Research Article

Hammill Institute on Disabilities

Journal of Learning Disabilities
2023, Vol. 56(4) 243–256

DOI: 10.1177/00222194221102644
journaloflearningdisabilities.sagepub.com
SAGE

Examining Interactions Across Instructional Tiers: Do Features of Tier 1 Predict Student Responsiveness to Tier 2 Mathematics Intervention?

Marah Sutherland, PhD[1,2], Taylor Lesner, MS[1], Derek Kosty, PhD[1,3], Cayla Lussier, BA[1], Keith Smolkowski, PhD[1,3], Jessica Turtura, PhD[1,2], Christian T. Doabler, PhD[4], and Ben Clarke, PhD[1,2]

Abstract

High-quality Tier 1 instruction is frequently conceptualized as the "foundation" for other tiers of intervention within multitiered systems of support (MTSS) models. However, the vast majority of Tier 2 intervention studies do not account for Tier 1 variables when examining intervention effectiveness. The purpose of this study was to examine Tier 1 predictors, or "quality indicators," of differential responsiveness to Tier 2 mathematics intervention. Data were drawn from a large-scale data set where all teachers taught the Early Learning in Mathematics (Tier 1) core program across the academic year, and a subset of students were selected for the ROOTS (Tier 2) mathematics intervention. We examined the following Tier 1 variables: (a) classroom-level mathematics gains, (b) Tier 1 fidelity of implementation, (c) Tier 1 classroom management and instructional support, and (d) class size. Response to Tier 2 intervention was not significantly predicted by any of the Tier 1 variables examined; however, the pattern of Hedges' *g* effect sizes suggested that students with higher quality of Tier 1 instruction tended to benefit less from the Tier 2 ROOTS intervention. Results are discussed in the context of implications for research and practice.

Keywords

multitiered systems of support (MTSS), Tier 2 mathematics intervention, core mathematics instruction, differential responsiveness, classroom management and instructional support, fidelity of implementation

One of the most significant shifts in school-based assessment and intervention efforts over the past two decades has been the introduction and widespread adoption of multitiered systems of support (MTSS). MTSS models are designed to support the learning of all students through a continuum of supports that increase in intensity based on student need. Such models are frequently depicted as a triangle or pyramid, with core (i.e., Tier 1) instruction as the "foundation" that is designed to meet the needs of roughly 80% of students in a school building (Balu et al., 2015; Gersten et al., 2009). Schools identify students at risk for academic or behavioral difficulties through universal screening and provide interventions matched to the intensity of student needs (i.e., Tiers 2 and 3), with the goal of reducing or eliminating achievement gaps (Gersten et al., 2009). Throughout the school year, screening and progress monitoring data are used to determine whether students are making adequate progress toward instructional goals and whether the intensity of supports should be adjusted.

MTSS models have become increasingly prevalent in U.S. schools. A recent review conducted by Berkeley and colleagues (2020) revealed that 47 states endorse MTSS implementation guidelines, representing a vast increase in state-level adoption of multitiered models since 2007 (Berkeley et al., 2009). Despite their widespread use, there is substantial variation in the implementation of MTSS in schools, including differences in the number of tiers in a given model, whether models address academics and behavior together or separately, and communication and

[1]University of Oregon, Eugene, OR, USA
[2]Center on Teaching and Learning, Eugene, OR, USA
[3]Oregon Research Institute, Eugene, OR, USA
[4]The University of Texas at Austin, Austin, TX, USA

Corresponding Author:
Marah Sutherland, PhD, University of Oregon, 1600 Millrace Drive, Suite 207, Eugene, OR 97403, USA.
Email: marahs@uoregon.edu

support for implementation (Berkeley et al., 2020; Lam & McMaster, 2014). Furthermore, successful adoption of multitier models is unlikely without organized and cohesive efforts spanning special and general education, allocation of resources to support effective assessment practices and intervention implementation, and buy-in from multiple stakeholders at the building and district level (Fuchs & Fuchs, 2017; Kratchowill et al., 2007; Nellis, 2012).

A national evaluation of MTSS, conducted by Balu and colleagues, illustrates many of these complexities. Balu et al. (2015) examined reading outcomes for "veteran" schools implementing MTSS across 13 states and 146 schools. A regression discontinuity design was used to compare reading outcomes of students identified to receive Tier 2 or 3 intervention and those just above the cut point who were not identified for additional intervention supports. Results indicated that overall, there were no statistically significant benefits for second- and third-grade students attending MTSS schools, and negative effects were found for students assigned to Tier 2 or Tier 3 interventions.

Since publication, several critiques regarding the study design, limitations of the analyses used, and the lack of implementation fidelity in Balu et al.'s (2015) study have been discussed in the literature (e.g., Fuchs & Fuchs, 2017; Gersten et al., 2017). Many of these limitations centered on the quality and consistency of Tier 1 instruction. For example, Gersten et al. (2017) pointed out that in 60% of classrooms, Tier 2 intervention supplanted, rather than supplemented, Tier 1 instruction, a practice that does not adhere to MTSS guidelines. In addition, information about the specific Tier 1 programs MTSS schools were using, and the degree of alignment between core and intervention programs, was not provided. Balu et al. (2015) suggest that poor alignment between interventions and core instruction may provide a plausible explanation for the negative impact of assignment to intervention in MTSS schools. These findings illustrate the complexities of implementing multitier systems in schools and highlight an underemphasized and under-researched component of MTSS across both research and practice: high-quality core instruction.

Tier 1 Mathematics Programs: The Importance of the Core

Kindergarten core mathematics instruction represents the first exposure to mathematical concepts and skills for many students. In the early elementary grades, students must develop conceptual understanding of and procedural fluency with foundational whole number content, including number identification and counting, addition and subtraction, and understanding numbers as quantities that can be composed (i.e., put together) and decomposed (i.e., taken apart; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). One central goal of core instruction is to remediate early mathematics difficulty (MD) by developing a strong foundation of math knowledge, potentially decreasing the need for more intensive intervention later on. The importance of core mathematics instruction in the early grades (e.g., K-2) cannot be understated, given the different levels of background knowledge that students bring to formal schooling (Jordan et al., 2010). With core instruction designed to serve all students in a classroom, general education teachers must teach to a wide variety of learners, including those at risk for MD. As a result, the research on core mathematics instruction has typically focused on at-risk populations of learners as well as the general student population.

Only a handful of studies have rigorously examined the effectiveness of Tier 1 mathematics programs. In perhaps the largest-scale evaluation of core mathematics programs to date, Agodini and colleagues (2010) randomly assigned 110 elementary schools across the United States to implement one of four widely used first- and second-grade mathematics curricula—*Investigations in Number, Data, and Space (Investigations); Math Expressions; Saxon Math*; and *Scott Foresman-Addison Wesley Mathematics (SFAW)*. The authors found evidence of differential mathematics achievement gains across these programs, with Hedges' g effect sizes ranging from 0.11 to 0.17 depending on the curriculum differential and grade level examined. Importantly, curriculum differentials tended to favor curricula that used teacher-directed or blended approaches (e.g., *Math Expressions* and *Saxon Math*) over student-centered or non-explicit approaches.

In another large-scale efficacy trial, Clarke et al. (2015, 2011) randomly assigned 129 kindergarten classrooms to implement Early Learning in Mathematics (ELM), a core program intentionally designed to support students at risk for MD, or business-as-usual (BAU) core math instruction. The authors found no overall differences in mathematics achievement gains across the kindergarten year between ELM and BAU classrooms. However, among students at risk for developing MD, those in ELM classrooms made significantly greater gains across the kindergarten year, resulting in decreased achievement gaps between at-risk students and their typically achieving peers.

In a smaller-scale study, Sood and Jitendra (2011) evaluated a supplement to core instruction designed to support students at risk for MD. The authors randomly assigned five kindergarten classrooms in a high-poverty area to a 4-week number sense program that replaced part of the core curriculum, or a BAU control condition, while holding instructional time constant. The number sense program yielded medium to large effects at posttest (Hedges' g = 0.55–0.87) and 3-week follow-up (Hedges' g = 0.68–1.20), with comparable gains regardless of students' mathematics risk status at pretest. This suggests that even small changes to core instruction may have immediate and lasting benefits for students. Taken together, these findings suggest that effective core instruction can boost student mathematics achievement, even for the most vulnerable learners, and may therefore lessen the need for intensive interventions.

Tier 2 Interventions Within the Context of Core Instruction

Even with an effective Tier 1 program designed to support a range of learners in place, some students will need more intensive and targeted support to progress at a comparable rate to their typically achieving peers. Within MTSS, Tier 2 should be delivered in addition to Tier 1. In grades K-2, Tier 1 and 2 mathematics instruction is typically situated within the general education classroom setting. Thus, Tier 2 should be a natural extension of Tier 1 designed to provide additional support to the students that require it.

An increased national focus on supporting students at risk for MD has led to a growing literature base on mathematics interventions, particularly in the area of whole number (e.g., Jitendra et al., 2021). Unfortunately, few mathematics intervention studies have attempted to account for variation in the quality of Tier 1 instruction. More often than not, intervention studies in both mathematics and reading are conducted without regard to contextual factors associated with core instruction, likely due to the difficulty of measuring and controlling for variability in Tier 1 practices across multiple schools and districts. Several researchers have noted the importance of capturing Tier 1 variability in the context of Tier 2 intervention, given that supports students receive across tiers likely interact to influence outcomes (e.g., Bailey et al., 2020; Coyne et al., 2018). Measuring the quality of core instruction and assessing its alignment with the intervention may yield important insights about intervention effectiveness across contexts (Hill et al., 2012).

Against that backdrop, a few researchers have attempted to control for the quality of Tier 1 in the context of Tier 2 intervention. In an efficacy study of a small-group mathematics problem-solving intervention, Fuchs et al. (2008) randomly assigned 119 third-grade classrooms to one of two core (i.e., Tier 1) mathematics programs: conventional problem-solving instruction, or a validated, schema-broadening problem-solving program. Students at-risk for MD were then randomly assigned to receive a schema-broadening tutoring program, Hot Math, or BAU math intervention supports. This design allowed the researchers to determine whether there was a differential positive impact for students who received aligned, schema-broadening problem-solving instruction across core and intervention supports. Findings indicated that Hot Math tutoring was significantly more effective for students who were also provided with aligned, validated core instruction, compared to those who received the intervention in the context of BAU core instruction. The authors concluded that two tiers of validated, highly aligned instruction were more effective than one. While researchers have begun to explore how Tier 1 practices may impact Tier 2 and 3 intervention effectiveness (Al Otaiba et al., 2014), more research is needed to better understand how variability in the quality of core instruction may differentially impact intervention outcomes.

Digging Deeper to Explore Tier 1 Predictors of Tier 2 Intervention Outcomes

One method to explore for whom and under what conditions an intervention is likely to be effective is evaluating potential moderators of intervention outcomes (Fuchs & Fuchs, 2019). Despite significant advances in the knowledge base supporting early mathematics intervention, many questions concerning predictors of intervention impact remain unanswered. Our purpose in this study was to examine associations among key Tier 1 variables and kindergarten students' responsiveness to an aligned Tier 2 mathematics intervention targeting whole number knowledge. To investigate this question, we analyzed data collected during an efficacy trial of a Tier 2 kindergarten mathematics intervention program (ROOTS) that was implemented in the context of a research-based core mathematics program (ELM; Clarke et al., 2016). That is, all participating classrooms used the core program ELM, and a subset of at-risk students were identified to receive the Tier 2 ROOTS program in addition to ELM core instruction.

In the original ROOTS efficacy trial conducted in Oregon schools, the researchers found significant positive effects for students at risk in mathematics (Hedges' g values ranged from .30 to .38; Clarke et al., 2016). However, in a conceptual replication of the original ROOTS study, conducted in Texas schools, the researchers did not find overall or differential impacts of ROOTS (Clarke et al., 2022). To explain this finding, the researchers pointed to several Tier 1 variables that differed between participating classrooms in the original ROOTS study and the replication study, though they did not specifically test Tier 1 variables as predictors of intervention effectiveness.

In this study, we proposed to further unpack Tier 1 variables that may have led to students' differential responsiveness to ROOTS, using combined data from the original efficacy study and the replication study. We refer to these Tier 1 variables as "quality indicators" throughout this paper. The term "quality indicator" originally stems from the medical field as a set of standards to measure and compare the quality of health care across settings (e.g., Stelfox & Straus, 2013) but has been broadly used to describe expectations for high-quality features and practices across disciplines. For our purposes, the term "quality indicator" is used to describe Tier 1 variables that are associated with increased student mathematics achievement, as summarized in the sections below. Note that these are distinct from the quality indicators described in other areas of education such as the quality indicators for group design research (e.g., Gersten et al., 2005).

Classroom-level mathematics gains. While they did not specifically test for Tier 1 differences, Clarke et al. (2022) hypothesized that the outcomes of the ROOTS conceptual replication study were nonsignificant due to differences in core instruction at the replication site. As described above, Fuchs et al. (2008) found that a third-grade mathematics tutoring intervention was significantly *more* effective when implemented in the context of validated core instruction. Thus, the role of core instruction and the gains that students make across the school year may affect their response to Tier 2 intervention. In this study, we conceptualized classroom-level mathematics gains as a proxy for general effectiveness of core instruction and hypothesized that ROOTS students participating in classrooms that made larger gains across the kindergarten year would experience smaller gains as a result of the intervention.

Tier 1 fidelity of implementation. Implementation fidelity is often conceptualized as adherence—or the degree to which an intervention or curricular program is delivered as planned (Harn et al., 2013; Moncher & Prinz, 1991). Researchers have advocated for the importance of high implementation fidelity within and across tiers of MTSS (Keller-Margulis, 2012; Scott et al., 2019). While measuring implementation fidelity has become increasingly recognized as an essential component of efficacy and effectiveness studies (DeFouw et al., 2019; O'Donnell, 2008), it is less commonly examined as a predictor of student outcomes. For example, O'Donnell (2008) conducted a systematic literature review of studies that investigated core K-12 programs and included a measure of implementation fidelity. Less than a quarter of studies included in the review examined associations between fidelity of implementation and student outcomes, though the few studies that did found positive associations (O'Donnell, 2008). In this study, we hypothesized that ROOTS students participating in classrooms of teachers with greater adherence fidelity to ELM would experience smaller gains from the intervention.

Tier 1 classroom management and instructional support. While implementation fidelity measures in the math literature most commonly target adherence (O'Donnell, 2008), fidelity can also be conceptualized as a measure of instructional quality (Nelson et al., 2019), including teaching behaviors such as classroom management and instructional skills. In mathematics, these constructs have been investigated primarily through direct observation using low-, moderate-, and high-inference measures to examine associations with student mathematics outcomes (Doabler et al., 2015, 2019; Pianta & Hamre, 2009). Some examinations of instructional quality have centered on direct observations of teaching behaviors, such as teacher models and individual and group student practice opportunities. For example, Doabler et al. (2019) found that teacher-facilitated individual practice opportunities moderated the relationship between students' initial skill and mathematics outcomes in the context of a Tier 2 mathematics intervention. Other investigations of implementation fidelity have demonstrated positive associations between observation instruments that rely on observer impressions and student mathematics achievement (Jiménez et al., 2021; Pianta & Hamre, 2009). While classroom management and instructional support are demonstrated predictors of student mathematics outcomes within Tier 1 and Tier 2 settings (Doabler et al., 2019; Pianta & Hamre, 2009), to our knowledge, Tier 1 instructional quality has not been examined as a predictor of student responsiveness to Tier 2 intervention. We hypothesized that ROOTS students participating in classrooms with greater instructional quality would experience a smaller benefit from the supplemental intervention.

Class size. Several studies demonstrate that increased class size is negatively associated with student mathematics achievement in the United States (Nye et al., 2000; Pong & Pallas, 2001). Of note is the large-scale Tennessee Class Size Experiment, which demonstrated that smaller classes were beneficial for a wide range of students across schools and districts (Nye et al., 2000). Researchers posit that the mechanism underlying the beneficial impact of smaller class sizes could be more frequent teacher–student interactions or the teacher's ability to devote more individual attention to differentiate instruction for students with diverse learning needs (Nye et al., 2000). We hypothesized that ROOTS students participating in classrooms with a smaller number of students would experience a smaller benefit from ROOTS.

Research Question

Our research question was as follows: Do Tier 1 quality indicators predict at-risk students' responsiveness to a Tier 2 (ROOTS) intervention? As described above, the Tier 1 quality indicators we examined included: (a) classroom-level gains on a broad mathematics achievement measure, (b) Tier 1 fidelity of implementation, (c) Tier 1 classroom management and instructional support, and (d) class size.

Method

Research Design

This study analyzed data from Years 2 (Oregon – 2009–2010) and 3 (Texas – 2010–2011) of a 4-year efficacy trial of ROOTS. Kindergarten teachers were randomly assigned to teach ELM or ELM + ROOTS (i.e., all students received ELM core instruction). Blocking, also known as stratification, was used to control for biases that might stem from systemic differences between conditions. For example, because some teachers had taught ELM the previous year, we blocked on teachers' prior experience teaching ELM. In schools with multiple kindergarten classrooms, we blocked on classrooms by assigning classrooms to condition within schools. Across Oregon and Texas, a total of 91 classrooms were included in the current analyses: 46 in the treatment condition (ELM + ROOTS) and 45 in the control condition (ELM only). All teachers were asked to nominate the five lowest-performing students or students who would most benefit from small-group mathematics instruction. Processes differed slightly in Texas, where teachers were provided with a list of students who scored below the 40th percentile on a pretest measure and a number sense screening measure to inform their decision (for additional details, see Method section of Clarke et al., 2016, 2022). These processes resulted in 141 identified ROOTS-eligible students in Oregon and 308 students in Texas. In both conditions, kindergarten teachers taught ELM throughout the academic year. Students who participated in ROOTS received the supplemental intervention as well as whole-class ELM instruction. To control for instructional time, ROOTS groups took place 3 days per week at the end of the ELM lesson during independent math practice worksheets. Nominated students in ELM-only (i.e., BAU) classrooms participated in the independent math practice worksheet component of ELM rather than the more intensive ROOTS intervention. In both sites, ROOTS instruction began in January and lasted through the end of May. ROOTS was delivered by trained instructional assistants (IAs) who were already employed by the participating school districts.

Participants

Kindergarten teachers. Across sites, 91 kindergarten teachers participated in the ELM only or ELM + ROOTS conditions. Of the 66 teachers who provided demographic information (72%), 96% identified as female, 62% White, 26% Hispanic, 11% African American, and 2% Asian American. Regarding teacher-reported credentials, 35% held a master's degree, 25% completed three or more college math courses, 53% completed college algebra, and 60% had taught kindergarten for 4 or more years. In the core math block, 38% of teachers reported spending 21 to 40 min per day and 62% of teachers reported spending 41 to 60 min per day.

Instructional assistants. Across sites, a total of 28 IAs already employed in participating schools taught ROOTS; 86% were female, 83% identified themselves as White, 11% identified themselves as Hispanic, and 18% identified themselves as African American. In total, 46% of the IAs had college degrees, of whom 25% held current teacher certifications in elementary education. Of the remaining 15 IAs, 14% held an Associate's degree, and 39% were high school graduates. Fourteen (50%) of the IAs had completed college-level coursework in mathematics. IAs had varying degrees of teaching experience; 18% had 10 or more years of experience, 18% had between 4 and 6 years of teaching experience, 25% had 1 to 3 years of experience, and 39% had less than 1 year of experience.

Students. Data from 448 students were analyzed in this study. Students were drawn from two districts in Oregon, and one district in Texas. Across sites, ROOTS-eligible students were 48% female, and 31% of students were English learners. Their average age was 66.6 months old (SD = 3.9). In Oregon, 16% of students were eligible for special education services overall, and 44% and 50% of students were eligible for free or reduced lunch (Districts A and B, respectively). In Texas, 87% of students were eligible for free or reduced lunch in the participating district. In addition, while special education eligibility information was unavailable in Texas, 53% of the 283 students with a Test of Early Mathematics Achievement–Third Edition (TEMA-3) score at pretest scored at or below the 10th percentile. Other student-reported demographic information was gathered at the school or district level and is reported in detail by Clarke et al. (2016, 2022).

Measures

ELM fidelity of implementation instrument. To measure teachers' implementation fidelity to ELM, we used the ELM Fidelity of Implementation Instrument. This standardized, researcher-developed instrument was designed to measure adherence to the ELM curriculum. For each ELM activity within a lesson, observers rated fidelity using a 3-point scale (0 = *did not implement*, 0.5 = *partial implementation*, 1 = *full implementation*). Each ELM classroom underwent three rounds of observations, scheduled in the fall, winter, and spring. In this study, ELM Fidelity of Implementation scores for each classroom are reported as the average score across observation points. On average, teachers implemented ELM with high levels of fidelity: fall (M = 0.93; SD = 0.14), winter (M = 0.96; SD = 0.09), and spring (M = 0.95; SD = 0.12).

Ratings of Classroom Management and Instructional Support. We used the Ratings of Classroom Management and Instructional Support (RCMIS; Doabler & Nelson-Walker, 2009) in this study as a broad measure of instructional quality. The RCMIS includes 11 items that address general features of effective mathematics instruction, including classroom management strategies, instructional delivery, and the learning environment. For each item, observers used a 4-point rating scale to indicate whether components of effective mathematics instruction were present (1 = *not present*, 4 = *highly present*). Observers used a detailed scoring rubric to differentiate between ratings. Observers completed the RCMIS at the end of each classroom observation. Internal consistency of the RCMIS was high, with Cronbach's alpha equal to .92. The authors report fairly stable ratings of instructional quality (intraclass correlation coefficient or ICC = .33), indicating that the three observations provided reasonable estimates. For each classroom, RCMIS scores are reported as the average item score across the three observation time points.

Test of Early Mathematics Ability–Third Edition. The Test of Early Mathematics Ability–Third Edition (TEMA-3; Ginsburg & Baroody, 2003) is a norm-referenced assessment designed to measure informal and formal mathematics skills, intended for use with children ages 3 to 8 years 11 months. The TEMA-3 includes a range of items, sampling across skills in the domains of numbering, comparing numbers, number facts, calculation skills, and related mathematical concepts. The TEMA-3 has high internal reliability (coefficient alphas range from .94 to .96) and moderate criterion-related validity with other measures of early mathematics skills (.54 to .91). The TEMA-3 was selected as a mathematics outcome measure in this study, given its sensitivity to students scoring at the lower end of the distribution. Test of Early Mathematics Ability–Third Edition scores are reported as standard scores.

Procedures

Data collection. The TEMA-3 was individually administered to students pre- and post-ROOTS implementation. Trained staff with extensive experience in administering educational assessments for research projects administered all student measures. Data collectors were required to obtain interrater reliability coefficients of .90 prior to collecting data. Follow-up trainings were conducted prior to each data collection period to ensure continued reliable data collection. Student assessment protocols were processed using Teleform, a form-processing program.

Observations. Classroom observations were conducted in the fall, winter, and spring. Observations were scheduled in advance but were not scheduled according to a particular lesson or specific content planned for instruction. The ELM fidelity of implementation instrument, as well as the RCMIS, were administered in addition to several other measures that were not the focus of this study. Trained observers conducted the observations, receiving approximately 14 hours of training from the project observation coordinator. Refresher trainings were conducted prior to each subsequent round of observations to minimize observer drift. Two reliability checkout were implemented. First, in the training observers coded a 5-min video of kindergarten math instruction and compared their coding to codes predetermined by the observation coordinator. Second, the observation coordinator conducted in-field reliability checks where the coordinator "shadow coded" an observation alongside the observer. All observers met minimum reliability criteria of 0.85 percent agreement across checkouts. Interobserver reliability data were collected on 61 occasions within ELM classrooms. For the RCMIS and ELM fidelity of implementation instrument, moderate to high interobserver reliability was found with ICCs of .86 and .63, respectively.

Early Learning in Mathematics curriculum. ELM is a 120-lesson, core kindergarten curriculum designed for whole-class instruction and focused on building early foundational mathematical skills (see Clarke et al., 2015 for more details). Lessons consist of a 15-min daily calendar routine and a 45-min mathematics lesson comprised of 4 to 5 activities. In addition, each lesson includes a student Math Practice worksheet and a "Note Home" in English and Spanish to encourage parental involvement. ELM covers critical content across three mathematical strands: number and operations, geometry, and measurement. Greater emphasis is placed on the development of whole number skills in comparison to the other two strands (National Council of Teachers of Mathematics (NCTM), 2006). Every fifth lesson focuses on problem-solving, incorporating content across the previous four lessons within that context. ELM content aligns with the Common Core State Standards in Mathematics (2010) and was selected based on the NCTM Focal Points for kindergarten (NCTM, 2006). ELM uses an explicit and systematic instructional design (Gersten et al., 2009), including curricular supports, such as teacher scripting to assist teachers in using precise mathematical language and representations.

ROOTS intervention. ROOTS is a scripted, 50-lesson, Tier 2 kindergarten mathematics intervention program that was designed to be delivered during the second half of the kindergarten year (see Clarke et al., 2016 for more details). In comparison to ELM which covers a wider scope of mathematical content, ROOTS is designed to build students' conceptual understanding and procedural fluency within whole number operations specifically. The focus on whole number understanding is aligned with the Common Core State Standards in Mathematics (2010) and calls from expert panels for meeting the needs of students with MD (Gersten et al., 2009). ROOTS lessons are designed to last for 20 min and include 4 to 5 brief math activities that center on three domains of mathematical understanding: (a) counting and cardinality, (b) number and operations, and (c) base 10 understanding/place value. ROOTS uses the concrete-representational-abstract sequence (Agrawal & Morin, 2016) and frequently incorporates mathematical representations, such as linking cubes, base 10 blocks, place value charts, finger representations of numbers, and 10 frames to build students' conceptual understanding. When introducing students to novel mathematics concepts and skills, the program includes clear teacher models and a gradual and systematic decrease in teacher supports across lessons to promote learner independence.

Professional development and coaching support. ELM teachers participated in four 6-hr days of professional development (PD) related to program implementation and research-based principles of kindergarten math instruction. The first workshop took place before the start of the school year, and the other three were distributed evenly throughout the year. During the workshops, teachers had opportunities to practice lessons and receive feedback on their instruction from members of the PD team. ROOTS interventionists participated in three 4-hr PD workshops organized around active participation and critical math content. The first workshop covered Lessons 1 to 25, and the subsequent two covered content in the second half of ROOTS. Interventionists received three coaching visits conducted by an expert teacher to increase fidelity to the program. Coaching visits consisted of direct observation and post-lesson feedback focused on instructional delivery and implementation fidelity. More PD and coaching support details can be found in Clarke et al.'s studies (2015, 2016).

Statistical Analysis

We explored Tier 1 quality indicators as predictors of differential response to ROOTS using an expanded version of the main effects statistical model presented by Clarke et al. (2016, 2022). In the previous studies, the authors examined main effects of ROOTS with mixed-model (multilevel) Time × Condition analyses (Murray, 1998) designed to account for the intraclass correlation associated with students nested within classrooms, the level of assignment to study condition. The models estimated differences between conditions (ELM versus ELM + ROOTS) on change in outcomes from pretest (T_1) to posttest (T_2), with gains for individual students clustered within classrooms. The model included time, condition, and the Time × Condition interaction, with time coded 0 at T_1 and 1 at T_2 and condition coded 0 for ELM and 1 for ELM + ROOTS. In this study, we examined whether Tier 1 quality indicators predicted differential response to the ROOTS intervention. Therefore, we expanded the original statistical model to include each Tier 1 quality indicator, separately, and its interaction with the condition, time, and the Time × Condition terms.

Differential response to intervention implies that the condition difference in student outcomes depends on student- or group-level predictor variables (e.g., Tier 1 quality). We hypothesized larger differences between conditions (favoring ROOTS) at lower levels of Tier 1 quality. To explore these differences, we estimated the ROOTS intervention effect and its 95% confidence interval (CI) at multiple points along the distribution of the predictor variables (Jaccard & Turrisi, 2003). We used these estimates to graph the results with the method recommended by Preacher et al. (2006) for interpretation. The graphs depict the condition effect size (Hedges' g) and its 95% CIs across the range of predictor variables.

Model estimation. We fit the statistical models to our data using SAS PROC MIXED version 9.4 (SAS Institute, 2016) with maximum likelihood estimation. Maximum likelihood estimation with all available data produces potentially unbiased results even in the face of substantial missing data, provided the missing data were missing at random (Graham, 2009). In this study, missing data did not likely represent a meaningful departure from the missing at random assumption, meaning that missing data did not likely depend on unobserved determinants of the outcomes of interest (Little & Rubin, 2002). Student-level attrition explains missing data (10%), which did not significantly differ by condition and the effect of attrition on outcomes did not vary by condition (Clarke et al., 2016, 2022).

The models assume independent and normally distributed dependent variables. We addressed the first, more important assumption (van Belle, 2008) by explicitly modeling the multilevel nature of the data. The data in this study also did not markedly deviate from univariate normality; skewness was 0.1 and kurtosis was −0.6 for the TEMA-3 outcome measure.

Interpretation of results. To interpret results, we focus on Hedges' g effect sizes, their 95% CIs, and model probabilities for hypothesis tests. As recommended by the American Statistical Association (Wasserstein & Lazar, 2016), we abstained from using bright-line rules such as claims of "statistical significance" when $p < .05$. P values measure the incompatibility between the observed data and all assumptions of the statistical model, including the null hypothesis, H_0 (Greenland et al., 2016). This awkward definition determines neither which assumptions are incorrect nor the importance of the association. To complement p values, we report effect sizes, g, and model probabilities, w. The model probabilities indicate the strength of evidence

Table 1. Descriptive Statistics for Student-Level Math Achievement and Classroom-Level Tier 1 Quality Indicators by Condition and Time.

Measure	Statistic	Intervention		Control	
		Pretest	Posttest	Pretest	Posttest
Student-level TEMA-3 percentile score	*M*	77.3	94.8	76.2	93.7
	(*SD*)	(16.7)	(13.2)	(15.4)	(14.5)
	n	206	206	199	196
Class-level TEMA-3 gains	*M*		14.5		15.1
	(*SD*)		(6.3)		(6.4)
	n		46		45
Fidelity of Tier 1	*M*		0.9		0.9
	(*SD*)		(0.1)		(0.1)
	n		46		45
RCMIS score	*M*		3.1		3.1
	(*SD*)		(0.4)		(0.4)
	n		46		45
Class size	*M*		21.6		21.4
	(*SD*)		(3.7)		(4.2)
	n		46		45

Note. TEMA-3 = Test of Early Mathematics Achievement–Third Edition; RCMIS = ratings of classroom management and instructional support.

for one model when compared with others, given the data at hand. Based on the Akaike Information Criterion, Burnham et al. (2011) describe w as the probability of selecting the same model with a "replicate data set from the same system" (p. 30) and allow statements such as "the probability of [H_A] is 0.78" (p. 26). Model probabilities better characterize the chance of a replicated result than p values. In this study, we compared models for two hypotheses: a model with the three-way Tier 1 Quality × Time × Condition interaction term (H_A) and one without the three-way interaction (H_0). We reported the model probability for the model with the three-way interaction effect (H_A), and with only two models, the model probability for H_0 is $1 - w$.

Results

Attrition analyses and main effects for the ROOTS intervention in the present sample were presented by Clarke et al. (2016, 2022).

Tier 1 Quality Predictors of Differential Response to ROOTS

Table 1 displays descriptive statistics for students receiving ROOTS + ELM (Intervention), and ELM only (Control), as well as descriptive statistics of Tier 1 quality indicators in their respective classrooms. Table 2 presents tests of differential response to ROOTS as a function of the Tier 1 quality indicators. Results suggested that response to the ROOTS intervention was not statistically significantly predicted by class-level gains in TEMA-3 percentile scores ($t_{89} = 0.50$, $p = .6197$, $w = .29$), fidelity of Tier 1 implementation ($t_{89} = -1.18$, $p = .2417$, $w = .42$), Tier 1 classroom management and instructional support ($t_{89} = -1.53$, $p = .1302$, $w = .53$), or class size ($t_{89} = 1.46$, $p = .1484$, $w = .50$). In these models, w describes the probability for the model with the test of differential response compared to an equivalent model without the Tier 1 Quality × Time × Condition interaction. Models with the three-way interaction for differential response were nearly equally as likely (w's $\geq .42$) as models without the interaction for three of the four Tier 1 quality indicators: fidelity of Tier 1 implementation, Tier 1 classroom management and instructional support, and class size (see Table 2). As demonstrated in Figure 1, the pattern of results indicated that students in classrooms with lower Tier 1 quality may have benefited more from the ROOTS intervention.

Figures 1 to 3 present the ROOTS intervention effect on student outcomes (dark line) and its 95% CI (light lines) across the range of Tier 1 quality indicator variables, plotted separately. Zero on the vertical axis represents no difference between conditions. Within the confidence bounds, vertical lines represent the 5th, 25th, 50th, 75th, and 95th sample percentiles, similar to a boxplot.

Figure 1 shows the ROOTS intervention effect across the range of Tier 1 fidelity of implementation scores. The vertical lines show that about 25% of classrooms had fidelity scores below 0.90, 50% below 0.96, and 75% below 1.0. The decreasing dark line suggests that the estimated ROOTS effect decreased as fidelity of Tier 1 implementation increased. To assist with interpretation, we estimated effect sizes of 0.08 at the 25th percentile, −0.02 at the median, and −0.09 at the 75th percentile of Tier 1 fidelity scores.

Table 2. Differential Response Results from Mixed Time × Condition Analysis of Student Outcomes.

Effect or statistic		Tier 1 quality indicator (predictor): Class-level TEMA-3 gains	Fidelity of Tier 1	RCMIS score	Class size
Model probability (*w*)		0.29	0.42	0.53	0.50
Fixed effects	Intercept	75.7 (1.6)	75.8 (1.7)	75.7 (1.6)	75.6 (1.6)
	Time	17.5 (1.0)	17.7 (1.3)	17.6 (1.3)	17.5 (1.3)
	Condition	0.8 (2.2)	1.0 (2.3)	0.9 (2.2)	1.5 (2.2)
	Time × condition	0.9 (1.4)	0.3 (1.8)	0.5 (1.8)	0.5 (1.8)
	Predictor	−0.4 (0.2)	26.3 (23.5)	7.0 (3.7)	−1.1 (0.4)
	Predictor × condition	−0.9 (0.4)	14.9 (30.5)	9.6 (5.6)	−0.5 (0.6)
	Predictor × time	0.9 (0.2)	24.4 (18.3)	4.3 (3.1)	−0.1 (0.3)
	Predictor × time × condition	0.1 (0.2)	−28.2 (23.9)	−7.1 (4.7)	0.7 (0.5)
Variances	Classroom-level intercept	81.5 (15.1)	77.6 (15.9)	64.1 (13.8)	65.0 (14.0)
	Classroom-level gain	3.5 (3.4)	19.6 (5.8)	19.1 (5.7)	18.8 (5.7)
	Student-level intercept	40.5 (7.3)	39.8 (7.3)	39.5 (7.3)	39.7 (7.3)
	Student-level gain	76.7 (6.4)	77.1 (6.5)	77.4 (6.5)	77.3 (6.5)
Predictor × time × condition effects	*p* value	0.6197	0.2417	0.1302	0.1484
	Degrees of freedom	89	89	89	89

Note. Table entries show parameter estimates with standard errors in parentheses except for model probabilities, *p* values, and degrees of freedom. Tier 1 quality indicators were centered at the mean. TEMA-3 = Test of Early Mathematics Achievement–Third Edition; RCMIS = ratings of classroom management and instructional support.

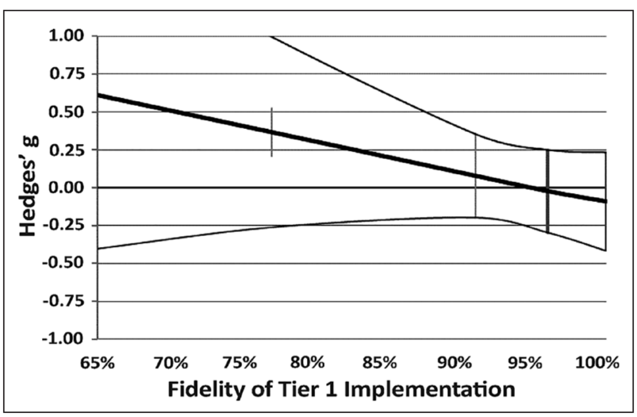

Figure 1. Differential effects of ROOTS on student outcomes across fidelity of Tier 1 implementation scores.

Note. Figures 1–3 depict differential effects of ROOTS on student math achievement as a function of (a) fidelity of Tier 1 implementation (see Figure 1), (B) RCMIS = ratings of classroom and instructional support (see Figure 2), and (C) class size (see Figure 3). The vertical axis shows Hedges' *g* effect sizes for the ROOTS intervention—zero on the vertical axis represents no difference between conditions—and the horizontal axis represents the range of each Tier 1 quality indicator. The heavy line depicts the mean difference between conditions at each level of the Tier 1 quality indicators. The two thinner, outer lines show the 95% confidence bounds around the effect estimate. To show the location of the sample on the graphs, the vertical lines within the confidence bounds depict the median (heavier vertical line), 25th and 75th percentiles (thinner long lines), and the 5th and 95th percentiles (short outer lines). For example, in Figure 1, an x-axis value of about 91% represents the lower 25th sample percentile in Tier 1 fidelity scores.

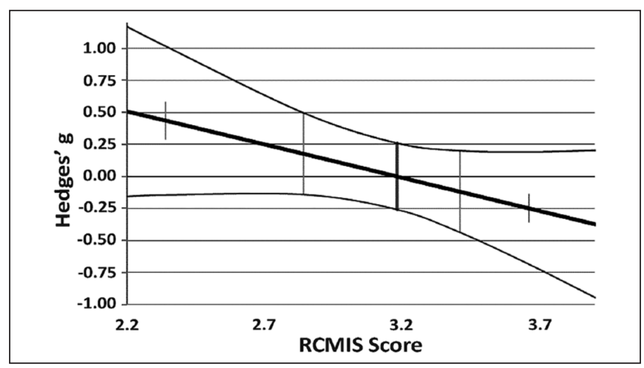

Figure 2. Differential effects of ROOTS on student outcomes across Tier 1 classroom management and instructional support (RCMIS) scores.

Figure 2 shows the ROOTS intervention effect across the range of Tier 1 classroom management and instructional support (RCMIS) scores. The vertical lines show that about 25% of classrooms had classroom management and instructional support scores below 2.8, 50% below 3.2, and 75% below 3.4. The decreasing dark line suggests that the estimated ROOTS effect decreased as Tier 1 classroom management and instructional support scores increased. We estimated effect sizes of 0.18 at the 25th percentile, 0.00 at the median, and −0.12 at the 75th percentile of classroom management and instructional support scores.

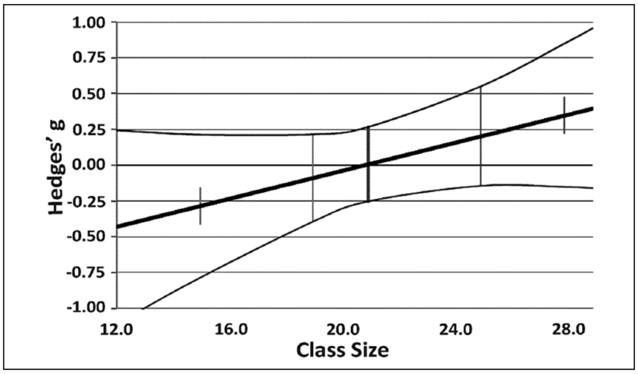

Figure 3. Differential effects of ROOTS on student outcomes across class size.

Figure 3 shows the ROOTS intervention effect across the range of class size (i.e., number of students per classroom). The vertical lines show that about 25% of classrooms had a class size below 19.0, 50% below 21.0, and 75% below 25.0. The increasing dark line suggests that the estimated ROOTS effect increased as class size increased. We estimated effect sizes of −0.09 at the 25th percentile, 0.01 at the median, and 0.20 at the 75th percentile of class size.

Discussion

The purpose of this study was to investigate whether the quality of Tier 1 instruction (i.e., ELM; as measured by four quality indicators) resulted in differential response to a Tier 2 mathematics intervention (ROOTS). We combined data from Years 2 (Oregon) and 3 (Texas) of the ROOTS efficacy trial to examine the following Tier 1 predictors of at-risk students' responsiveness to ROOTS: (a) classroom-level mathematics gains, (b) Tier 1 fidelity of implementation, (c) Tier 1 classroom management and instructional support, and (d) class size. Given the relative dearth of research investigating student responsiveness to research-based interventions in the context of Tier 1 instruction, this study builds on the extant literature by examining the interaction between two tiers of support.

Overall, we found that response to the ROOTS intervention was not significantly predicted by any of the Tier 1 variables that we examined. While our results were not statistically significant, the pattern of results and Hedges' *g* effect sizes indicated that students with higher quality of Tier 1 instruction tended to benefit less from the ROOTS intervention. Specifically, we found that the effect of ROOTS decreased as both Tier 1 fidelity and classroom management and instructional support increased. This pattern was most pronounced when examining classroom management and instructional support, where the effect of ROOTS was 0.18 for students in classrooms at 25th percentile of management and instructional support, and −0.12 in classrooms at the 75th percentile. As hypothesized, the opposite pattern emerged when examining class size, where the effect of ROOTS was −0.09 for students in classrooms at the 25th percentile, and 0.20 for classrooms at the 75th percentile. This pattern of findings was in line with our hypotheses and supports previous research on the role of fidelity (DeFouw et al., 2019; O'Donnell, 2008), classroom management and instructional support (Doabler et al., 2015; Jiménez et al., 2021; Pianta & Hamre, 2009), and class size (Nye et al., 2000).

While these findings are compelling, we must acknowledge several limitations that should be taken into consideration when interpreting our results. First, the ROOTS efficacy study was not sufficiently powered to detect the smaller interaction effects that were of interest in this study, given its design to test the overall efficacy of ROOTS. Future research should investigate the interaction between Tier 1 variables and Tier 2 response as a primary research question, where the number of participants allows for sufficient power to detect more nuanced effects. Second, a ceiling effect occurred on the ELM Fidelity of Implementation Instrument which may have interfered with our ability to detect differences across the continuum of implementation fidelity. The average score on the fidelity measure across both ELM and ELM + ROOTS classrooms was 0.90 (*SD* = 0.1). Teachers' high fidelity to ELM may be a product of the four PD sessions across the school year, the teacher supports built into the program, or the ongoing involvement of the research team across participating sites. We would not expect such high levels of implementation fidelity in typical school practice. In addition, while we conceptualized fidelity as the extent to which ELM was delivered as planned (i.e., adherence to ELM), definitions of fidelity vary, and best practices represent a more multifaceted approach to examining this construct (Harn et al., 2013). Future research should investigate Tier 1 fidelity more comprehensively to allow for a deeper understanding of the role of fidelity in core instruction and its effect on intervention response. Third, while we conceptualized the Tier 1 variables of interest in this study as Tier 1 "quality indicators," other variables may also play an interacting role between Tier 1 instruction and Tier 2 interventions. For example, PD, teacher content knowledge, and student–teacher interactions have all been identified as important contributors to the quality of Tier 1 instruction (Doabler et al., 2015; Garet et al., 2011; Hill et al., 2005; Pianta & Hamre, 2009; Sutherland et al., in press).

Implications and Future Research Directions

Findings of this study highlight the critical importance of implementing high-quality core mathematics instruction in the early grades. Patterns of effects consistently indicated that students who received higher-quality core instruction made smaller gains from the ROOTS intervention. Thus, ensuring high-quality Tier 1 mathematics instruction is central to improving outcomes for students at risk for MD. This may be particularly relevant for students who are on the “cusp” of needing intensive intervention, where remediation may be accomplished through modifications to core instruction. Given the role of Tier 1 and the findings from this study, an important direction for future research is following at-risk students to determine if developing key foundational skills in kindergarten leads to a decrease in the need for Tier 2 supports as well as intensive intervention.

In addition, the instructional context of this study is important when considering these results in the context of typical school practice. Reviews of core mathematics curricula have reported that widely used Tier 1 programs rarely incorporate instructional supports and design features aligned with the needs of at-risk learners (Bryant et al., 2008; Doabler et al., 2012). In contrast, ELM was specifically designed to support at-risk learners (Clarke et al., 2011), and includes evidence-based design features linked to improved outcomes for students with MD (e.g., Gersten et al., 2009; NMAP, 2008). It is therefore possible that our findings would have differed if a more typical core program less tailored to the needs of at-risk learners was in place.

Of equal importance is the consideration of alignment between the ELM core program and ROOTS intervention (Hill et al., 2012). Alignment may be conceptualized as *procedural* (e.g., intervention corresponds to what is taught in Tier 1 on a given day), *instructional* (e.g., specific teaching strategies and materials are aligned), and *philosophical* (e.g., general instructional approaches are aligned; Walp & Walmsley, 1989). While ELM and ROOTS are strongly aligned in their use of mathematical representations, mathematical content, and instructional approaches (i.e., *instructional* and *philosophical* alignment), they are not *procedurally* aligned. Research indicates that the level of alignment between the core and intervention program may impact student response to the Tier 2 intervention (Fuchs et al., 2008). Future research should further unpack the relations between program alignment and Tier 2 intervention effectiveness within the context of other Tier 1 quality indicators.

Conclusion

While results from this study were not statistically significant, findings point to the potential of high-quality Tier 1 instruction to mitigate the need for more intensive interventions. In addition, while class size is largely an unalterable variable, our findings suggest that smaller class sizes may decrease the need for Tier 2 intervention. Teachers may combat larger class sizes by intensifying aspects of Tier 1 or strategically differentiating supports to better meet the needs of students with MD. With time, personnel, and financial resources in short supply, schools face difficult decisions about how to allocate resources effectively to remediate MDs (Ochsendorf, 2016). Questions regarding what content should be prioritized, when, and with what degree of alignment with core instructional content and delivery are essential for efforts to match services to student needs, a central component of MTSS. Research that evaluates math intervention within a multitier context may ultimately provide guidance to help schools use resources more effectively. As much as possible, we recommend that intervention researchers capture what is occurring in Tier 1 to allow for more nuanced investigations into the role of core instruction and whether it may enhance or eliminate the need for more intensive interventions. This may lead to insights regarding situations in which less-intensive alternatives to intervention, such as a supplemental class-wide “intervention” for all students, are a better fit given available school resources. Continued research in this vein is needed to better understand the conditions under which intensive interventions are likely to be effective.

Declaration of Conflicting Interests

B.C. and C.T.D. are eligible to receive a portion of royalties from the University of Oregon’s distribution and licensing of certain ROOTS-based works. Potential conflicts of interest are managed through the University of Oregon’s Research Compliance Services. An independent external evaluator and coauthor of this publication completed the research analysis described in the article.

Funding

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education (grant nos. R305K040081 and R305A080699) to the Center on Teaching and Learning at the University of Oregon and Pacific Institutes for Research. The opinions expressed are those of the authors and do not represent views of the institute or the U.S. Department of Education.

ORCID iD

Marah Sutherland https://orcid.org/0000-0002-6108-8515

References

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders (NCEE 2011-4001)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Agrawal, J., & Morin, L. L. (2016). Evidence-based practices: Applications of concrete representational abstract framework across math concepts for students with mathematics disabilities. *Learning Disabilities Research & Practice*, *31*(1), 34–44. http://dx.doi.org/10.1111/ldrp.12093

Al Otaiba, S., Wagner, R. K., & Miller, B. (2014). “Waiting to fail” redux: Understanding inadequate response to intervention. *Learning Disability Quarterly*, *37*(3), 129–133. http://dx.doi.org/10.1177/0731948714525622

Bailey, D. H., Fuchs, L. S., Gilbert, J. K., Geary, D. C., & Fuchs, D. (2020). Prevention: Necessary but insufficient? A 2-year follow-up of an effective first-grade mathematics intervention. *Child Development*, *91*(2), 382–400. http://dx.doi.org/10.1111/cdev.13175

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE 2016-4000). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Berkeley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). A snapshot of progress toward implementation of responsiveness to intervention (RTI) throughout the United States. *Journal of Learning Disabilities*, *42*(1), 85–95. http://dx.doi.org/10.1177/0022219408326214

Berkeley, S., Scanlon, D., Bailey, T. R., Sutton, J. C., & Sacco, D. M. (2020). A snapshot of RTI implementation a decade later: New picture, same story. *Journal of Learning Disabilities*, *53*(5), 332–342. http://dx.doi.org/10.1177/0022219420915867

Bryant, B. R., Bryant, D. P., Kethley, C., Kim, S. A., Pool, C., & Seo, Y. J. (2008). Preventing mathematics difficulties in the primary grades: The critical features of instruction in textbooks as part of the equation. *Learning Disability Quarterly*, *31*(1), 21–35. http://dx.doi.org/10.2307/30035523

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35. http://dx.doi.org/10.1007/s00265-010-1029-6

Clarke, B., Baker, S., Smolkowski, K., Doabler, C. T., Strand Cary, M., & Fien, H. (2015). Investigating the efficacy of a core kindergarten mathematics curriculum to improve student mathematics learning outcomes. *Journal of Research on Educational Effectiveness*, *8*, 303–324. http://dx.doi.org/10.1080/19345747.2015.1116034

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Strand Cary, M. (2016). Examining the efficacy of a Tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities*, *49*(2), 152–165. http://dx.doi.org/10.1177/0022219414538514

Clarke, B., Smolkowski, K., Baker, S. K., Fien, H., Doabler, C. T., & Chard, D. (2011). The impact of a comprehensive Tier 1 core kindergarten program on the achievement of students at risk in mathematics. *The Elementary School Journal*, *111*(4), 561–584. http://dx.doi.org/10.1086/659033

Clarke, B., Turtura, J., Lesner, T., Cook, M., Smolkowski, K., Kosty, D., & Doabler, C. T. (2022). A conceptual replication of a kindergarten math intervention within the context of a research-based core. *Exceptional Children*. Advance online publication. http://dx.doi.org/10.1177/00144029221088938

Coyne, M. D., Oldham, A., Dougherty, S. M., Leonard, K., Koriakin, T., Gage, N. A., Burns, D., & Gillies, M. (2018). Evaluating the effects of supplemental reading intervention within an MTSS or RTI reading reform initiative using a regression discontinuity design. *Exceptional Children*, *84*(94), 350–367. http://dx.doi.org/10.1177/0014402918772791

DeFouw, E. R., Codding, R. S., Collier-Meek, M. A., & Gould, K. M. (2019). Examining dimensions of treatment intensity and treatment fidelity in mathematics intervention research for students at risk. *Remedial and Special Education*, *40*(5), 298–312. http://dx.doi.org/10.1177/0741932518774801

Doabler, C. T., Baker, S. K., Kosty, D. B., Smolkowski, K., Clarke, B., Miller, S. J., & Fien, H. (2015). Examining the association between explicit mathematics instruction and student mathematics achievement. *The Elementary School Journal*, *115*(3), 303–333. http://dx.doi.org/10.1086/679969

Doabler, C. T., Fien, H., Nelson-Walker, N. J., & Baker, S. K. (2012). Evaluating three elementary mathematics programs for presence of eight research-based instructional design principles. *Learning Disability Quarterly*, *35*(4), 200–211. http://dx.doi.org/10.1177/0731948712438557

Doabler, C. T., & Nelson-Walker, N. J. (2009). *Ratings of classroom management and instructional support* [Unpublished observation instrument]. Center on Teaching and Learning, College of Education, University of Oregon.

Doabler, C. T., Stoolmiller, M., Kennedy, P. C., Nelson, N. J., Clarke, B., Gearin, B., Fien, H., Smolkowski, K., & Baker, S. K. (2019). Do components of explicit instruction explain the differential effectiveness of a core mathematics program for kindergarten students with mathematics difficulties? A mediated moderation analysis. *Assessment for Effective Intervention*, *44*(3), 197–211. http://dx.doi.org/10.1177/1534508418758364

Fuchs, D., & Fuchs, L. S. (2017). Critique of the national evaluation of response to intervention: A case for simpler frameworks. *Exceptional Children*, *83*(3), 255–268. http://dx.doi.org/10.1177/0014402917693580

Fuchs, D., & Fuchs, L. S. (2019). On the importance of moderator analysis in intervention research: An introduction to the special issue. *Exceptional Children*, *85*(92), 126–128. http://dx.doi.org/10.1177/0014402918811924

Fuchs, L. S., Fuchs, D., Craddock, C., Hollenbeck, K. N., Hamlett, C. L., & Schatschneider, C. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students’ math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology*, *100*(3), 491–509. http://dx.doi.org/10.1037/0022-0663.100.3.491

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011-4024). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools. *NCEE* 2009-4060. What Works Clearinghouse. https://eric.ed.gov/?id=ED504995

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*(2), 149–164. http://dx.doi.org/10.1177/001440290507100202

Gersten, R., Jayanthi, M., & Dimino, J. (2017). Too much, too soon? Unanswered questions from national response to intervention evaluation. *Exceptional Children*, *83*(3), 244–254. http://dx.doi.org/10.1177/0014402917692847

Ginsburg, H., & Baroody, A. (2003). *Test of early mathematics ability-Third edition*. Pro-Ed.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. http://dx.doi.org/10.1146/annurev.psych.58.110405.085530

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, p-values, confidence intervals, and power: A guide to misinterpretations. *The American Statistician*, *70*(2, supplement to Wasserste & Lazar), 1–12. http://dx.doi.org/10.1007/s10654-016-0149-3

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children*, *79*(2), 181–193. http://dx.doi.org/10.1177/001440291307900205l

Hill, D. R., King, S. A., Lemons, C. J., & Partanen, J. N. (2012). Fidelity of implementation and instructional alignment in response to intervention research. *Learning Disabilities Research & Practice*, *27*(3), 116–124. http://dx.doi.org/10.1111/j.1540-5826.2012.00357.x

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. http://dx.doi.org/10.3102/00028312042002371

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). SAGE.

Jiménez, J. E., Gutiérrez, N., & de León, S. C. (2021). Analyzing the role of fidelity of RTI Tier 2 reading intervention in Spanish kindergarten and first grade students. *Annals of Dyslexia*, *71*(1), 28–49. http://dx.doi.org/10.1007/s11881-021-00221-5

Jitendra, A. K., Alghamdi, A., Edmunds, R., McKevett, N. M., Muanoutoua, J., & Roesslein, R. (2021). The effects of tier 2 mathematics interventions for students with mathematics difficulties: A meta-analysis. *Exceptional Children*, *87*(3), 307–325. http://dx.doi.org/10.1177/0014402920969187

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, *39*(2), 181–195. https://doi.org/10.1080/02796015.2010.12087772

Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools*, *49*(4), 342–352. http://dx.doi.org/10.1002/pits.21602

Kratchowill, T. R., Volpiansky, P., Clements, M., & Ball, C. (2007). Professional development in implementing and sustaining multitier prevention models: Implications for response to intervention. *School Psychology Review*, *36*(4), 618–631. http://dx.doi.org/10.1080/02796015.2007.12087921

Lam, E. A., & McMaster, K. L. (2014). Predictors of responsiveness to early literacy intervention: A 10-year update. *Learning Disability Quarterly*, *37*(3), 134–147. http://dx.doi.org/10.1177/0731948714529772

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). John Wiley & Sons.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*, 247–266. http://dx.doi.org/10.1016/0272-7358(91)90103-2

Murray, D. (1998). *Design and analysis of group-randomized trials*. Oxford University Press.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Nellis, L. M. (2012). Maximizing the effectiveness of building teams in response to intervention implementation. *Psychology in the Schools*, *49*(3), 245–256. http://dx.doi.org/10.1002/pits.21594

Nelson, P. M., Normal, E. R., Parker, D. C., & Cormier, D. C. (2019). An examination of interventionist implementation fidelity and content knowledge as predictors of math intervention effectiveness. *Journal of Applied School Psychology*, *35*(3), 234–256. http://dx.doi.org/10.1080/15377903.2019.1568334

Nye, B., Hedges, L. V., & Konstantantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee Class Size Experiment. *American Educational Research Journal*, *37*(1), 123–151. http://dx.doi.org/10.3102/00028312037001123

Ochsendorf, R. (2016). Advancing understanding of mathematics development and intervention: Findings from NCSER-funded efficacy studies. *Journal of Research on Educational Effectiveness*, *9*(4), 570–576. http://dx.doi.org/10.1080/19345747.2016.1222144

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. http://dx.doi.org/10.3102/0034654307313793

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*(2), 109–119. http://dx.doi.org/10.3102/0013189X09332374

Pong, S., & Pallas, A. (2001). Class size and eighth-grade math achievement in the United States and abroad. *Educational Evaluation and Policy Analysis*, *23*(3), 251–273. http://dx.doi.org/10.3102/01623737023003251

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437–448.

SAS Institute. (2016). *Base SAS® 9.4 procedures guide* (6th ed., Product Documentation). http://support.sas.com/documentation/index.html

Scott, T. M., Gage, N. A., Hirn, R. G., Lingo, A. S., & Burt, J. (2019). An examination of the association between MTSS implementation fidelity measures and student outcomes. *Preventing School Failure: Alternative Education for Children and Youth*, *63*(4), 308–316. http://dx.doi.org/10.1080/1045988X.2019.1605971

Sood, S., & Jitendra, A. K. (2011). An exploratory study of a number sense program to develop kindergarten students' number proficiency. *Journal of Learning Disabilities*, *46*(4), 328–346. http://dx.doi.org/10.1177/0022219411422380

Stelfox, H. T., & Straus, S. E. (2013). Measuring quality of care: Considering measurement frameworks and needs assessment to guide quality indicator development. *Journal of Clinical Epidemiology*, *66*(12), 1320–1327. http://dx.doi.org/10.1016/j.jclinepi.2013.05.018

Sutherland, M., Clarke, B., Kosty, D. B., Baker, S. K., Doabler, C. T., Smolkowski, K., Fien, H., & Goode, J. (in press). Investigating the interaction between teacher mathematics content knowledge and curriculum on instructional behaviors and student achievement. *The Elementary School Journal*.

van Belle, G. (2008). *Statistical rules of thumb* (2nd ed.). John Wiley & Sons.

Walp, T. P., & Walmsley, S. A. (1989). Instructional and philosophical congruence: Neglected aspects of coordination. *The Reading Teacher*, *42*(6), 364–368.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133.