**		**Model 4**	
	β	**SE**	**β**	**SE**	**β**	**SE**	**β**	**SE**
Fixed Parameters								
Intercept	1.887*	.120	1.869*	.141	1.866*	.130	1.862*	.136
Student Variables								
Sex			-.009	.074	-.003	.074	-.003	.074
SES			-.059	.078	-.062	.078	-.066	.078
SBAch-Math[a]: Fall			-.392*	.069	-.393*	.067	-.398*	.068
Math Support: Fall			.472*	.065	.480*	.063	.474*	.064
Teacher Variables: MQI								
Whole Lesson					-.823*	.239		
Ambitious Math Instruction							-1.307*	.592
Random Parameters								
Student (within)	.738*	.064	.371*	.033	.370*	.032	.371*	.032
Teacher (between)	.230*	.094	.141*	.059	.080*	.037	.114*	.050
Model Fit								
-2LL (deviance)	754.527		575.976		**547.327**		553.040	
ICC	23.76%							
R²within			49.73%[b]		0.27%[c]		--	
R²between			--		43.26%[c]		19.15%[c]	

Note. [a] SBAch-Math = standards-based mathematics achievement.
[b] R^2 values represent the variance explained in comparison to the variance in Model 1.
[c] R^2 represents the variance explained in comparison to the variance in model 2. R^2 values are not presented when variances did not change or increased. Bold deviance value indicates the preferred final model given the results.
* $p < .05$.