American Educational Research Journal
August 2015, Vol. 52, No. 4, pp. 787–821
DOI: 10.3102/0002831215579484

Mathematical Knowledge for Teaching, Standards-Based Mathematics Teaching Practices, and Student Achievement in the Context of the *Responsive Classroom* Approach

Erin R. Ottmar
Indiana University and University of Richmond
Sara E. Rimm-Kaufman
University of Virginia
Ross A. Larsen
Brigham Young University
Robert Q. Berry
University of Virginia

This study investigates the effectiveness of the Responsive Classroom (RC) *approach, a social and emotional learning intervention, on changing the relations between mathematics teacher and classroom inputs (mathematical knowledge for teaching [MKT] and standards-based mathematics teaching*

Erin R. Ottmar is an assistant professor at Worcester Polytechnic Institute, Department of Social Science and Policy Studies, Salisbury Laboratories, 100 Institute Road, Worcester, MA 01609, e-mail: *erin.ottmar@gmail.com*. Her research focuses on the intersection of mathematics teaching and learning, classroom interventions, and social and emotional factors in elementary schools.

Sara E. Rimm-Kaufman is a professor at the University of Virginia in the Curry School of Education and the Center for Advanced Study of Teaching and Learning. She directs the Educational Psychology: Applied Developmental Science program. Her research examines the teaching practices that support elementary school students' social and emotional skill development and engagement in learning.

Ross A. Larsen is an assistant professor of instructional psychology and technology at Brigham Young University. His work focuses on using quantitative methodology to study the effects of school interventions on student achievement.

Robert Q. Berry is an associate professor in the Curry School of Education at the University of Virginia. He focuses on using qualitative metasynthesis as a methodological approach to understand the mathematics experiences of learners. Specifically, his research has focused on understanding the mathematics experiences of Black boys and examining factors that mitigate differentials in mathematical experiences and outcomes.

practices) and student mathematics achievement. Work was conducted in the context of a randomized controlled trial. Participants were 88 third-grade teachers and their 1,533 students. A multigroup path analysis accounting for fidelity of implementation revealed no direct or indirect effects linking MKT to student achievement in the RC *or control condition. The same analysis revealed different findings for the* RC *versus control teachers. In the* RC *group only: (a) Teachers trained in* RC *who used more* RC *practices showed higher use of standards-based mathematics teaching practices, and (b) higher use of standards-based mathematics teaching practices related to greater improvements in math achievement. No comparable findings were evident in the control condition. Results demonstrate the importance of building social and emotional capacity in teachers by helping create a supportive classroom that helps teachers provide stronger mathematics teaching practices that lead to improved student learning.*

KEYWORDS: MKT, standards-based math teaching practices, interventions, achievement

Results from the 2011 National Assessment on Education Progress (NAEP) indicate that only 40% of American fourth-grade students reach math proficiency standards. National reform efforts (e.g., Common Core Initiative, Race to the Top) direct attention toward the effectiveness of teachers and their role in reversing poor mathematics achievement among American students. There is no doubt that teachers play a critical role in students' mathematics learning and achievement (Chetty, Friedman, & Rockoff, 2011; Darling-Hammond, 2002; Hiebert & Grouws, 2007; Nye, Konstantopolous, & Hedges, 2004; Rowan, Correnti, & Miller, 2002). Teachers produce changes in students' achievement through the quality of classroom instructional interactions, including their selection of instructional tasks and opportunities afforded to students (Charalambous, 2010; Kilpatrick, Swafford, & Findell, 2001; Pianta & Hamre, 2009; Stein, Smith, Henningsen, & Silver, 2000). However, teachers and classroom conditions vary; thus, there is a need to explore what occurs *inside* of the classroom during mathematics instruction and understand how certain teacher strengths and classroom contextual factors can improve teacher practice and produce mathematics learning in students.

Teaching mathematics is multifaceted and requires a variety of teacher skills to produce gains in student achievement. Specific to mathematics content, teachers need knowledge of mathematics and the ability to translate that knowledge into effective teaching practices to promote student learning (Hill, Rowan, & Ball, 2005). In addition to mathematics content, teachers need to be capable of organizing their classroom environment and managing student behavior in a way that sets the stage for children's learning (Spillane & Zueli, 1999). It is quite plausible that challenges associated with classroom management and creating a positive social environment may influence teachers' instructional decisions and hinder their ability to teach effectively and impact student learning (Cohen, 1996). Further complicating matters, strategies to facilitate classroom management and effective social interactions among students are notably absent from most mathematics curricula, state or national standards, and professional development (PD) efforts (Scheuell, 1996). This reality turns attention to interventions outside of mathematics and their promise to improve teachers' ability to create classroom conditions conducive to learning.

A recent meta-analysis suggests that social and emotional (SEL) interventions hold promise for contributing to children's classroom behavior (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). SEL interventions are designed to teach and model social and emotional skills that help children interact productively with each other and develop self-management skills necessary for learning, among other skills. Given the focus of most SEL interventions, it is quite possible that SEL interventions may create classroom environments that promote more positive social conditions for learning. In turn, these positive social conditions may alter classroom dynamics in a way that strengthens the inputs and outputs of the instructional system by facilitating more effective mathematics teaching practices. It is plausible that these changes ultimately may lead to improved student achievement. Thus, in this study, we turn attention toward the *Responsive Classroom* (*RC*) approach (Northeast Foundation for Children [NEFC], 2007, 2009), an intervention designed to enhance teachers' ability to create well-managed, positive classroom social environments and build teachers' capacity for effective management of the classroom and responsive teaching and learning. We explore whether the changes in classroom social interactions as a result of *RC* account for improvements in student mathematics achievement and strengthen the relations in the classroom instructional system among mathematics teacher knowledge, instructional practice, and student mathematics achievement.

Mathematical Knowledge for Teaching, Standards-Based Mathematics Teaching Practices, and Student Mathematics Achievement

The recently released *Common Core Standards in Mathematics* (Common Core State Standards Initiative [CCSSI-M], 2010) describes standards for student understanding at each grade level in clear and specific language to create coherence across grade levels. In doing so, CCSSI-M describes content that students are expected to understand at each grade level (e.g., multiply and divide up to 100, represent and solve multiplication and division problems), as well as corresponding mathematical practices that students are expected to demonstrate (e.g., perseverance in solving problems, construction of mathematical arguments, ability to critique reasoning of others). Effective standards-based lessons (resembling those embodied in CCSSI-M) help students develop mathematics thinking, concepts, and skills by promoting cognitive depth (Stein et al., 2000), linking mathematical concepts together in a conceptually coherent manner (Charalambous, 2010; Stein, Grover, & Henningsen, 1996), and emphasizing concept-focused rather than answer-focused instruction (Kilpatrick et al., 2001). However, research suggests that many teachers face difficulty implementing these standards-based mathematics teaching practices (Hiebert & Stigler, 2000; Ottmar, Rimm-Kaufman, Berry, & Larsen, 2013; Spillane & Zeuli, 1999). Many teachers offer little opportunity for discourse, and opportunities for students to explain or justify their thinking rarely occur (Hiebert, 2003). Further, few teachers can maintain a high level of cognitive demand throughout the lesson (Stein et al., 2000). Although teachers sometimes use tools to represent mathematical ideas, often the use of those tools is not aligned with the conceptual purpose of the lesson and students have few opportunities to translate between varied representations (Moyer, 2001).

According to the National Council of Teachers of Mathematics (2000), "Teachers must know and understand deeply the mathematics they are teaching *and* be able to draw on that knowledge with flexibility in their teaching tasks" (p. 17). The construct of mathematical knowledge for teaching (MKT) (Hill, Schilling, & Ball, 2004) builds on Shulman's (1986) model of pedagogical content knowledge (PCK) and encompasses both subject knowledge and pedagogy domains required to teach mathematics effectively (Ball, Thames, & Phelps, 2008). The Mathematical Knowledge for Teaching assessment (Hill et al., 2004) requires teachers to evaluate student understanding of mathematical concepts, interpret unusual student answers, understand nontraditional algorithms, and anticipate student responses that occur commonly in elementary mathematics classrooms.

Confirming the importance of these abilities, studies describe how teachers who can successfully integrate their knowledge of mathematics into instruction are able to teach mathematics concepts with greater depth, be aware of children's thinking and understanding of mathematical concepts, analyze various methods and solutions, and choose appropriate representations and models for instruction (Boaler, 2002; Borko & Putnam 1995; Hill et al., 2004). Studies have found that higher amounts of MKT related positively to the mathematics instruction they provided. Further, teachers with limited mathematical knowledge provided students with lower quality mathematics instruction (Borko & Putnam, 1995; Fennema & Franke, 1992; Hill, Kapitula, & Umland, 2011). In one study, Hill and colleagues (2008) found that teachers who exhibited components of strong mathematics instruction scored higher on the MKT (r = .77). In contrast, teachers with lower MKT were more likely to define concepts incorrectly and chose inappropriate instructional materials and tasks. Similarly, Charalambous (2010) found evidence for positive links between teachers' MKT, the cognitive level of the task, and the enactment of the lesson. A third study found that teachers with higher levels of MKT provided students with more mathematically challenging tasks and presented content with greater mathematical accuracy. In contrast, teachers with lower MKT provided more procedural tasks, misconceptions, and inaccurate instruction (Walkowiak, 2010).

Further, some work has found that teachers' content knowledge about mathematics can positively enhance student achievement (Hill, Ball, Blunk, Goffney, & Rowan, 2007; Hill et al., 2005, 2011). Hill and colleagues found that higher scores on the MKT assessment related to student mathematics achievement gains. Specifically, a one standard deviation increase in MKT resulted in a 0.10 standard deviation gain in achievement (Hill et al., 2005). However, these positive associations between MKT and achievement were not replicated in other samples (Kersting, Givvin, Sotelo, & Stigler, 2010; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Shechtman, Roschelle, Haertel, & Knudsen, 2010), suggesting that what teachers know (mathematical knowledge) and the ways that teachers are able to use their knowledge to provide strong instruction in the classroom may be distinct. These inconsistent findings set the stage for further investigation that validates links between MKT, teachers' use of instructional practices, and mathematics achievement and/or examines potential interventions or explanatory mechanisms that lead to improved teacher practice and student learning.

Recent efforts have strived to improve teacher knowledge, teaching practice, and student mathematics achievement through high-quality, intensive, and ongoing PD and teacher preparation (Darling-Hammond, 1996; U.S. Department of Education, 2010). Recently, Santagata, Kersting, Givven, and Stigler (2011) implemented a randomized-controlled trial of a PD intended to strengthen teacher knowledge, improve mathematics practices, and promote student learning. Although the intervention did not significantly impact teacher knowledge or teacher practices, an interaction effect was found where students in the treatment group who had teachers with higher levels of mathematical knowledge for teaching achieved at higher levels than those with lower knowledge. These results suggest that on its own having higher knowledge may not be sufficient but that other classroom processes may contribute to creating positive change in classroom practice and student achievement (Pianta, Belsky, Houts, & Morrison, 2007; Santagata et al., 2011). The presence of social and emotional factors and other contextual supports (e.g., teacher collaboration, administrative support, and providing materials and examples of high-quality teaching) may promote teacher change and have potential to help teachers direct their knowledge toward improved mathematical teaching practices. This leads us to examine the role of the *Responsive Classroom* (*RC*) approach, a professional development intervention that emphasizes improving the social, emotional, and organizational barriers that are often present in mathematics classrooms.

The *Responsive Classroom* Approach: Strengthening Standards-Based Mathematics Teaching Practices and Student Mathematics Achievement

The *RC* approach is a widely used approach designed to help teachers create a safe, supportive, and well-managed classroom environment to encourage academic learning and support children's social and emotional skills (NEFC, 2007). Unique to the *RC* approach, the intervention does not have a set curriculum and is not a mathematics-specific intervention; the *RC* approach focuses on how teachers interact with their students, manage materials in their classroom, and prevent behavior problems. As a result, the *RC* approach can be used alongside any academic curricular material. *RC* recommends a set of principles and practices for integrating social and academic learning across the school day, creating classroom management processes well aligned with children's social and emotional needs, and fostering a caring and responsive environment for students. For example, in *Morning Meeting*, teachers and children gather as a class, greet each other by name, engage in an activity designed to be fun, and exchange news. Teachers and students collaborate in the beginning of the year to produce classroom rules, and subsequent misbehaviors are addressed in ways that are consistent with the nature of the misbehavior and preserve children's feelings of dignity. By exposing children to *Rules and Logical Consequences*, they become accustomed to a set of positive expectations for their behavior, thus minimizing later behavioral problems and maintaining children's sense of self-respect. In doing so, the *RC* approach may reduce social climate barriers that prevent effective mathematics teaching.

Existing research conducted on the *RC* approach in mathematics classrooms informed our hypothesis that the *RC* approach may help facilitate strong standards-based mathematics teaching practices that may support improved student achievement. A recent study found impacts of training in the *RC* approach, where teachers in schools receiving *RC* training provided stronger standards-based mathematics teaching practices to students (Ottmar et al., 2013). Specifically, *RC* teachers provided standards-based mathematics teaching practices that were .39 standard deviations higher than teachers in the control condition. Further, *RC* teachers were observed spending more time teaching in small groups and more time facilitating higher order analysis and inference (Skibbe, Decker, & Rimm-Kaufman, 2006).

Although it remains unclear *how* the *RC* approach improves standards-based mathematics teaching practices, it is plausible that the intervention provides teachers with the necessary social, organizational, and instructional skills to support more challenging inquiry-based learning. Many of the principles and practices of the *RC* approach are consistent with and support the mathematical processes and requirements outlined in the Common Core standards.

For example, both *RC* and the Common Core standards prioritize improved discourse and communication in the classroom (Chapin, O'Connor, & Anderson, 2009; Sherin, 2002; Yackel, Cobb, & Wood, 1991). The facilitation of a strong mathematical discourse community depends on the establishment of social norms that foster students' freedom of expression of mathematical ideas (Kazemi, 1998). Mathematical discourse arises when teachers serve as the classroom facilitator by orchestrating discussions, asking appropriate questions, and soliciting student ideas to support students' mathematical thinking (Sherin, 2002; Yackel et al., 1991). Classrooms characterized by effective discourse offer ample opportunities for students to "talk mathematics" and share ideas (Fennema, et al., 1996) in large and small groups, writing, and class presentations. Discourse involves explanation of mathematical strategies and students' own justification of their thinking and reasoning (Cobb et al., 1991). Such explanations focus on not only *how* students obtain solutions but *why* strategies are appropriate. Because of the positive social norms that are created in the classroom, an *RC* teacher may be more likely to facilitate frequent open-ended discussions about mathematical concepts and press students for conceptual depth.

Further, standards-based mathematics teaching practices require teachers to create and maintain highly supportive and organized classroom environments in which children feel safe taking risks (Ryan & Patrick, 2001), interact with and form relationships with students, and encourage student autonomy so they can direct their attention to individual students, as needed (Grossman, Schoenfeld, & Lee, 2005). The inclusion of *RC* practices in mathematics classrooms may allow teachers to better match classroom experiences, instruction, and feedback to meet individual students' needs. As a result of these social supports, student readiness to learn mathematics may increase (Borman & Overman, 2004; Hughes & Kwok, 2007). Taken together, the *RC* approach may address many of the social and management challenges inherent in classroom teaching. We posit that training and *use* of the *RC* approach builds teachers' capacity for effective and responsive teaching that may help produce gains in student mathematics achievement.

Fidelity of Implementation: Use of *RC* Practices

A recent randomized-controlled trial examined the three-year impact of the *RC* approach on student reading and mathematics achievement (Rimm-Kaufman et al., 2014). The first finding showed that students at schools assigned randomly to the *RC* condition did not outperform students at schools assigned to the control condition. However, the second finding revealed that fidelity of implementation, operationalized as the extent to which the intervention is used as intended by the developers, played an important role for producing changes in teacher practice and student achievement. A mediation analysis demonstrated that training in *RC* increased the use of *RC* practices in teaching, which in turn related to higher mathematics and reading achievement. More specifically, students in *RC* schools who had teachers who scored one standard deviation higher on use of *RC* practices performed, on average, .44 standard deviations higher on their fifth-grade mathematics assessment. Further, the magnitude of the associations between use of *RC* practices and achievement were .89 standard deviations higher for students in the lowest quartile in math achievement three years earlier. While this work demonstrates that training in the *Responsive Classroom* approach can produce large changes in teacher practices (d = 2.47 to 3.88), the actual use of the *RC* practices varied considerably among schools, leading to treatment heterogeneity (Rimm-Kaufman et al., 2014).

Existing work on the *RC* approach and SEL interventions more broadly also provides additional rationale for measuring fidelity of implementation. For example, Abry, Rimm-Kaufman, Larsen, and Brewer (2013) found that teachers who were trained in *RC* used more *RC* practices in their teaching, and in turn, this higher use of SEL practices related to greater improvement in teacher-student interaction quality. In a second study, higher use of *RC* practices was related to stronger standards-based mathematics teaching practices (Ottmar et al., 2013). Next, quasi-experimental work demonstrated that teachers using more *RC* practices reported closer relationships with their students (Rimm-Kaufman & Chiu, 2007), and use of *RC* practices related to more prosocial behavior, higher achievement, and more positive feelings (by students) about school (Brock, Nashida, Chiong, Grimm, & Rimm-Kaufman, 2008; Rimm-Kaufman & Chiu, 2007; Rimm-Kaufman, Fan, Chiu, & You, 2007). Further, a meta-analysis by Durlak and colleagues (2011) demonstrate that links between SEL and social and academic skills are stronger (or in some cases, only present) when the intervention is actually implemented by teachers. More recent research on specific SEL interventions including RULER, 4Rs, and PATHS also demonstrate the importance of fidelity in relation to outcomes (Brown, Jones, LaRusso, & Aber, 2010; Conduct Problems Prevention Research Group [CPPRG], 1999; Hagelskamp, Brackett, Rivers, & Salovey, 2013; Jones, Brown, & Aber, 2011). Taken together, these findings highlight the need to include measures of fidelity of implementation when considering the role of an intervention on teacher practice and student mathematics learning.

The Present Study

The present study integrates and extends existing research in mathematics education (Hill et al., 2005) and social and emotional learning (Durlak et al., 2011; Rimm-Kaufman et al., 2014) by investigating the mechanisms by which teacher mathematical knowledge and mathematics instructional practice contribute to student mathematics achievement in the presence and absence the *RC* approach. Using classroom observations and direct assessments from a randomized controlled trial of the *RC* approach, the following four research questions were examined separately in intervention and control conditions in third grade:

Research Question 1: To what extent does MKT relate to standards-based mathematics teaching practices and student mathematics achievement?
Research Question 2: To what extent do standards-based mathematics teaching practices relate to student mathematics achievement?
Research Question 3: How do these relations differ in intervention versus control conditions?
Research Question 4: To what extent do the associations vary depending on fidelity of implementation, given the salience of fidelity of implementation in prior work (Abry et al., 2013; Rimm-Kaufman et al., 2014)?

While we are well aware of the varied definitions and measures that exist for identifying effective mathematics instruction, in this study, we conceptualize standards-based mathematics teaching practices as the presence and implementation of the following: tasks that teachers select and the way in which these tasks are enacted in the classroom (cognitive demand, problem solving, and connections and applications), mathematical discourse between teachers and children and among children about mathematics (mathematical discourse community and explanation and justification), mathematical representations and tools used by teachers and students to represent and translate mathematical ideas, and the structure of the mathematics instruction (Berry, Rimm-Kaufman, Ottmar, Walkowiak, & Merritt, 2010; Borko, Stecher, Alonzo, Moncure, & McClam, 2005). We advance the notion that the presence of these practices defined and measured here as standards-based mathematics teaching practices offer one way to conceptualize teacher practice but do not exhaustively measure all practices beneficial to students' mathematical learning.

Method

The present study uses Year 1 data from a three-year longitudinal cluster randomized-controlled trial, the *Responsive Classroom* Efficacy Study (RCES), examining the impact of the *RC* approach on student reading and mathematics achievement (Rimm-Kaufman et al., 2014). Twenty-four schools in a large ethnically and socioeconomically diverse mid-Atlantic suburban school district were enrolled in RCES because of their interest in implementing the *RC* approach and their willingness to participate in a research study. District policy required all elementary schools to select a social and emotional learning approach, and district administrators collaborated with the research team to invite school principals to participate in the study. Participating schools were matched and assigned randomly into intervention (n = 13) and control (n = 11) schools. After randomization, independent t tests and logistic regression were conducted on various demographic and school characteristics that have been shown to relate to student achievement in past educational research. Tests were conducted to ensure that the groups were comparable at baseline (on gender, free and reduced priced lunch [FRPL], ethnic composition, and English language learner [ELL] status). FRPL and ethnic composition were comparable across intervention (27.63% FRPL, 59% ethnic minority) and control schools (24.53% FRPL, 53% ethnic minority).

Participants

Teachers

All third-grade teachers who taught mathematics at the 24 schools during the 2008–2009 school year were invited to participate in this study (N = 100); 94 teachers consented (94% response rate). Six third-grade teachers were excluded because they taught mathematics in a foreign language (Spanish or French). Thus, the final sample was comprised of 88 third-grade teachers from 24 schools (83 female, 5 male; 83% White, 6.8% African American, 1.1% Hispanic, and 9.1% of another ethnicity). Teachers had, on average, 9.27 years of teaching experience (range, 1–35 years, SD = 8.10). Forty-one percent of the teachers had a bachelor's degree only, 54% held a master's or other graduate degree, and 5% did not report their education level. Half (n = 44; 50%) of the teachers taught in schools receiving the training in the *RC* approach. Teachers in the present sample were, on average, comparable to national samples in terms of ethnicity and education level but were slightly less experienced (Goldring, Gray, & Bitterman, 2013).

Children

All-third grade children enrolled in the 24 schools (n = 2,464) participated in the larger three-year longitudinal RCES study (Rimm-Kaufman et al., 2014). Of these students, 1,533 were considered eligible for the current study. Students were considered eligible if they were enrolled in the classroom of 1 of the 88 mathematics teacher participants and took the state standardized assessment in mathematics in the spring of 2009. On average, students in the present sample were slightly more ethnically diverse than national samples but comparable in terms of FRPL, individualized education plan (IEP), and ELL (National Center for Education Statistics [NCES], 2011). About half (50.8%) of the students were male (n = 779). In terms of ethnicity, 636 (41.5%) were Caucasian, 165 (10.8%) were African American, 338 (22%) were Hispanic, 285 (18.6%) were Asian, and 6 (0.4%) were of other ethnicities. One hundred and three of the students (6.7%) did not report their ethnicity. Further, 208 (13.5%) of the students had an IEP, 661 (43.1%) were ELL, 497 (32.4%) of the students received FRPL, and 742 (48.4%) of the children attended schools assigned to the treatment condition (and thus had teachers who were trained in the *RC* approach).

Procedures

RC Training

During the summer of 2008, the teachers assigned to the *RC* condition completed a one-week long *RC* 1 training institute consisting of 35 hours of instruction, where they were introduced to and taught *RC* practices. Trained consultants from the Northeast Foundation for Children provided training to teachers in groups of 20. During the 2008–2009 school year, teachers received three consultations and classroom visits by their *RC* coach, attended one-day workshops, and had access to e-mail and phone communication. During coaching visits, teachers were observed and given feedback about their use of *RC* practices. In addition, each coach conducted a lesson in the teachers' classroom, held debriefing sessions and mini-workshops, and led meetings with teachers and administrators. The 44 control teachers received no training on *RC* practices and continued their instruction using "business as usual" approaches. The schools reported using a combination of textbooks (i.e., *Everyday Mathematics*, Silver Burdett, Scott-Foresman) and district testing frameworks and pacing guides as tools to direct mathematics instruction.

Data Collection

Data were collected from three sources: school record data, online teacher-report questionnaires, and classroom observations conducted and coded by research assistants. Student demographic information was gathered at baseline in May 2008. Students' mathematics achievement was assessed during both second and third grade in May 2008 and May 2009, respectively. In spring 2009, teachers completed an online demographics questionnaire and an online assessment of their mathematical knowledge for teaching. Questionnaires and MKT assessment batteries required roughly one hour of teachers' time, and teachers received $100 for study participation. Research assistants videotaped and conducted fidelity of implementation observations (described in the following) on all third-grade teachers in both control and *RC* groups for a full mathematics lesson (usually about 60 minutes) at three time points during the 2008–2009 school year. Three 3-month windows of observation were established to ensure representative sampling at different time points during the year: fall (September to November), winter (December to February), and spring (March to May). Upon the completion of videotaping, tapes were then sent to the laboratory for off-site observational coding.

Coder Training and Interrater Reliability for Observational Measures

All classroom observations were coded for use of standards-based mathematics practices and use of *RC* practices by trained research assistants using the Mathematics Scan (M-Scan) measure (Berry et al., 2010) and the Classroom Practices Observational Measures (CPOM) measure (Abry, Brewer, Nathanson, Sawyer, & Rimm-Kaufman, 2010), described in the following. Several steps were followed to ensure that the coders were properly trained and achieved high levels of reliability. First, coders worked independently to learn about the respective measure and protocol. Each coder watched three videos, took notes, considered scores they would assign for each item, and then reviewed the master codes justifications for the videos. Coders then met with a master coder to review the readings, the coding protocol, and the coding form. Together, they watched video clips that exemplified different scores on each of the dimensions and practiced coding at least two one-hour segments. To establish interreliability, coders independently coded six one-hour classroom observations without conferring with the master coder. After coding the eight observations, they met with the master coder to confer on scores. The master coder computed reliability scores (at least 80% exact match was required), identified items that were sources of systematic error, and looked at convergence of ratings. If a coder did not meet the 80% exact match, the coder was required to code an additional two tapes. After the two additional tapes were coded and added to the set, the exact match percentage was recalculated. All coders became reliable before their codes were used as data. Every month, the master coder conducted a meeting in which coders viewed, coded, and discussed one or two one-hour segments. These meetings were used to assess coders' drift from master codes, and all scores were tracked to determine coders' ongoing status.

Measures

Student Demographic Information

Demographic information about students was collected directly from the school administrators. Information was included in the state standards of learning database and verified by the State Department of Education. Ethnicity/minority status, gender, free-reduced priced lunch status, and English language learner status (ELL) were used as covariates at the child level.

Second-Grade Achievement

Second-grade mathematics achievement was measured using the *Stanford Achievement Test–10th Edition* (Harcourt Educational Measurement, 2002). The assessment was administered to all children in participating third-grade classrooms (n = 1,533) in April of their second grade. A standardized score was generated for each child by Harcourt (range, 1–999) and was used as an indicator of baseline mathematics achievement prior to entering third grade. It was also used as a student-level covariate to isolate the effects of teacher knowledge and instructional practice on student achievement.

Third-Grade Mathematics Achievement

The third-grade mathematics state standardized assessment was administered to all third-grade students in May 2009. Tests were administered by district-level testing professionals and classroom teachers using the standard district protocols. The test was composed of 50 items, covering five content areas (number and number sense, 13 items; computation and estimation, 11 items; measurement and geometry, 12 items; probability and statistics, 7 items; and patterns, functions, and algebra, 7 items). The number of items correct was calculated and converted into a scale score (range, 0–600). The pass cutoff score for this assessment was 404 (37 items correct).

Teacher Demographic Measure

Teachers completed a demographic questionnaire, providing information about their age, ethnicity, years teaching experience, and educational training.

Mathematical Knowledge for Teaching Assessment

Teachers' mathematical knowledge for teaching was assessed using the Mathematical Knowledge for Teaching Assessment (Hill et al., 2004). The MKT is a multiple-choice assessment requiring teachers to evaluate multiple solution methods, represent mathematical concepts and ideas, and identify appropriate mathematical explanations. Teachers completed a 13-item assessment related to Number and Operations (K–6). The MKT scores are reported using item response theory (IRT) scales for the number and operations items, ranging from–2 to 2. Each teacher's score represented the number of standard deviation units he or she scored from the larger national sample mean. Reliability values ranging from .71 to .84 were comparable to other reported samples (Hill et al., 2004).

Mathematics Scan

Mathematics teaching practices in the classroom were assessed using the M-Scan (Berry et al., 2010). The M-Scan is an observational measure designed to assess the enactment of standards-based mathematics teaching practices. This instrument was developed by our research team from 2008 to 2010 to assess dimensions of mathematics teaching practices recommended by National Council of Teachers of Mathematics (2000), CCSSI-M, and prior research in mathematics education. Measurement dimensions were based on and adapted from observational work by Borko et al. (2005). The

Classroom Assessment Scoring System (CLASS) (Pianta, LaParo, & Hamre, 2008) served as a framework for designing the dimensions and rating scales. The CLASS is an observational measure that is designed to measure teacher-child instructional, emotional, and organizational interactions in the classroom. The CLASS is not mathematics specific but can be used in any elementary classroom content or setting.

The M-Scan measure includes eight dimensions of mathematics teaching practices, each coded on a Likert scale of 1 to 7, representing *low* (1, 2), *mid* (3, 4, 5), and *high* (6, 7) enactment of these standards-based practices. The eight dimensions include: structure of the lesson, multiple representations, students' use of mathematical tools, cognitive depth, mathematical discourse community, explanation and justification, problem solving, and connections and applications. A multidimensional coding guide provides key word descriptors and anchors for each dimension, characterizing low, medium, and high enactment of mathematics teaching practices (see Appendix in the online journal for a description of each dimension).

Evidence of score reliability of the M-Scan instrument was established using four methods in prior work (Walkowiak, Berry, Meyer, Rimm-Kaufman, & Ottmar, 2014). First, content experts in mathematics and mathematics education reviewed the M-Scan dimensions and indicators and confirmed that they represented important mathematics teaching practices. Second, a qualitative review of coders' responses indicated that 87.7% of the coders' rationales for assigning scores aligned with the M-Scan descriptors. Third, patterns of convergence and divergence were found with existing measures of mathematics instruction. For example, analyses showed moderate to strong correlations (.46–.79) with the 11 items that loaded onto inquiry-based instruction measured by the Reformed Teaching Observation Protocol (RTOP) (Sawada & Piburn, 2002). The RTOP is an observational instrument designed to measure the extent to which instruction is consistent with reformed mathematics and science teaching. The M-Scan also showed moderate relations (.22–.44) with the instructional support items on the CLASS (Pianta et al., 2008).

A G-Study estimating variance components for each source of variance (i.e., coder, classroom, and items) and a D-Study computing the measurement error variance and reliability coefficients support the reliability of the measure (Walkowiak et al., 2014). Findings indicated the largest portion of variance was at the classroom level, suggesting that the M-Scan differentiated among the mathematical teaching practices in different classrooms. Individual raters accounted for only 5% of the variance, and reliability coefficients indicated evidence of interrater reliability (generalizability coefficients ranged from .74 [one coder] to .85 [four coders]). Additional reliability analyses were conducted to ensure the interrater reliability of coders (all intraclass correlation coefficients [ICCs] $>.76$; weighted kappa, .59; 88% exact agreement; and correlations ranging from .77 to .85).

Every mathematics lesson was observed by a trained research assistant who was blind to teachers' intervention status and unfamiliar with the *RC* approach. Coders watched the first 30 minutes of the tape, took anecdotal notes, and wrote down example problems throughout the 30-minute segment, to record what occurred during the lesson. After 30 minutes, coders reflected on what they saw and assigned preliminary codes to rate the first part of the lesson. Next, coders watched the remainder of the tape, following the same procedures as the first 30-minute segment. Upon completion of the lesson, coders reviewed their notes and assigned final codes from 1 to 7 for each dimension on the M-Scan measures. Coders watched three lessons from each teacher, and no coder watched more than two lessons taught by the same teacher.

Exploratory and confirmatory factor analyses indicated that all eight M-Scan dimensions loaded onto one latent construct (standards-based mathematics teaching practices). Thus, a composite score of overall standards-based mathematics teaching practices was created for each teacher. First, a mean score for each dimension was calculated by averaging the scores for each dimension over the three sampled days. Next, these composite means for each of the eight dimensions were averaged, resulting in an M-Scan value for each teacher. Potential values of M-Scan scores could range from 1 (*low mathematics teaching practices*) to 7 (*high mathematics teaching practices*). Measures of internal consistency were strong (Cronbach alphas ranged from .90 to .93).

Classroom Practices Observational Measure

The CPOM (Abry et al., 2010) was designed to measure a teacher's observed level of fidelity of implementation to *RC* practices (in the *RC* group) and the use of *RC*-like practices in the control group. The CPOM measures observed fidelity of implementation to the *RC* approach by assessing 10 classroom practices, including routines, rules and procedures, and the organization of physical classroom space. The measure uses a 3-point rating scale (1 = *not at all characteristic*, 2 = *moderately characteristic*, 3 = *very characteristic*). Each item was written to ensure that the measure tapped teachers' use of specific *RC* classroom practices without using *RC*-specific language. Trained coders present in the classroom watched and took notes during each mathematics lesson (60 minutes). At the end of the lesson, coders assigned codes for each item, using the criteria in the CPOM coding protocol. The internal consistency of the measure was strong (Cronbach alpha = .84).

Although fidelity of implementation is usually thought of as a construct measured only in the treatment condition, many of the *RC* principles are considered "best practices" in education, and thus, it is possible that some teachers in the control group were using some of these practices (Abry et al., 2013). Therefore, the decision was made to include CPOM to account for teachers in the control group who were using strategies similar to *RC* practices in their everyday lessons. Measuring of use of *RC* practices in both the control and intervention group allowed us to estimate the effect of social and emotional processes on teacher practice. In the intervention group, fidelity of implementation assesses variability in use of *RC* practices, whereas in the control group, fidelity of implementation measures the degree to which teachers in the control group may be using practices so closely resembling *RC* practices that they are likely to have comparable effects. For the remainder of the article, we will refer to fidelity of implementation as the use of *RC* practices both in relation to the intervention and control conditions.

Approach to Data Analysis

First, three sets of descriptive statistics (means and standard deviations) and bivariate correlations were calculated to determine the variation of and relation between variables for the whole sample as well as the *RC* and control groups separately. Independent *t* tests, using the Bonferroni method to account for multiple comparisons ($p < .05/5$ or .01) were used to examine whether there were differences between any of the variables of interest in the *RC* and control group at baseline (second-grade mathematics achievement) and during the third-grade year (MKT, use of *RC* practices, standards-based mathematics teaching practices, third-grade mathematics achievement). All data were checked for normality of distributions, homoscedacity, and outliers (Tabachnick & Fidell, 2001).

The data were analyzed using a series of multilevel, multigroup path analyses to test the relations between variables and determine whether there were differences in the relations between variables across the *RC* and control samples. This was done using Mplus 7.11 (Muthén & Muthén, 2010), with type = TWOLEVEL to control for the clustering of students within classrooms. Preliminary examination of the data suggested that a hierarchical two-level model (students nested in classrooms) was most appropriate. A three-level model was not necessary as the design effects were less than 2 when clustering at the school level (Muthén & Satorra, 1995), and the ICC values at the school level were less than .02 after controlling for baseline achievement. Intraclass correlation coefficients at the classroom level for the outcome of mathematics achievement were .29 for the whole group, .30 for the *RC* group, and .24 for the control group. For the *RC* group, 69.7% of the total variance was at the child level, and 30.3% was at the teacher level. For the control group, 75.6% of the variance in student mathematics achievement was at the child level, and 24.4% was at the teacher level.

Multigroup analysis was conducted with the groups being determined by treatment status, and the invariance of the groups was then tested (as explained in the following). Full information maximum likelihood (FIML) and robust maximum likelihood (MLR) estimation procedures in Mplus were used to estimate the models and account for data that were determined to be missing at random. In all models, Level 1 controlled for student-level characteristics that have been found in prior research to relate to student achievement, including free and reduced priced lunch, English language learner and minority status, and second-grade mathematics achievement. All Level 1 dichotomous variables (FRPL, minority status, ELL) were uncentered, and second-grade achievement was grand mean centered. All Level 2 variables (MKT, M-Scan, and CPOM) were grand mean centered.

To establish a baseline model, an initial fully constrained model between treatment and control groups was tested. Next, a series of models were conducted, releasing the constraints one path at a time. Model fit indices for each model were evaluated by calculating the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and chi-square (χ^2). Good model fit was indicated by the following criterion: $\chi^2/df < 8.0$, RMSEA less than .06 (Hu & Bentler, 1999), CFI and TLI greater than .90 (Bakker, Hakanen, Demerouti, & Xanthopoulu, 2007; Bentler, 1990), and SRMR (within and between) <.05 (Sivo, Fan, Witta, & Willse, 2006). After each path was released, the model was compared to the previous model by a chi-square difference test with a Satorra-Bentler correction for using the MLR estimator. If the chi-square difference test demonstrated model fit improvement, then the less constrained model was accepted. This was repeated until the best fit was found. Model 1 functioned as the baseline model and initial source for comparison and tested the direct and indirect relations between MKT, standards-based mathematics teaching practice, and student mathematics achievement in the whole sample. In this model, we fully constrained the paths to be equal across groups so that the estimates for each path were the same for both the control and intervention groups. In order to test the effect of fidelity of implementation, we also constrained the paths to and from fidelity to be 0.

Next, to estimate the effects for intervention and control samples separately, we tested the same relations as Model 1 but allowed the loadings to be unequal across groups (with the exception of the paths to and from fidelity, which were still constrained to be 0). Model 2 allowed for: (a) simultaneous testing of the direct and indirect effects of each variable for predicting teacher practice and student achievement for each group separately and (b) a comparison of path coefficients across the two groups of teachers (*RC* and control) (Park & Huebner, 2005). It also allowed us to determine the effect of the *RC* approach on the overall instructional system. In order to test whether the relations were invariant across groups, a chi-square difference test with a Satorra-Bentler correction for using the MLR estimator was run between the fully constrained model (Model 1) and the model where the paths were allowed to vary freely across groups (Model 2).

Third, to examine the relations between MKT, use of *RC* practices, use of standards-based mathematics teaching practices, and student mathematics achievement in the whole group, the paths to and from fidelity of implementation (use of *RC* practices) at the teacher level were freely estimated (Model 3). This continuation of model building and the inclusion of fidelity of implementation as a key variable allowed us to test empirically driven relations from prior work on the *RC* approach (Abry et al., 2013; Rimm-Kaufman et al., 2014). A chi-square difference test with a Satorra-Bentler correction for using the MLR estimator was then run to compare Models 1 and 3 to justify the inclusion of the fidelity variable in the model.

Next, a fourth model was conducted, testing the same relations as Model 3 but allowing the loadings to be unequal across groups. Finally, a test of model invariance using a chi-square difference test with a Satorra-Bentler correction for using the MLR estimator was conducted to determine if a multigroup model accounting for use of *RC* practices (Model 4) was a better fit to the data than a whole group model accounting for use of *RC* practices (Model 3). Our four research questions were then examined and interpreted using the best fitting model.

Results

Descriptive Statistics and Correlations

Descriptive statistics and bivariate correlations for all variables of interest for the *RC* and control group as well as the full sample are presented in Table 1. Independent t tests using the Bonferroni correction ($p < .01$) indicated that there were differences on some variables between the control and *RC* group in the third-grade year. First, use of standards-based mathematics teaching practices was higher in *RC* classrooms than control classrooms ($p < .01$). Next, MKT was lower in *RC* classrooms than control classrooms ($p < .01$). Third, the use of *RC* practices was higher in *RC* classrooms than control classrooms, indicating that *RC* teachers used many of the practices that were taught in the *RC* training in their mathematics teaching ($p < .01$).

In both groups, higher scores on the MKT assessment were positively related to the use of standards-based mathematics teaching practices. MKT scores were not correlated to student achievement scores in third grade in either condition; however, group differences were found with regards to mathematics teaching practices and student achievement: M-Scan scores were positively related to third-grade mathematics achievement in the *RC* intervention group but were not significantly related to third-grade mathematics achievement in the control group.

Table 1

Correlations and Descriptive Statistics for All Variables

	1	2	3	4	5	6	7	8
1. FRPL (1 = FRPL)	—	*0.52***	*0.50***	*–0.36***	*–0.02*	*0.17***	*0.18***	*–0.33***
	—							
2. ELL (1 = ELL)	**0.53****	—	*0.35***	*–0.36***	*–0.04*	*0.10***	*0.10***	*–0.27***
	0.50**	—						
3. Minority status (1 = minority)	**0.46****	**0.33****	—	*–0.35***	*–0.01*	*0.13***	*0.15***	*–0.29***
	0.52**	0.36**	—					
4. Second-grade math achievement	**–0.33****	**–0.37****	**–0.33****	—	*–0.00*	*0.01*	*–0.05**	*0.71***
	–0.38**	–0.34**	–0.36**	—				
5. Mathematical knowledge for teaching (MKT)	**–0.06**	**–0.04**	**–0.08**	**0.00**	—	*0.21***	*0.16***	*0.02*
	0.04	0.03	0.06	–0.02	—			
6. Mathematical teaching practices (M-Scan)	**0.11****	**0.09***	**0.03**	**0.16****	**0.22****	—	*0.28***	*0.02*
	0.19**	0.10**	0.20**	–0.13**	0.24**	—		
7. Use of *RC* practices (CPOM)	**0.21****	**0.16****	**0.18****	**–0.04**	**0.05**	**0.23****	—	*–0.02***
	0.04	0.01	0.05	0.03	0.41**	0.22**	—	
8. Third-grade math achievement	**–0.37****	**–0.32****	**–0.30****	**0.76****	**0.01**	**0.13****	**–0.09***	—
	–0.25**	–0.21**	–0.26**	0.65**	–0.01	–0.05	0.04	—
RC mean	**0.39**	**0.46**	**0.45**	**574.24**	**–0.19**	**3.38**	**1.52**	**466.39**
RC SD	**0.49**	**0.50**	**0.50**	**45.12**	**0.70**	**0.90**	**0.23**	**79.63**
Control mean	0.26	0.40	0.35	584.31	–0.02	3.10	1.34	490.89
Control *SD*	0.44	0.49	0.48	38.88	0.79	0.84	0.19	68.35
Full sample mean	*0.32*	*0.43*	*0.40*	*579.44*	*–0.10*	*3.24*	*1.43*	*479.08*
Full sample *SD*	*0.47*	*0.50*	*0.49*	*42.30*	*0.75*	*0.88*	*0.23*	*74.99*

Note. For the top part of the table showing correlation coefficients, results for the *RC* group are listed first in bold, and the results for the control group are listed second in regular type. Correlation coefficients for the full sample (both *RC* and control) are presented on the top right diagonal in italics. FRPL = free and reduced price lunch; ELL = English language learner; M-Scan = Mathematics Scan; *RC* = *Responsive Classroom*; CPOM = Classroom Practices Observational Measures.

$^{*}p < .05$. $^{**}p < .01$.

Table 2
Model Fit Indices for Models 1 Through 4

Model Fit Indices	Model 1	Model 2	Model 3	Model 4
CFI	0.99	0.99	0.99	1.00
TLI	0.99	0.97	0.98	1.00
RMSEA	0.04	0.06	0.04	0.05
SRMR (within)	0.03	0.02	0.02	0.02
SRMR (between)	0.15	0.15	0.07	0.00
χ^2 (*df*)	62.85 (26)	33.94 (10)	51.36 (23)	17.05 (4)
p	<.01	<.01	<.01	<.01
Change χ^2 (*df*)[a]		28.73 (16)	11.49 (3)	32.76 (19)

Note. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.
[a]Values show the Satorra-Bentler correction.

Model Comparisons

A series of four models were conducted to test the relations between variables and determine the most parsimonious model for the data. Model fit indices for all four models are reported in Table 2. Standardized and unstandardized path coefficient estimates, standard errors, and percentage variance explained for Models 1 and 2 and Models 3 and 4 are presented in Tables 3 and 4, respectively.

The first step was to build a baseline one-group model (Model 1) that estimated the relations between MKT, standards-based mathematics teaching practices, and student achievement but did not account for the use of *RC* practices. Model 1 had acceptable fit (χ^2 = 62.85, *df* = 26, $p < .01$, RMSEA = .04, CFI = .99, TLI = .99, SRMR within = .03, SRMR between = .15). In Model 1, MKT was positively related to standards-based mathematics teaching practices: Teachers who scored one standard deviation higher on the MKT assessment implemented .30 standard deviations more standards-based mathematics teaching practices ($p < .05$). Further, students who had teachers who implemented stronger mathematics teaching practices performed at higher levels on the mathematics assessment (effect size = .21, $p < .05$). No direct or indirect relations were found between MKT and student learning. Model 1 accounted for 4% of the variance in student achievement and 9% of mathematics teaching practices for the teacher level.

The second model tested the same relations as Model 1 but allowed the paths to vary by treatment group. Model 2 also had adequate fit (χ^2 = 33.94, *df* = 10, $p < .01$, RMSEA = .06, CFI = .99, TLI = .97, SRMR within = .02, SRMR between = .15). The results of the chi-square difference test with a Satorra-Bentler correction for using the MLR estimator favored Model 2 over Model 1 ($\Delta\chi^2$ = 28.73, Δdf = 16, p = .026), indicating that the groups were not

Table 3

Standardized and Unstandardized Path Coefficients for Models 1 and 2

	Model 1			Model 2					
	One Group, Fidelity Relations Constrained to 0			Multigroup, Fidelity Relations Constrained to 0					
				Control			*RC* Intervention		
Level 1 Predictors	Standardized *B*	Standardized *SE*	β	Standardized *B*	Standardized *SE*	β	Standardized *B*	Standardized *SE*	β
FRPL	–0.09**	0.02	–14.14**	–0.05	0.03	–6.60	–0.14**	0.03	–21.70**
ELL	0.02	0.02	3.09	0.05	0.03	6.71	0.00	0.04	0.29
Minority	–0.04	0.02	–5.29	–0.09**	0.03	–11.85**	0.00	0.02	0.18
Second-grade achievement	0.69**	0.02	1.16**	0.67**	0.03	1.08**	0.70**	0.03	1.23**
Level 2 predictors									
MKT to M-Scan	0.30*	0.14	0.36*	0.27	0.19	0.30	0.34	0.18	0.44
Predictors of third-grade achievement									
MKT to achievement	–0.01	0.11	–0.11	–0.01	0.15	–0.21	0.00	0.16	0.08
M-Scan to achievement	0.21*	0.10	4.38*	0.16	0.13	3.92	0.26	0.17	4.23
Indirect effects									
MKT to M-Scan to achievement	0.06	0.05	1.57	0.04	0.05	1.17	0.09	0.08	1.86
ICC third-grade achievement		0.29			0.30			0.24	
Level 1 % variance explained (achievement)		53			50			56	
Level 2 % variance explained (achievement)		4			3			7	
Level 2 % variance explained (M-Scan)		9			7			12	

Note. *RC* = *Responsive Classroom*; FRPL = free and reduced price lunch; ELL = English language learner; MKT = mathematical knowledge for teaching; M-Scan = Mathematics Scan; ICC = intraclass correlation coefficients.
*$p < .05$. **$p < .01$.

Table 4

Standardized and Unstandardized Path Coefficients for Models 3 and 4

	Model 3			Model 4					
	One Group, Fidelity Relations Estimated			Multigroup, Fidelity Relations Estimated					
				Control			*RC* Intervention		
Level 1 Predictors	Standardized *B*	Standardized *SE*	β	Standardized *B*	Standardized *SE*	β	Standardized *B*	Standardized *SE*	β
FRPL	–0.09**	0.02	–13.94**	–0.05	0.03	–6.59	–0.14**	0.02	–21.21**
ELL	0.02	0.02	3.17	0.05	0.03	6.71	0.00	0.04	0.65
Minority	–0.04	0.02	–5.11	–0.09**	0.03	–11.84**	0.00	0.03	0.70
Second-grade achievement	0.69**	0.02	1.16**	0.67**	0.03	1.08**	0.69**	0.02	1.23**
Level 2 predictors									
MKT to CPOM	0.24*	0.10	0.07*	0.39**	0.13	0.11**	0.06	0.17	0.02
MKT to M-Scan	0.24	0.14	0.28	0.14	0.22	0.16	0.32	0.17	0.42
CPOM to M-Scan	0.26*	0.10	1.00*	0.30	0.19	1.91	0.26*	0.13	0.99*
Predictors of third-grade achievement									
MKT to achievement	0.02	0.12	0.61	–0.01	0.22	–0.15	0.00	0.16	0.08
M-Scan to achievement	0.25*	0.10	5.23*	0.16	0.11	3.95	0.34*	0.16	5.67*
CPOM to achievement	–0.18	0.13	–15.17	0.00	0.19	–0.59	–0.33	0.18	–21.45
Indirect effects									
MKT to CPOM to M-Scan	0.06	0.04	0.07	0.12	0.09	0.13	0.02	0.04	0.02
MKT to M-Scan to achievement	0.06	0.05	1.48	0.02	0.05	0.64	0.11	0.08	2.38
MKT to CPOM to M-Scan to achievement	0.02	0.01	0.38	0.02	0.02	0.52	0.01	0.01	0.11
CPOM to M-Scan to achievement	0.06	0.04	5.22	0.05	0.05	4.70	0.09	0.06	5.62
Total effect	0.30*	0.15	0.36*	0.26	0.19	0.30	0.34	0.18	0.42
ICC third-grade achievement		0.29			0.30			0.24	
Level 1 % variance explained (achievement)		53			50			56	
Level 2 % variance explained (achievement)		7			3			17	
Level 2 % variance explained (M-Scan)		15			14			18	
Level 2 % variance explained (CPOM)		6			16			0	

Note. RC = *Responsive Classroom*; FRPL = free and reduced price lunch; ELL = English language learner; MKT = mathematical knowledge for teaching; CPOM = Classroom Practices Observational Measures; M-Scan = Mathematics Scan; ICC = intraclass correlation coefficients.

$^{*}p < .05$. $^{**}p < .01$.

invariant, and it was necessary to use a multigroup model allowing the paths to vary across *RC* and control groups. Model 2, estimating the paths separately for control and intervention samples, revealed no significant direct or indirect relations between mathematical knowledge for teaching, standards-based mathematics teaching practices, or student achievement in either group (all ps $> .05$). Model 2 accounted for 3% and 7% of the variance in student achievement and 7% and 12% of mathematics teaching practices at the teacher level for the control and intervention groups, respectively.

The next step in building the model was to freely estimate the paths relating to fidelity of implementation, specifically, from MKT to use of *RC* practices and use of *RC* practices to M-Scan, while constraining the paths across groups to be equal (Model 3). Model 3 had acceptable fit (χ^2 = 51.36, df = 23, $p < .01$, RMSEA = .04, CFI = .99, TLI = .98, SRMR within = .02, SRMR between = .15). Several indicators supported that Model 3, the one-group model that included a fidelity variable to account for the use of *RC* practices, was preferred over Model 1. First, the Satorra-Bentler corrected chi-square test favored Model 3 ($\Delta\chi^2$ = 11.49, Δdf = 3, $p < .01$). Next, both the Akaike Information Criterion (AIC) and the sample-size adjusted Bayesian Information Criterion (BIC) were lower in Model 3 than Model 1 (Model 1: AIC = 37,282.66, BIC = 37,496.06; Model 3: AIC = 37,276.65, BIC = 37,369.46). Further, the addition of fidelity of implementation to the model explained an additional 6% of variance in standards-based mathematics teaching practices and 3% of Level 2 variance in student achievement. Additionally, the associations between fidelity of implementation and M-Scan were statistically significant in both models, indicating that leaving that term out would be underfitting. In Model 3, after accounting for the use of *RC* practices, MKT no longer significantly predicted standard-based mathematics teaching practices ($p > .05$). However, higher MKT was related to greater use of *RC* practices (effect size = .24, $p < .05$) and accounted for 6% of the variance in use of *RC* practices. Next, teachers who used more *RC* practices used more standards-based mathematics teaching practices (effect size = .26, $p < .01$). Further, greater use of standards-based mathematics teaching practices was related positively to student achievement ($p < .01$). In other words, students whose teachers scored one standard deviation higher on the M-Scan performed, on average, one-fourth of a standard deviation higher on their mathematics achievement assessment. No significant indirect effects were found linking MKT to student achievement ($p > .05$). However, the total effect was significant (.30, $p < .05$). For every one standard deviation increase on MKT, student achievement increased (on average) by .30 standard deviations through both the use of *RC* practices and mathematics teaching practices pathways.

The fourth model tested a multigroup model that accounted for fidelity of implementation (use of *RC* practices) and estimated *RC* and control samples separately. Model 4 had strong fit (χ^2 = 17.05, df = 4, $p < .01$, RMSEA =

.05, CFI = 1.00, TLI = 1.00, SRMR within = 0.02, SRMR between = .00) and outperformed the other three models. Results from the chi-square difference test with a Satorra-Bentler correction ($\Delta\chi^2 = 34.31$, $\Delta df = 19$, $p = .025$) confirmed that the *RC* and control groups were not invariant, and it was necessary to allow the paths to vary across groups. Based on these results, Model 4 was determined to represent the data the best and was used to interpret the study findings.

Relations Between MKT, Use of *RC* Practices, Standards-Based Teacher Practice, and Student Learning

Results from Model 4 show significantly different patterns of relations between MKT, use of *RC* practices, standards-based mathematics teaching practices, and student mathematics achievement across the two groups. Figure 1 contains the final model with all significant standardized path coefficient estimates and explained variances for both groups.

In the control group, MKT did not relate to mathematics teaching practices or student mathematics achievement. However, teachers who scored one standard deviation higher on MKT used .39 standard deviations more *RC* practices in their teaching ($p < .01$). No direct or indirect relations between use of *RC* practices and standards-based mathematics teaching practices or mathematics teaching practices and student mathematics achievement were found.

MKT did not significantly predict use of *RC* practices, standards-based mathematics teaching practices, or student achievement in the intervention group ($p > .05$). However, *RC* teachers who used *RC* practices with greater fidelity implemented stronger standards-based mathematics teaching practices (effect size = .26, $p < .05$). Further, greater use of standards-based mathematics teaching practices was positively related to achievement in the *RC* group. For every one standard deviation increase in standards-based mathematics teaching practices, students performed .34 standard deviations higher on the mathematics assessment. No significant indirect effects were found linking MKT to student achievement.

Model 4 accounted for the largest amount of variance in teacher practice and mathematics achievement. In the *RC* group, the model explained 56% of the child-level variance and 17% of the teacher-level variance in student mathematics achievement. However, for the control group, this model explained 50% of the child-level variance and only 3% of the teacher-level variance in student mathematics achievement. MKT explained 16% of the variance in the use of these social and emotional practices for control teachers but 0% of the variance in the use of these practices for those teachers trained in the *RC* intervention. Further, the model explained 14% and 18% of the variance in standards-based mathematics teaching practices in the control and *RC* conditions, respectively.

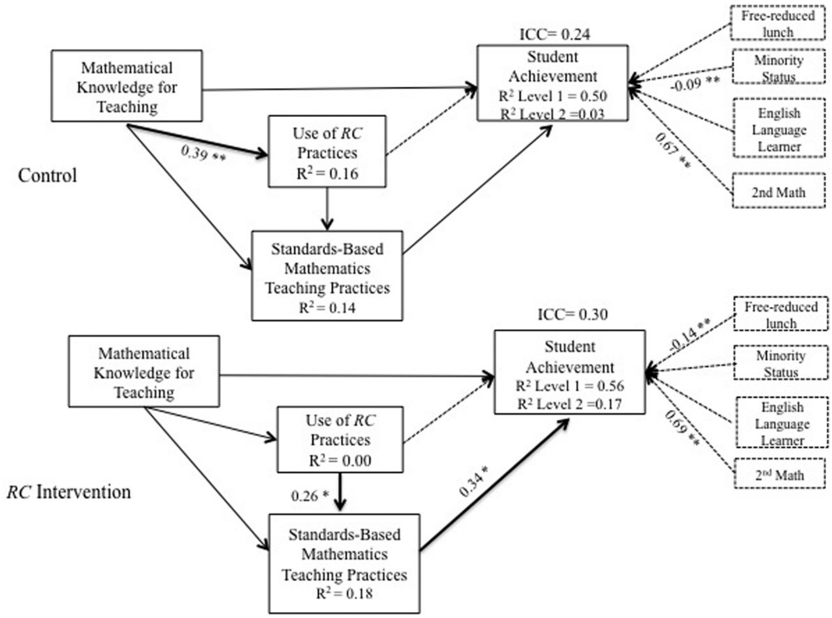

Figure 1. **Significant standardized path coefficients for Model 4, the multigroup model comparing control versus *RC* with fidelity (use of *RC* practices). Dashed lines represent covariates. ICC = intraclass correlation coefficients; *RC* = *Responsive Classroom*.**

$*p < .05$. $**p < .01$.

Discussion

The goals of the present study were to examine the direct and indirect relations between mathematical knowledge for teaching, standards-based mathematics teaching practices, and student mathematics achievement and consider the effect of the *RC* intervention for changing these relations. Multigroup path analysis results (from Model 4) indicate different patterns of relations in the *RC* and control conditions. No direct or indirect effects were found from MKT to standards-based mathematics teaching practices or student achievement in either group. However, a significant positive relation from standards-based mathematics teaching practices to student achievement was found in the *RC* group after accounting for use of *RC* practices. Further, direct effects linking higher use of *RC* practices to stronger mathematics teaching practices in the *RC* condition confirm the importance of actual use of *RC* practices, above and beyond training in the intervention.

Taken together, these findings raise questions about the ways in which the *RC* approach sets the stage for improving classroom practice and student mathematics achievement. The findings can be viewed from two different vantage points. The first considers the relations between MKT, standards-based mathematics teaching practices, and student mathematics achievement that have been theorized and tested previously in other samples. The second considers the role of the *RC* approach (and actual use of the *RC* practices) for supporting standards-based mathematics instruction and improving student achievement. In addressing the latter role, the work considers the challenges of enacting standards-based mathematics instruction successfully and improving student mathematics achievement within the socially complex nature of the typical third-grade classroom.

A longstanding goal in education research is to identify classroom inputs and aspects of teachers that account for shifts and improve student learning (Rivkin, Hanushek, & Kain, 2005). Existing work suggests that higher teacher knowledge is a strong predictor of teacher practice and student learning (Ball et al., 2008; Rockoff, Jacob, Kane, & Staiger, 2008). It is worth noting that none of the direct or indirect relations between MKT, standards-based mathematics teaching practice, and student mathematics achievement that have been theorized and reported in other samples (Hill et al., 2007) were significant in our final model. However, the findings are consistent with more recent work that finds varying relations between teacher knowledge, teacher practice, and student mathematics learning in different classroom conditions (Kersting et al., 2010, 2012; Shechtman et al., 2010). The findings presented here do not imply that teacher knowledge is not important. Rather, the results support moving beyond testing direct effects of teacher knowledge and instead examining additional classroom contextual factors that could alter these relations within the diverse classroom instructional system (Shechtman et al., 2010).

Fidelity of implementation, measured here as teachers' use of *RC* practices both in intervention and control conditions, provides insights about how teachers and classroom organization can influence student learning in different classroom settings. The presence of *RC* practices in both the intervention and control group adds support to the idea that many of the practices promoted by the *RC* intervention represent "best practices" in teaching and can occur in traditional classrooms (Abry et al., 2013). Although theoretically we would expect to see positive links from standards-based mathematics teaching practices to third-grade mathematics achievement in both intervention and control samples, these were only evident in the *RC* condition in analyses accounting for fidelity of implementation. Specifically, in the intervention condition, findings show a positive relation between use of *RC* practices and standards-based mathematics practices. These findings are consistent with recent work by Abry and colleagues (2013), Rimm-Kaufman and colleagues (2014), and Goodwin

(2011), suggesting that training in an intervention, on its own, is not sufficient for changing teacher practice and child outcomes. The intervention needs to be used as intended to convey benefits to teachers and students.

The use of multigroup path analyses provides a richer understanding about how contextual conditions (e.g., receiving training in *RC* vs. not) can alter the relations between key teacher and student variables in the instructional system. The significant relations between mathematics teaching practices and student mathematics achievement were not present in the control conditions. Thus, findings reveal an important pattern in the intervention condition only; that is, using *RC* practices (taught in *RC* training) results in tighter associations among teacher and student variables within the classroom instructional system and accounts for a larger percentage of variance in student achievement. The inclusion of use of *RC* practices to the instructional system accounted for 17% of the teacher-level variance in student achievement in the *RC* group, compared with only 3% in the control condition.

While this study does not allow us to determine which elements of *RC* are most responsible for these shifts in mathematics teaching practices and student mathematics achievement or why these relations did not exist in control schools, we can speculate about potential mechanisms that could help explain these higher relations in treatment schools. Professional development and training in *RC* may trigger a process of behavioral change that makes teachers more aware of their mathematical and pedagogical weaknesses and their students' individual differences (Rimm-Kaufman, Storm, Sawyer, Pianta, & LaParo, 2006; Turner, Warzon, & Christensen, 2011), resulting in teachers' improved efforts to create a supportive and caring classroom environment. Through enhancing the social, emotional, and organizational dynamics of the classroom, many of the barriers and demands that exist in the classroom can be removed and teachers may be better able to provide higher quality mathematical practice, regardless of their levels of mathematical content knowledge. This is consistent with SEL work suggesting that embedding social and emotional learning interventions in content-specific instruction can lead to higher quality instructional practice and student learning (Brown et al., 2010).

The strategies taught in the *RC* intervention may also create teacher and student readiness for using standards-based teaching practices by modifying behavioral management approaches, decreasing chaos in the classroom, and emphasizing the need to meet children's social and emotional needs. These findings are consistent with Spillane and Zeuli (1999), who argue that social and behavioral aspects of teaching, as well as epistemological regularities of the classroom, must both be addressed when improving teacher practice and that these behavioral regularities may be more responsive to reform than content-specific aspects. This has also been noted in several studies examining the role of teacher knowledge, instructional practice, and effective professional development programs for educators (Osterman & Kottkamp, 2004; Santagata et al., 2011; Silverman & Thompson, 2008; Simon, Tzur, Heinz, & Kinzel, 2004). Although the process of teacher change after professional development has been documented in SEL research (Abry et al., 2013; Adalbjarnardottir & Selman, 1997; Brown et al., 2010; Rimm-Kaufman et al., 2006), considering the ways in which the process of learning new techniques to facilitate a positive classroom climate opens up opportunities for teachers to reflect on their mathematics instructional practice is a new area of inquiry.

Another potential explanation is that *RC* provides specific positive strategies for improving teacher language and teacher behavior that may stimulate more challenging classroom tasks, facilitate more classroom discourse, and foster autonomous student efforts. For example, *RC* encourages student-teacher and student-student interactions in *Morning Meetings* and throughout the day, which create more opportunities for students to participate freely in discussions, ask questions, and provide explanations for their actions (NEFC, 2007). These social and communication skills may carry over into mathematics instruction, making the mathematics classroom more interactive and student focused. Through *Positive Teacher Language*, the *RC* approach focuses on increasing the quality of teachers' language and feedback to students. As teachers learn how to use language more effectively, they create more back-and-forth communication between students and themselves. The increase of classroom math discussions may change the social and mathematical interactions between teachers and students. In environments rich with student discourse, questioning, and participation, teachers may be better able to address student questions (Ball, Hill, & Bass, 2005; Borko & Putman, 1995; Walkowiak, 2010) and provide more varied and cognitively demanding standards-based instruction. This is consistent with other professional development work that finds treatment teachers allow students more time to grapple with higher demanding mathematics problems than teachers in the control group (Santagata et al., 2011). Perhaps these increased opportunities for mathematical problem solving, collaboration, and discourse may positively impact teacher instruction and student learning. This explanation is consistent with work by Chapin and O'Connor (2012), who suggest that interventions and practices that increase classroom community, knowledge, and reasoning and develop classroom norms and discourse can provide students with greater opportunities to learn, resulting in improved student learning for many students.

Limitations

Two limitations require mention. First, despite the randomized design and large sample of students, statistical power in the current study was limited due to the multigroup models, randomization at the school level (n =

24), the small number of teachers (44 in each group), and the nested design. The present findings are still evident despite the concern about low power, suggesting the robustness of these results. However, it is possible that the direct effects from MKT to teacher practice and the indirect effects from MKT to achievement through use of *RC* practices and mathematics teaching practices might become more robust in a larger sample. Future research at a large scale is needed to investigate these direct and indirect paths in more detail. A second limitation is that our estimates of standards-based mathematics teaching practices are only a sampling of three school days during the year. The decision to sample three days was based on recommendations for observational work in elementary classrooms (Pianta et al., 2008; Stuhlman, Hamre, Downer, & Pianta, 2009) and speaks to resource limitations associated with such intensive data collection efforts. Future mixed-methods is needed to address the implications of our sampling decision.

Implications

These findings can be interpreted in the context of contemporary national efforts intended to improve children's mathematics achievement (National Mathematics Advisory Panel, 2008). The majority of the states in the United States have adopted the Common Core Standards (CCSSI, 2010). The adoption requires major changes to teachers' day-to-day instructional practices, and thus, change efforts will need to help teachers learn how to effectively integrate these new standards into strong mathematics teaching practices and interactions between the content, students, and teachers.

A current emphasis in teacher education is to improve preservice and in-service teachers' content knowledge (Fennema et al., 1996; Hill & Ball, 2004; U.S Department of Education, 2010) through additional course work, changes to the curriculum, and professional development programs. However, the present findings suggest that solely focusing on enriching teachers' content knowledge without helping teachers understand how to provide more supportive and effective instructional practices (Fennema & Franke, 1992) may fall short of the goal of improving instruction to raise student achievement. Part of the challenge of implementing standards-based practices may lie in the difficulties of managing the social, behavioral, and organizational aspects of the classroom, suggesting the need for focused scrutiny on ways to decrease these classroom barriers. Efforts that strive to improve contextual supports in the classroom hold potential to help teachers improve day-to-day classroom practices and interactions with students that can lead to improved student learning. The current work suggests the importance of professional development that supports the process of blending teacher knowledge components with useful social and emotional learning and classroom management practices to improve mathematics teaching and learning.

Note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grants R305A070063, R305B090002, and R305B100009 to the University of Virginia. The opinions expressed here are those of the authors and do not represent views of the U.S. Department of Education.

References

Abry, T., Brewer, A., Nathanson, L., Sawyer, B., & Rimm-Kaufman, S.E. (2010). *Classroom practices observation measure* (Unpublished measure). Charlottesville, VA: University of Virginia.

Abry, T., Rimm-Kaufman, S. E., Larsen, R. A., & Brewer, A. J. (2013). The influence of fidelity of implementation on teacher–student interaction quality in the context of a randomized controlled trial of the *Responsive Classroom* approach. *Journal of School Psychology*, *51*(4), 437–453.

Adalbjarnardottir, S., & Selman, R. L. (1997). "I feel I received a new vision": An analysis of teachers professional development as they work with students on interpersonal issues. *Teaching and Teacher Education*, *13*, 409–428.

Bakker, A. B., Hakanen, J. J., Demerouti, E., & Xanthopoulou, D. (2007). Job resources boost work engagement, particularly when job demands are high. *Journal of Educational Psychology*, *99*(2), 274–284.

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching. *American Educator*, *29*(3), 14.

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

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246.

Berry, R. Q., Rimm-Kaufman, S. E., Ottmar, E. R., Walkowiak, T. A., & Merritt, E. G. (2010). *The Mathematics-Scan coding guide* (Unpublished measure). Charlottesville, VA: University of Virginia.

Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning*. New York, NY: Routledge.

Borko, H., & Putman, R. (1995). Expanding a teachers' knowledge base: A cognitive psychological perspective on professional development. In T. R. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms & practices* (pp. 35–65). New York, NY: Teachers College Press.

Borko, H., Stecher, B. M., Alonzo, A. C., Moncure, S., & McClam, S. (2005). Artifact packages for characterizing classroom practice: A pilot study. *Educational Assessment*, *10*(2), 73–104.

Borman, G., & Overman, L. (2004). Academic resilience in mathematics among poor and minority students. *The Elementary School Journal*, *104*(3), 177–195.

Brock, L., Nashida, T., Chiong, C., Grimm, K., & Rimm-Kaufman, S. (2008). Children's perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the *Responsive Classroom* approach. *Journal of School Psychology*, *46*(2), 129–149.

Brown, J., Jones, S., LaRusso, M., & Aber, L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4R's program. *Journal of Educational Psychology*, *102*(1), 153–167.

Chapin, S., & O'Connor, C. (2012). Project Challenge: Using challenging curriculum and mathematical discourse to help all students learn. In C. Dudley-Marling & S. Michaels (Eds.), *High expectation curricula: Helping all students succeed with powerful learning* (pp. 113–127). New York, NY: Teachers College Press.

Chapin, S., O'Connor, C., & Anderson, N. (2009). *Classroom discussions: Using math talk to help students learn, grades K–6* (2nd ed.). Sausalito, CA: Math Solutions Publications.

Charalambous, C. (2010). Mathematical knowledge for teaching and task unfolding: An exploratory study. *The Elementary School Journal*, *110*(3), 247–278.

Chetty, R., Friedman, J., & Rockoff, J. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (NBER Working Paper 17699). Retrieved from http://www.nber.org/papers/w17699.pdf

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education*, *22*(1), 3–29.

Cohen, D. K. (1996). Standards-based school reform: Policy, practice, and performance. In H. F. Ladd (Ed.), *Holding schools accountable* (pp. 99–127). Washington, DC: Brookings Institution.

Common Core State Standards Initiative. (2010). *Common Core State Standards for mathematics*. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology*, *67*(5), 648–657.

Darling-Hammond, L. (1996). The quiet revolution: Rethinking teacher development. *Educational Leadership*, *53*(6), 4–10.

Darling-Hammond, L. (2002). How teacher education matters. *Journal of Teacher Education*, *51*, 166–173.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27*, 403–434.

Fennema, E., & Franke, M. (1992). Teachers' knowledge and its impact. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147–164). New York, NY: Macmillan Publishing Company.

Goldring, R., Gray, L., & Bitterman, A. (2013). *Characteristics of public and private elementary and secondary school teachers in the United States: Results from the 2011-2012 schools and staffing survey* (NCES 2013-314). Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Goodwin, B. (2011). Bullying is common and subtle. *Educational Leadership*, *69*(10), 82–84.

Grossman, P., Schoenfeld, A., & Lee, C. (2005). Teaching subject matter. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 201–231). San Francisco, CA: Jossey-Bass.

Hagelskamp, C., Brackett, M., Rivers, S., & Salovey, P. (2013). Improving classroom quality with the RULER approach to social and emotional learning: Proximal and distal outcomes. *American Journal of Community Psychology*, *51*(3), 530–543.

Harcourt Educational Measurement. (2002). *Stanford achievement test* (10th ed.). San Antonio, TX: Author.
Hiebert, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. Washington, DC: NCES.
Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Greenwich, CT: Information Age.
Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classroom teaching: Lessons from the TIMSS video study. *The Elementary School Journal*, *101*(1), 3–20.
Hill, H., & Ball, D. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal of Research in Mathematics Education*, *35*(5), 330–351.
Hill, H., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*(3), 794–831.
Hill, H. C., Ball, D. L., Blunk, M., Goffney, I. M., & Rowan, B. (2007). Validating the ecological assumption: The relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research and Perspectives*, *5*(2), 107–117.
Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430–511.
Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406.
Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11–30.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.
Hughes, J., & Kwok, O. (2007). Influence of student-teacher and parent-teacher relationships on lower achievement readers' engagement and achievement in the primary grades. *Journal of Educational Psychology*, *99*(1), 39–51.
Jones, S., Brown, K., & Aber, J. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, *82*, 533–554.
Kazemi, E. (1998). Discourse that promotes conceptual understanding. *Teaching Children Mathematics*, *4*, 410–414.
Kersting, N., GIvvin, K., Sotelo, F., & Stigler, J. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, *61*, 172–181.
Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, *49*(3), 568–589.
Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Moyer, P. S. (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics*, *47*(2), 175–197.

Muthen, L. K., & Muthen, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthen & Muthen.

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267–316.

National Center for Education Statistics. (2011). *The nation's report card: Mathematics 2011* (NCES 2012–458). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

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

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Northeast Foundation for Children, Inc. (2007). *Responsive classroom level 1 resources book*. Turners Falls, MA: Author.

Northeast Foundation for Children, Inc. (2009). *Responsive classroom level 2 resources book*. Turners Falls, MA: Author.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Osterman, K. F., & Kottkamp, R. B. (2004). *Reflective practice for educators*. Thousand Oaks, CA: Corwin Press.

Ottmar, E. R., Rimm-Kaufman, S. E., Berry, R. Q., & Larsen, R. A. A. (2013). Results from a randomized controlled trial: Does the *Responsive Classroom* approach affect the use of standards-based mathematics teaching practices? *Elementary School Journal*, *113*(3), 434–457.

Park, N., & Huebner, E. S. (2005). A cross-cultural study of the levels and correlates of life satisfaction among adolescents. *Journal of Cross-Cultural Psychology*, *36*, 444–456.

Pianta, R., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science*, *315*, 1795–1796.

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

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System [CLASS] manual: Pre-k*. Baltimore, MD: Brookes Publishing.

Rimm-Kaufman, S. E., & Chiu, Y. I. (2007). Promoting social and academic competence in the classroom: An intervention study examining the contribution of the *Responsive Classroom* approach. *Psychology in the Schools*, *44*(4), 397–413.

Rimm-Kaufman, S. E., Fan, X., Chiu, Y. I., & You, W. (2007). The contribution of the *Responsive Classroom* approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology*, *45*, 401–421.

Rimm-Kaufman, S. E., Larsen, R. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J .B., & . . . DeCoster, J. (2014). Efficacy of the *Responsive Classroom* approach: Results from a 3-year longitudinal randomized controlled trial. *American Educational Research Journal*, *51*(3), 567–603.

Rimm-Kaufman, S. E., Storm, M., Sawyer, B., Pianta, R. C., & La Paro, K. (2006). The Teacher Belief Q-Sort: A measure of teachers' priorities and beliefs in relation to disciplinary practices, teaching practices, and beliefs about children. *Journal of School Psychology*, *44*, 141–165.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2008). *Can you recognize an effective teacher when you recruit one?* (NBER Working Paper 14485). Cambridge, MA: National Bureau of Economic Research.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools. *The Teachers College Record*, *104*(8), 1525–1567.

Ryan, A., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, *38*(2), 437–460.

Santagata, R., Kersting, N., Givven, K. B., & Stigler, J. W. (2011) Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students' mathematics learning. *Journal of Research on Educational Effectiveness*, *4*, 1–24.

Sawada, D., & Piburn, M. (2002). *Reformed Teaching Observation Protocol (RTOP): Training Guide*. Phoenix, AZ: Arizona Board of Regents.

Scheuell, T. J. (1996). Teaching and learning in a classroom context. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726–764). New York, NY: Simon & Schuster MacMillian.

Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (2010). Investigating links from teacher knowledge to classroom practice to student learning in the instructional system of the middle school mathematics classroom. *Cognition and Instruction*, *28*(3), 317–359.

Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction*, *20*(2), 119–150.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Silverman, J., & Thompson, P. W. (2008). Toward a framework for the development of mathematical knowledge for teaching. *Journal of Mathematics Teacher Education*, *11*, 499–511.

Simon, M., Tzur, R., Heinz, K., & Kinzel, M. (2004). Explicating a mechanism for conceptual learning: Elaborating the construct of reflective abstraction. *Journal for Research in Mathematics Education*, *35*(5), 305–329.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education*, *74*(3), 267–288.

Skibbe, L., Decker, L., & Rimm-Kaufman, S. (2006, June). *The contribution of the responsive classroom approach to classroom quality and teaching practices*. Poster presented at the Institute for Education Sciences, Washington, DC.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy*, *21*(1), 1–27.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488.

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

Stuhlman, M., Hamre, B., Downer, J., & Pianta, R. (2009). *A practitioner's guide to conducting classroom observations: What the research tells us about choosing and using observational systems*. New York, NY: W. T. Grant Foundation.

Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics* (4th ed.). New York, NY: Harper Collins.

Turner, J., Warzon, K., & Christensen, A. (2011). Motivating mathematics learning: Changes in teachers' practices and beliefs during a nine-month collaboration. *American Educational Research Journal*, *48*(3), 718–762.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2010). *ESEA blueprint for reform*. Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint/index.html.

Walkowiak, T. (2010). *Third grade mathematics instructional quality, mathematical knowledge for teaching, and mathematics teaching self-efficacy: A quantitative and qualitative analysis*. Unpublished doctoral dissertation.

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.

Yackel, E., Cobb, P., & Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education*, *22*(5), 2–33.

Manuscript received May 9, 2012
Final revision received September 26, 2014
Accepted January 18, 2015