Exceptional Children
2019, Vol. 85(2) 129–146

DOI: 10.1177/0014402918799503
journals.sagepub.com/home/ecx
SAGE

Exploring the Relationship Between Initial Mathematics Skill and a Kindergarten Mathematics Intervention

Ben Clarke[1], Christian T. Doabler[2], Keith Smolkowski[3], Jessica Turtura[1], Derek Kosty[3], Evangeline Kurtz-Nelson[1], Hank Fien[1], and Scott K. Baker[4]

Abstract

This study examined the role of initial skill in moderating intervention effects of a 50-lesson mathematics intervention program, ROOTS, for at-risk kindergarten students focused on developing whole-number concepts and skills. The study utilized a randomized block design with at-risk students (n = 592) within classrooms (n = 60) randomly assigned to one of two treatment conditions (a small group of two to five students) or control condition. Proximal and distal measures were collected in the fall (pretest), spring (posttest), and winter of first grade (delayed posttest). Analyses examined the moderating effects of initial student achievement level on mathematics outcomes. Results indicated that initial skill moderated student outcomes but the relationship did not differ by group size. Implications for tiered mathematics instruction are discussed.

Concern with the mathematics thinking and knowledge of students in the U.S. educational system has been a prominent issue extending back to the start of the century (National Research Council, 2001). Driven partly by concerns for international competitiveness and forecasted growth within mathematics-related fields (National Science Board, 2008), efforts to improve mathematics instruction have garnered increased national attention (National Mathematics Advisory Panel, 2008) and resulted in greater expectations for all students (Common Core State Standards Initiative, 2010). To some extent, those efforts have been successful, with relatively consistent and sustained improvements on national assessments of mathematics achievement (National Center for Education Statistics, 2015). However, overall levels of proficiency remain low, with persistent gaps remaining for minority students, students with disabilities, and English language learners.

Given the current state of mathematics achievement and an understanding that long-term mathematics difficulties begin early (Jordan, Kaplan, & Hannich, 2002) and are relatively intractable (Duncan et al., 2007; Morgan, Hillemeier, Farkas, & Macuga, 2014), there has been increasing interest in ensuring that early elementary students are exposed to and learn foundational mathematics content (Frye et al., 2013). To that end,

[1]University of Oregon
[2]University of Texas at Austin
[3]Oregon Research Institute
[4]Southern Methodist University

Corresponding Author:
Ben Clarke, University of Oregon, 1600 Millrace Dr., Suite 108, Eugene, OR 97403.
E-mail: clarkeb@uoregon.edu

several intervention programs have been developed and studied focused on transitioning students' early understanding of number (Berch, 2005; Gersten & Chard, 1999) to formal school mathematics. As called for by experts in the field (Gersten et al., 2009), intervention programs have focused on building an understanding of whole-number content and the use of systematic and explicit instructional design and delivery principles (Archer & Hughes, 2011). The result has been an emerging and growing research base on intervention programs targeting kindergarten whole-number content (e.g., Clarke, Doabler, Smolkowski, Baker, et al., 2016; Dyson, Jordan, & Glutting, 2013). Across these programs, results showed generally positive significant affects on student mathematics outcomes, demonstrating that targeted intervention can be effective in teaching students essential whole-number concepts and skills.

Understanding Response Variation

While significant advances have been made in the development and validation of early mathematics intervention programs, challenges remain in ensuring that the learning needs of all students are met. A range of studies have consistently documented that between 5% and 10% of the school-age population is classified as having persistent low achievement in mathematics (e.g., Geary, 2011). For this set of students in particular, the need for effective intervention is paramount. However, emerging evidence suggests that not all students will respond to a generally effective intervention program. For example, L. S. Fuchs, Fuchs, and Compton (2012), summarizing a program of research on the efficacy of mathematics tutoring, estimated that the modal rate of unresponsiveness to well-designed mathematics interventions that were implemented with fidelity was approximately 4% of all students. Within a response-to-intervention model (RTI) or multitier-system-of-support (MTSS) model of service delivery, it is assumed that as students exhibit nonresponse, instruction is altered to provide a more targeted intensive experience (National Center on Response to Intervention, 2010).

One proposed mechanism to accomplish this goal is to gain a better understanding of what student-level variables, including academic, cognitive, and behavioral, are associated with unresponsiveness to generally efficacious Tier-2 interventions (Miller, Vaughn, & Freund, 2014). Such insight might enable the deployment of more efficient screening systems and serve to inform the design and delivery of instruction that meets the needs of students through customized or more intensive interventions to address the specific at-risk student's needs (L. S. Fuchs, Fuchs, & Compton, 2013). Given that schools often struggle to implement RTI (Balu et al., 2015) as theoretically conceptualized, including implementing all three tiers of RTI systems with fidelity, schools may benefit from more straightforward systems of RTI or MTSS that allow for students to be screened directly into intensive, customized intervention (D. Fuchs & Fuchs, 2017).

One proposed mechanism to accomplish this goal is to gain a better understanding of what student-level variables, including academic, cognitive, and behavioral, are associated with unresponsiveness to generally efficacious Tier-2 interventions (Miller, Vaughn, & Freund, 2014).

Research in reading has shown that students' responsiveness to intervention can be predicted by performance on measures of specific reading-related domains (i.e., word identification, alphabetic principle, fluency, and phonemic awareness) prior to intervention (Lam & McMaster, 2014). In recent years, researchers have also started to examine student characteristics associated with lack of responsiveness to intervention in mathematics; however, a paucity of research remains in this area. Toll and Van Luit (2013) examined the effects of a kindergarten mathematics intervention, looking specifically at effects for subgroups of students with low (between the 25th and 50th percentiles) and very low (below the 25th percentile) early numeracy skills. Analyses showed that the intervention was effective only for students who scored between the 25th and 50th percentiles on pretest measures of early numeracy and was not beneficial for students with very low early number ability (<25th percentile) at the start of the intervention. Similarly, L. S. Fuchs, Sterba, Fuchs, and Malone (2016) examined whether responsiveness to a fraction intervention was associated with start-of-fourth-grade whole-number calculation skill. The authors did not find a significant moderator effect on posttest fraction understanding or calculation performance. In contrast, L. S. Fuchs, Malone, et al. (2016) determined that responsiveness to fraction word-problem intervention was associated with start-of-fourth-grade reasoning ability. Although most students responded adequately to the intervention, those with very low reasoning scores did not.

Ultimately, this body of research suggests that a greater understanding of student-level characteristics predictive of responsiveness to intervention can allow for more efficient and effective procedures for screening students. There is a small body of research that suggests screening students into more intensive interventions shows promise in creating more dynamic RTI models by placing students with the weakest initial skill in areas predictive of responsiveness directly into intensive, Tier-3 interventions rather than "waiting to fail" in Tier 2 (Lam & McMaster, 2014). Al Otaiba and colleagues (2014) investigated a variation of the typical RTI model by utilizing screeners to immediately place students in a more intensive tiered intervention experience and found significantly higher reading outcomes. Besides utilizing screeners to fast-track students to appropriate interventions, another potential mechanism to increase intervention intensity is by reducing the size of instructional groups (Baker, Fien, & Baker, 2010). A meta-analysis examining group size in reading found larger effect sizes for smaller groups (Wanzek & Vaughn, 2007), and research in which instructional content was held constant found that smaller groups (1:1 or 1:3 teacher-student ratio) had greater effects when compared to a larger smaller group (1:10 teacher-student ratio). However, results also indicated that the two smaller small groups (1:1, 1:3) did not differ from each other (Vaughn, Thompson, Kouzekanani, and Dickson, 2003). Few studies in mathematics have manipulated group size as an independent variable to examine its effects on learning. For example, Bryant et al. (2016) examined the effects of a systematic, explicit, strategic Tier-3 intervention on the mathematics performance of students in the second grade who had identified severe mathematics difficulties. As compared to the author's Tier-2 intervention, the Tier-3 intervention was intensified in dosage (every day), group size (1:2 or 1:3), and the use of explicit instructional practices. The majority (75%) of students responded to the intensified intervention and qualified to exit at the end of treatment. However, this manipulation was across studies rather than within a study and included multiple elements in addition to a manipulation of group size.

Few studies in mathematics have manipulated group size as an independent variable to examine its effects on learning.

Clarke et al. (2017) investigated similar questions related to the general efficacy of a Tier-2 early mathematics kindergarten intervention program, ROOTS, and a treatment manipulation of group size. Blocking on classroom, at-risk kindergarten students ($n = 10$) were randomly assigned to one of three conditions: ROOTS delivered in a 2:1 student-to-teacher-ratio small group, ROOTS delivered in a 5:1 student-to-teacher-ratio small group, or control ($n = 3$). Overall, results indicated significant positive effects on three of six outcome measures, favoring ROOTS students over controls. However, no significant difference was found between ROOTS small-group conditions (2:1 and 5:1 student-teacher ratio).

Purpose and Research Questions

While the field's understanding of factors related to intervention response is growing, continued research is needed to expand our understanding of specific student-level predictors of responsiveness, such as initial skill, and how initial skill status interacts with approaches, such as modifying group size, to increase instructional intensity. The purpose of this study was to expand previous work by conducting secondary analyses of the data from Clarke et al. (2017), which investigated the efficacy of a Tier-2 kindergarten mathematics intervention, with a focus on two specific research questions:

1. Did students benefit differentially from the ROOTS intervention by initial early mathematics skill?
2. Did students benefit differentially from the two treatment conditions, large group versus small group, by initial mathematics skill?

Our previous research in early mathematics (Clarke et al., 2015) indicated a greater effect for a kindergarten core mathematics curriculum for students with lower initial skills, and a previous investigation of the ROOTS curriculum indicated no difference in student outcomes by group size (Clarke et al., 2017). Given these results, we hypothesized that there may be differential effects of the ROOTS intervention based on initial skill, with a stronger overall effect for students who began lower in initial skill, but we did not expect the same differential effect by group size. To contextualize overall results, we also explored the ability of ROOTS to help close the achievement gap.

Method

This study analyzed data collected from two cohorts of the federally-funded ROOTS Efficacy Project (Clarke, Doabler, Fien, Baker, & Smolkowski, 2012). A partially nested randomized controlled trial was employed (Baldwin, Bauer, Stice, & Rohde, 2011), randomly assigning kindergarten students within classrooms to one of three conditions: (2:1 ROOTS group, 5:1 ROOTS group, and a no-treatment control condition). Full methods for this study were previously published (Clarke et al., 2017).

Participants

Schools. Fourteen elementary schools from four Oregon school districts participated in the study. Within the 14 schools, 0% to 12% of students were American Indian or Native Alaskan, 0% to 16% were Asian, 0% to 9% were Black, 0% to 74% were Hispanic, 0% to 2% were Native Hawaiian or Pacific Islander, 19% to 92% were White, and 0% to 15% were more than one race. Eight percent to 25% of students received special education services, 5% to 69% were English language learners, and 17% to 87% were eligible for free or reduced lunch.

Classrooms. A total of 69 classrooms participated in the study (n = 37 classrooms in Year 1; n = 32 classrooms in Year 2). Classrooms had an average of 25.06 students (SD = 5.60). The 69 classrooms were taught by 31 teachers. Twenty teachers participated in both years of the study. Nine Year 1 teachers and seven Year 2 teachers taught two participating half-day (morning and afternoon) classrooms. Of the 31 teachers, 100% identified as female, 84% as White, and 10% as Asian American/Pacific Islander. Teachers averaged 16.45 years of teaching experience and 8.81 years of kindergarten teaching experience.

Criteria for participation. In each participating classroom, all students with parental consent were screened in the late fall of their kindergarten year. The screening process included the Assessing Student Proficiency in Early Number Sense (ASPENS; Clarke, Gersten, Dimino, & Rolfhus, 2011) and the Number Sense Brief Screener (NSB; Jordan, Glutting, & Ramineni, 2008). Students were eligible for the ROOTS intervention if they received an NSB score of 20 or less and an ASPENS composite score in the strategic or intensive ranges. After being determined eligible for the ROOTS intervention, students' NSB and ASPENS scores were separately converted into standard scores with the full sample and then combined to form an overall composite score for each at-risk student. All data management was conducted by the project's independent evaluator. Composite scores within each classroom were then rank ordered, and the 10 ROOTS-eligible students with the lowest composite scores were randomly assigned to one of three conditions: (a) 2:1 ROOTS group, (b) 5:1 ROOTS group, or (c) no-treatment control condition. Out of the 69 participating classrooms, 53 had at least 10 students who met ROOTS eligibility criteria. Fourteen classrooms in Year 1 and two classrooms in Year 2 had fewer than 10 ROOTS-eligible students, and in these instances, classrooms were combined to create virtual ROOTS "classrooms." After these procedures were applied, a total of 60 ROOTS classrooms participated in this study.

Students. A total of 1,550 kindergarten students were screened for ROOTS eligibility. Of these students, 592 met eligibility criteria and were randomly assigned within each of the 60 classrooms to the two-student group condition ($n = 120$), the five-student group condition ($n = 295$), or the no-treatment control condition ($n = 177$).

Interventionists. ROOTS intervention groups were taught by instructional assistants employed by the district and by interventionists hired specifically for this study. Eighty nine percent identified as female, 93% as White, 4% as Hispanic, and 2% as another ethnicity. Interventionists had an average of 8 years of teaching experience, previous experience providing small-group instruction (93%), and a bachelor's degree or higher (58%), and 63% had taken a college-level algebra course.

ROOTS. ROOTS is a 50-lesson, Tier-2 mathematics program designed to build students' proficiency in whole-number concepts and skills. The ROOTS intervention was delivered in 20-min small-group sessions (2:1 or 5:1) 5 days per week for approximately 10 weeks. Instruction for all students began in the late fall and ended in the spring, and this start date was selected to provide students with the opportunity to respond to initial core mathematics instruction and to therefore minimize the identification of typically achieving students. ROOTS was designed to supplement core mathematics instruction and thus was delivered at times that did not conflict with students' core instruction in mathematics.

ROOTS instruction is aligned with Common Core State Standards for Mathematics (CCSS-M; Common Core State Standards Initiative, 2010) and recommendations from expert panels to focus on whole-number concepts and skills (Gersten et al., 2009). Specifically, ROOTS instruction emphasizes concepts from the Counting and Cardinality and Operations and Algebraic Thinking domains of the CCSS-M. The ROOTS instructional approach is drawn from principles of explicit and systematic mathematics instruction (Gersten et al., 2009), including explicit teacher modeling, deliberate practice, visual representations of mathematics, and academic feedback. Frequent opportunities for students to verbalize their mathematical thinking and discuss problem-solving methods are also embedded throughout the program's lessons.

Implementation fidelity. Each ROOTS group was observed three times during the course of the intervention. On a 4-point scale (4 = *all*, 3 = *most*, 2 = *some*, 1 = *none*), observers rated the extent to which the interventionist (a) met the lesson's instructional objectives, (b) followed the provided teacher scripting, and (c) used the prescribed mathematics models for that lesson. Observers also recorded whether the interventionist taught the number of activities prescribed in the lesson. Observations indicated that interventionists delivered prescribed activities as specified ($M = 4.03$ out of 5 activities, $SD = 0.87$). Observations indicated that interventionists met mathematics objectives ($M = 3.43$, $SD = 0.74$), followed teacher scripting ($M = 3.20$, $SD = 0.77$), and used prescribed mathematics models ($M = 3.58$, $SD = 0.67$). Intraclass correlation coefficients (ICCs) were calculated across observers, with ICCs for individual fidelity ratings indicating moderate to nearly perfect agreement: 0.92 for number of activities delivered, 0.72 for met mathematics objectives, 0.72 for followed teacher scripting, and 0.59 for used prescribed mathematics models. Landis and Koch (1977) characterize ICCs of 0.41 to 0.60 as moderate, 0.61 to 0.80 as substantial, and 0.81 to 1.00 as nearly perfect.

Control Condition

Core (Tier-1) mathematics instruction served as the control condition in this study, as both treatment and control students continued to receive their daily core mathematics instruction. For treatment students, ROOTS instruction was provided in addition to core mathematics instruction. The control condition was documented through teacher surveys and direct observations of classroom instruction. Teachers reported that they used a variety of published mathematics curricula during their mathematics instruction, primarily Scott Foresman, enVisionmath, Houghton Mifflin, and Everyday Mathematics, and supplemented these core curricula with teacher-created materials. Teachers reported that they provided approximately 31(31.32) min of daily mathematics instruction (SD = 9.88). Survey data also identified that all teachers included mathematics topics during calendar time. All teachers reported that they provided whole-group and teacher-led mathematics instruction. Majorities of teachers reported that they provided opportunities for peer or group work, independent student work, and mathematics centers. Information about the control condition was also gathered from direct observations of core mathematics instruction by trained project staff. No evidence of treatment diffusion during core mathematics instruction was identified.

Outcome Measures

Students were administered five measures of whole-number sense at pretest (T1) and posttest (T2). These measures included a proximal assessment of whole-number understanding that measured skills taught during ROOTS, two distal measures of whole-number sense, and a set of curriculum-based measures of discrete early number sense skills. A distal outcome measure was administered 6 months into students' first-grade year (T3). Interscorer reliability criteria of .95 or above were met for all assessments.

ROOTS Assessment of Early Numeracy Skills (RAENS; Doabler, Clarke, & Fien, 2012).

RAENS is a researcher-developed measure that includes 32 items assessing aspects of counting and cardinality, number operations, and the base-10 system. RAENS' predictive validity ranges from .68 to .83 for the Test of Early Mathematics Ability–Third Edition (TEMA-3) and the NSB (Clarke, Doabler, Smolkowski, Kurtz Nelson, et al., 2016) and interrater scoring agreement reported at 100% (Clarke, Doabler, Smolkowski, Baker et al., 2016).

Oral Counting–Early Numeracy Curriculum-Based Measurement (Clarke & Shinn, 2004). This curriculum-based measure has students orally count in English for 1 min. Test-retest reliability and alternate-form reliability are reported at above .80, and predictive validity with standardized measures of mathematics ranges from .46 to .72.

ASPENS (Clarke et al., 2011). This set of three 1-min curriculum-based measures focuses on numeral identification, comparing quantities, and strategic counting. Test-retest reliabilities of kindergarten ASPENS measures are in the moderate to high range (.74 to .85). Predictive validity from fall to spring scores on the TerraNova 3 is reported as ranging from .45 to .52.

NSB (Jordan et al., 2008). The NSB is an individually administered measure with 33 items covering a range of early numeracy skills, including counting knowledge and principles, number recognition, number comparisons, nonverbal calculation, story problems, and number combinations. The NSB has a coefficient alpha of .84.

TEMA-3 (Ginsburg & Baroody, 2003). This individually administered measure of early mathematical ability assesses whole-number understanding for students ranging in age from 3 years to 8 years 11 months. Alternate-form and test-retest reliabilities of the TEMA-3 are .97 and .93, respectively. The TEMA-3 has concurrent validity with other mathematics measures ranging from .54 to .91.

The Stanford Achievement Test–Tenth Edition (SAT-10; Harcourt Educational Measurement, 2002) and the Stanford Early School Achievement Test (SESAT). The SAT-10 and SESAT are group-administered standardized achievement tests with two mathematics subtests, Problem Solving and Procedures. Both measures are multiple choice and have two mathematics subtests: Problem Solving and Procedures. The SESAT is administered in the kindergarten year and the SAT-10 in first grade. The SAT-10 is a standardized achievement test with adequate and well-reported validity (r = .67) and reliability (r = .93). All treatment and control students were administered the SESAT at posttest (T_2) and the SAT-10 midway through their first-grade year (T_3).

Statistical Analysis

We conducted two sets of analyses to address our research questions about differential response to ROOTS and the group size based on initial TEMA-3 scores. Previously, we examined overall effects of the ROOTS intervention on mathematics achievement and the overall effects of group size (Clarke et al., 2017) with an analysis designed to account for students partially nested within small groups (Baldwin et al., 2011). Because the ROOTS groups, but not the unclustered controls, required a group-level variance, the analyses accounted for the potential heterogeneity among variances across conditions (Roberts & Roberts, 2005).

In this study, we examined whether initial mathematics achievement based on TEMA-3 scores predicted differential response to the ROOTS intervention or to the different group sizes. We expanded the statistical model to include the pretest TEMA-3 as a predictor of differential response and its interaction with the condition effect, either ROOTS versus control or small versus large groups. For the analysis of the group size condition, the models used a standard analysis for nested data rather than the partially nested models. For this question, we included only students who were nested in small groups.

For the tests of ROOTS versus control with partially nested data, the residual variances may have differed between conditions. We therefore tested whether the homoscedastic and heteroscedastic models could be assumed equivalent with a likelihood ratio test and reported the simpler model if we were able to accept the equivalence of the two models. For the test of variances (only), we followed the logic of noninferiority trials and reversed the null and alternative hypotheses and the associated Type I and Type II error rates (Dasgupta, Lawson, & Wilson, 2010). For this reason, and because tests of variance structures have limited power (Kromrey & Dickinson, 1996), we set $\alpha = .20$ as our Type I error rate and reported the more complex heteroscedastic model unless we were relatively certain the two variances were equivalent.

A moderated effect implies that the condition difference depends on the moderator. For a statistically significant moderator, the difference between the two conditions will be larger (or smaller) at higher (or lower) levels of the moderator. To detect these differences, we estimated the difference between conditions and confidence bounds at multiple points along the moderator (Jaccard & Turrisi, 2003). We used these estimates to compute the regions of statistical significance based on the confidence intervals and graphed the results with the method recommended by Preacher, Curran, and Bauer (2006) for interpretation. The graphs depicted the condition effect and its 95% confidence intervals across the range of moderator scores. For the regions in which both confidence bounds exclude zero value for condition differences, we interpret that as a statistically significant condition difference across that region.

We tested an additional set of mixed models that extended those discussed above to account for students clustered within classrooms. Results were similar in both sets of models and condition effects did not vary by classroom, so we omitted the results.

We fit the statistical models to our data using SAS PROC MIXED Version 14.2 (SAS Institute, 2016) with restricted 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 (Schafer & Graham, 2002). In the present study, the missing data do 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) or that missingness is detrimental to the internal validity. The majority of missing data involved students who were absent on the day of assessment (e.g., due to illness) or transferred to a new school (e.g., due to their families moving). Nonrandom missingness, however, "is often not sufficient to affect the internal validity of an experimental study to any practical extent" (Graham, 2009, p. 568).

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. Multilevel regression methods are also quite robust to violations of normality (e.g., Hannan & Murray, 1996). We also corrected for multiple tests with the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) and reported the original p values as well as the Benjamini-Hochberg adjusted p values for each outcome. We adjusted p values separately within the set of analyses for each research question.

Results

Demographic characteristics, the potential for differential attrition, and main effects for ROOTS intervention and group sizes were presented in Clarke et al. (2017).

Predictors of Differential Response to ROOTS

Clarke et al. (2017) found statistically significant condition differences—main effects—for the ASPENS, TEMA-3, and RAENS. Tables 1 and 2 present tests of differential response to ROOTS as a function of pretest TEMA-3 scores. The bottom two rows of the tables show the likelihood ratio test results that compared homoscedastic residuals with heteroscedastic residuals, and the tables report a different number of variances depending on the results. The homoscedastic model, which assumed equivalent residual and pre-/posttest covariance estimates between conditions, fit the data for Oral Counting, TEMA-3, and SESAT.The data fit the heteroscedastic model ($p < .20$) for NSB, SESAT, and SAT-10.

The TEMA-3 moderated condition effects for the RAENS (Benjamini-Hochberg adjusted $p_{BHa} = .0252$) and TEMA-3 $p_{BHa} = .0252$). The moderation effect for the NSB ($p_{BHa} = .0730$) produced a raw p value of .0313. Figure 1 presents the condition differences across the range of baseline TEMA-3 values for the RAENS, TEMA-3, and NSB at posttest. The graphs show the main effects (dark line) with the 95% confidence intervals (light lines) across the range of pretest TEMA-3 scores (Preacher et al., 2006). The difference between conditions is considered statistically significant when the confidence bounds exclude zero.

Students in the ROOTS condition outperformed those in the control condition on the RAENS across nearly all pretest TEMA-3 scores (those below the 97th sample percentile or a TEMA-3 score of 36). Conditions differed for 97% of the sample, and students with lower TEMA-3 scores at pretest appeared to benefit most from ROOTS on the RAENS. The region of statistical significance includes TEMA-3 score of 36 or less, which corresponds to the 95th national percentile for students who entered kindergarten at ages 5 years to 5 years 2 months and the 84th national percentile for students who entered kindergarten at ages 5 years 6 months to 5 years 8 months.

Table 1. Results of Partially Nested Time × Condition Analyses That Tested Pretest as a Moderator of Differences in Fall-to-Spring Gains in Mathematics Scores Between ROOTS Students Nested Within Groups and Unclustered Control Students.

Variable	RAENS	Oral Counting	ASPENS Composite	NSB Total	TEMA-3
Fixed effects					
Intercept	17.07**** (.27)	21.34**** (1.22)	22.32**** (1.69)	12.28**** (.24)	17.07**** (.27)
Time	6.75**** (.38)	18.11**** (1.62)	35.31**** (2.09)	6.21**** (.29)	6.75**** (.38)
Condition (ROOTS)	.00 (.33)	1.53 (1.47)	1.06 (2.10)	.20 (.30)	.00 (.33)
Time × Condition	2.03**** (.48)	2.73 (1.97)	17.77**** (2.63)	.33 (.38)	2.03**** (.48)
Pretest TEMA-3	1.00**** (.04)	1.12**** (.19)	1.97**** (.26)	.35**** (.04)	1.00**** (.04)
Pretest TEMA-3 × Condition	.00 (.05)	.17 (.22)	−.40 (.31)	−.04 (.04)	.00 (.05)
Pretest TEMA-3 × Time	−.02 (.06)	.59* (.25)	1.14*** (.33)	.22**** (.04)	−.02 (.06)
Pretest TEMA-3 × Condition × Time	−.19** (.07)	−.03 (.30)	.05 (.38)	−.12* (.05)	−.19** (.07)
Variances					
Time × Condition between ROOTS groups	1.26** (.48)	6.56 (8.03)	43.64* (18.33)	.90* (.43)	1.26** (.48)
Pre-/posttest covariance	.00 (.49)	37.58*** (11.06)	130.43**** (21.48)		.00 (.49)
Residual	10.87**** (.79)	211.72**** (14.34)	315.59**** (23.60)		10.87**** (.79)
ROOTS residual				7.90**** (.66)	
ROOTS pre-/posttest covariance				2.14*** (.56)	
Control residual				5.97**** (.87)	
Control pre-/posttest covariance				3.01*** (.80)	
Pretest TEMA-3 × Condition × Time, *p*	.0041	.9198	.8899	.0313	.0072
Pretest TEMA-3 × Condition × Time, BH *p*	.0252	.9198	.9198	.0730	.0252
Pretest TEMA-3 × Condition × Time, *df*	587	551	551	334	587
ICC ROOTS groups	.09	.03	.12	.10	.10
Likelihood ratio χ^2	0.12	1.65	0.35	3.36	0.14
p	.9405	.4377	.8388	.1865	.9323

Note. Fixed effects and variances shown with standard errors in parentheses. The models nested only ROOTS students within groups. *P* values also provided with the Benjamini-Hochberg (BH) correction. Degrees of freedom for tests of fixed effects based on the Satterthwaite approximation. Likelihood ratio tests compared homoscedastic to heteroscedastic residuals ($\alpha = .20$, 2 degrees of freedom). ASPENS = Assessing Student Proficiency in Early Number Sense; ICC = intraclass correlation; NSB = Number Sense Brief Screener; RAENS = ROOTS Assessment of Early Numeracy Skills; TEMA-3 = Test of Early Mathematics Ability–Third Edition.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$. ****$p < .0001$.

Table 2. Results of Partially Nested Mixed-Model Analyses of Covariance on Posttest and Follow-Up Mathematics Scores That Tested Pretest as a Moderator of Differences Between ROOTS Students Nested Within Groups and Unclustered Control Students.

Variable	Posttest SESAT	Follow-Up SAT-10
Fixed effects		
Intercept	449.00****	495.62****
	(2.13)	(1.85)
Condition (ROOTS)	4.35†	.45
	(2.52)	(2.58)
Pretest TEMA-3	3.16****	2.01****
	(.33)	(.29)
Pretest TEMA-3 × Condition	−.10	−.19
	(.38)	(.36)
Variances		
ROOTS group intercept	19.89	154.55**
	(26.54)	(51.66)
ROOTS residual	573.11****	459.80****
	(48.97)	(50.17)
Control residual	711.25****	252.49***
	(85.94)	(74.17)
Pretest TEMA-3 × Condition, *p*	.7880	.5924
Pretest TEMA-3 × Condition, BH *p*	.9198	.9198
Pretest TEMA-3 × Condition, *df*	262	256
ICC ROOTS groups	.03	.25
Likelihood ratio χ^2	2.49	6.37
p	.1145	.0116

Note. Fixed effects and variances shown with standard errors in parentheses. The models nested only ROOTS students within groups. *P* values also provided with the Benjamini-Hochberg (BH) correction. Degrees of freedom for tests of fixed effects based on the Satterthwaite approximation. Likelihood ratio tests compared homoscedastic to heteroscedastic residuals (α = .20, 1 degree of freedom). ASPENS = Assessing Student Proficiency in Early Number Sense; ICC = intraclass correlation; NSB = Number Sense Brief Screener; RAENS = ROOTS Assessment of Early Numeracy Skills; SAT-10 = Stanford Achievement Test–Tenth Edition; SESAT = Stanford Early School Achievement Test; TEMA-3 = Test of Early Mathematics Ability–Third Edition.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$. ****$p < .0001$.

For the TEMA-3 at posttest, we found statistically significant condition differences for students with pretest TEMA-3 scores below the 77th sample percentile (score of 22). This region of significance included students below the 63rd national percentile at ages 5 years to 5 years 2 months and below the 37th national percentile at ages 5 years 6 months to 5 years 8 months.

Students in the ROOTS condition outperformed those in the control condition on the NSB with pretest TEMA-3 scores below the 26th sample percentile (score of 12). The region of significance includes students ages 5 years to 5 years 2 months below the 19th national percentile and students ages 5 years 6 months to 5 years 8 months below the 6th national percentile. The moderation effect was not statistically significant for the NSB after correcting for multiple tests.

ROOTS Group Size Differences

Clarke et al. (2017) found no differences between large and small groups for any variables ($p > .15$). Analyses of initial skill moderating group differences found all unadjusted *p* values exceeded .05 except

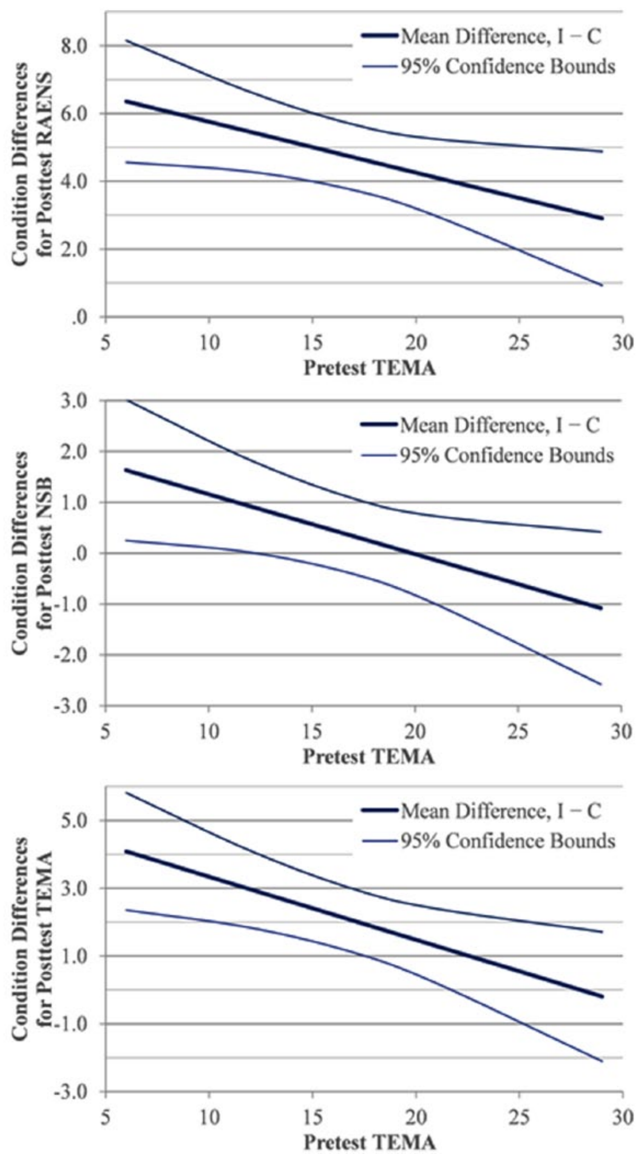

Figure 1. Interactions between condition and initial Test of Early Mathematics Ability–Third Edition (TEMA-3) scores. The figures plot the differences between conditions (vertical axis) as a function of pretest TEMA-3 values (horizontal axis). Heavy lines show the predicted difference between conditions across the range of TEMA-3 values. The two light, outer lines depict the 95% confidence bounds on the mean difference. Statistically significant differences between conditions occur when both confidence intervals fall either above or below zero. The differences between condition is considered statistically significant for the ROOTS Assessment of Early Numeracy Skills below the pretest TEMA-3 score of 32 (97th sample percentile), for the Number Sense Brief Screener below the pretest TEMA-3 score of 12 (26th sample percentile), and for the TEMA-3 below the pretest TEMA-3 score of 22 (77th sample percentile).

for the ASPENS, where the TEMA-3 appeared to moderate the group differences (p = .0280), but the Benjamini-Hochberg adjusted p value was .1960.

Closing the Achievement Gap

To investigate whether the ROOTS closed the achievement gap, we examined the proportion of students who exceeded the 25th national percentile on the posttest TEMA-3, SESAT, and SAT-10 by treatment condition. We did not have specific ages for individual students, so we examined two relevant age ranges for the TEMA-3. The average student begins kindergarten between ages 5 and 6, and ends about 9 months later. We therefore calculated the number of students above the 25th percentile for students in the age ranges of 5 years 8 months ± 1 month and 6 years 4 months ± 1 month. For students around age 5 years 8 months, a standard score of 90 translates to a raw TEMA-3 score of 22. Approximately 70% of the ROOTS sample but only 50% of the control sample exceeded this criterion for the 25th national percentile. For students around age 6 years 4 months, a standard score of 90 translates to a raw TEMA-3 score of 29, and about 34% of ROOTS students and 26% of control students exceeded the criterion. For the SESAT, 25.1% of the ROOTS sample but only 22.7% of the control sample scored above the 25th national percentile. For the SAT-10, 21.3% of students in the ROOTS condition and 15.7% of the control sample exceeded the 25th national percentile.

Discussion

The purpose of this study was to explore the moderating role of initial skill on student outcomes. Results related to this research question indicated that students with lower initial skills, as measured by the TEMA-3, showed greater benefit from the intervention on two out of six kindergarten outcome measures. We did not find a difference in this relationship by group size. Based on our previous work in mathematics (Clarke et al., 2015) and with the ROOTS curriculum (Clarke et al., 2017), the results matched our study hypotheses. Overall, most students differed on the TEMA-3 at pretest by less than 1 point (Clarke et al., 2017), yet by posttest, substantially greater numbers of students in the ROOTS condition scored within the average range on the TEMA-3.

Results related to this research question indicated that students with lower initial skills, as measured by the TEMA-3, showed greater benefit from the intervention on two out of six kindergarten outcome measures. We did not find a difference in this relationship by group size.

Collectively, the results provide a range of interesting points to consider and discuss as the field advances and attempts to better understand the intricacies of mathematics intervention work in the early elementary grades. Findings from this study add to a number of studies that have attempted to examine the role of initial skill status and intervention response. To date, clear patterns have not emerged, with some work suggesting similar effect across initial skill (e.g., L. S. Fuchs, Sterba, et al., 2016) or differential positive effects for lower initial skill (e.g., L. S. Fuchs, Malone, et al., 2016) or greater initial skill (Toll & Van Luit, 2013). Results from this study show a general pattern of differential effect for students with initial lower skill on a general measure of mathematics achievement. Collectively, these results are difficult to interpret. In part, the disparate results across studies may be due to the differences in the independent variable studied, the intervention. For example, while ROOTS was delivered as a Tier-2 program, the intensity of the ROOTS intervention may more closely align with conceptualizations of what a Tier-3 experience should look like. Original conceptualizations of RTI models noted that by design, many intensive intervention experiences represented the intensification of instruction reserved for special education, and perhaps general education with modifications should serve as the platform by which to engage response in a RTI model (L. S. Fuchs, 2003). Thus, the intensive nature of the ROOTS intervention may have been better aligned with the learning needs of students with significant skill deficits. Correspondingly, within the current study, it seems reasonable to hypothesize students on the upper end of the sample skill spectrum may have benefited from a less intensive intervention delivered as a supplement to general education. Given the lack of long-term effects of mathematics interventions in general (Starkey & Klein, 2008) and for ROOTS specifically (Clarke et al., 2017), the finding that initial skill did not moderate outcomes on our first-grade follow-up is not surprising. Results may also reflect the nature of the measure used to define initial skill. In this study, the TEMA-3 was used to operationalize initial skill. The TEMA-3 is a general measure of early mathematics and, in this case, also distal to the intervention. A more proximal measure to the intervention may have resulted in different findings (e.g., a differential benefit for students with greater initial skill).

A fundamental premise of RTI models is that instruction is modified or adjusted based on intervention response with the provision of more intensive intervention to students as they exhibit nonresponse and move from Tier 1 to Tier 3 (Coyne et al., 2013; Vaughn & Fuchs, 2003). However, the concept of intervention intensity remains elusive and difficult to define despite widespread recognition that premising decision making related to services necessitates defining overall intervention intensity and its subcomponents (Codding & Lane, 2015; D. Fuchs & Fuchs, 2017). When designing the ROOTS studies, our initial conceptualization of intervention intensity focused on group size, a common variable hypothesized to relate to intervention intensity (Baker et al., 2010). Findings from previous work (Clarke et al., 2017) indicated that greater rates of student behaviors hypothesized to increase learning (i.e., practice opportunities) were increased in smaller small groups but that those groups did not show a differential positive affect on mathematics outcomes. It could be hypothesized that although decreasing group size will increase desired teacher and student behaviors, those differences may not be relevant to student growth beyond a certain threshold. And if that is the case, and it holds true across other teacher and student behaviors (e.g., modeling) hypothesized to represent leverage points to increase intervention intensity, is doing more of the same an intensification of an intervention? Mixed findings in both mathematics and readings related to group size suggest that although group size may be an important hypothesized mechanism to increase the intensity of intervention, further research is necessary to explicitly contrast effects of groups with varying sizes and to investigate other mechanisms to increase intervention intensity. In addition, such efforts should consider the moderating role of crucial student-level variables, including but not limited to initial skill (Gersten, 2016; Ochsendorf, 2016; Woodward, 2016).

We consider work in this vein to represent a new wave of intervention research focused on how to better tailor the interventions schools deploy for students at risk. However, efforts by the field to better design and deliver interventions invoke a number of questions that are often deeply embedded within RTI models and special education. For example, the advent and development of curriculum-based measurement was driven by a general belief that progress monitoring was essential to monitor intervention response because it was not possible a priori to know how an individual student would respond to a given intervention (Deno, 1985). Service delivery models that propose skipping a tier of delivery seem at some level to contradict a long-standing theoretical orientation toward intervention delivery. To what extent do we need to have confidence (e.g., 50%, 80%, 90%) that a student will not respond to an intervention such that moving directly to a more intensive intervention is warranted? And how should those questions be considered in light of the concerns that RTI models may become another wait-to-fail model (Fletcher, Coulter, Reschly, & Vaughn, 2004; L. S. Fuchs & Vaughn, 2012). As measured by our primary distal measure (TEMA-3), results from this study suggest that between 34% and 70% of the treatment sample exited the intervention program with skills in the range that would enable them to benefit from Tier-1 instruction. Given previous results (Clarke et al., 2017) indicating no treatment effects for the SESAT (posttest) or SAT-10 (delayed first-grade posttest), the finding that approximately 21% to 25% of treatment students (slightly more than the percentage of control students) met a similar threshold on the SESAT and SAT-10 is not surprising. Collectively, given the results across measures related to the treatment normalizing student mathematics such that students could theoretically benefit from Tier-1 instruction, variations to service delivery models warrant consideration. For example, would service delivery models that considered immediate and direct placement in a Tier-3 intervention be better suited to meet the learning needs of severely-at-risk learners? Initial investigations related to the ability to accurately predict response show promise (e.g., Compton et al., 2012), as do models where students are placed immediately in more intensive instructional settings or where interventions are systematically altered based on ongoing student data (e.g., Al Otaiba et al., 2014; Coyne et al., 2013; Denton, Fletcher, Taylor, Barth, & Vaughn, 2014). The relatively consistent finding related to effects fading over time (Starkey & Klein, 2008) and found in our previous work (Clarke et al., 2017) suggests that even for treatment responders, addressing mathematics learning needs of at-risk learners warrants consideration of models that approach intervention from a multiyear, multitier perspective. How should the field consider these models as it relates to intervention response and the interaction of intervention response and crucial variables, such as initial skill?

Service delivery models that propose skipping a tier of delivery seem at some level to contradict a long-standing theoretical orientation toward intervention delivery.

The advancements in our understanding of mathematics intervention for young students have increased remarkably in the last decade. We believe a continued focus on questions related to mediators and moderators of interventions, aspects of intervention intensity, and modifications to RTI service delivery models warrants additional empirical investigation and discussions among leaders and researchers in the field of special education. Efforts in this regard are essential to moving the field forward and ensuring that all students are successful in acquiring essential mathematical knowledge.

References

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in Tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children*, *81*, 11–27. doi:10.1177/0014402914532234

Archer, A. L., & Hughes, C. A. (2011). *Explicit instruction: Effective and efficient teaching*. New York, NY: Guilford Press.

Baker, S. K., Fien, H., & Baker, D. L. (2010). Robust reading instruction in the early grades: Conceptual and practical issues in the integration and evaluation of Tier 1 and Tier 2 instructional supports. *Focus on Exceptional Children*, *42*(9), 1–20. doi:10.17161/fec.v42i9.6693

Baldwin, S. A., Bauer, D. J., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. *Psychological Methods*, *16*, 149–165. doi:10.1037/a0023464

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. M. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE No. 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED560820.pdf

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300. Retrieved from http://www.jstor.org/stable/2346101

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, *38*, 333–339. doi:10.1177/00222194050380040901

Bryant, B. R., Bryant, D. P., Porterfield, J., Dennis, M. S., Falcomata, T., Valentine, C., . . . Bell, K. (2016). The effects of a Tier 3 intervention on the mathematics performance of second grade students with severe mathematics difficulties. *Journal of Learning Disabilities*, *49*, 176–188. doi:10.1177/0022219414538516

Clarke, B., Baker, S. K., 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. doi:10.1080/19345747.2014.980021

Clarke, B., Doabler, C. T., Fien, H., Baker, S. K., & Smolkowski, K. (2012). *A randomized control trial of a tier 2 kindergarten mathematics intervention* (Project ROOTS). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Clarke, B., Doabler, C. T., Kosty, D., Kurtz Nelson, E., Smolkowski, K., Fien, H., & Turtura, J. (2017). Testing the efficacy of a kindergarten mathematics intervention by small group size. *AERA Open*, *3*(2), 1–16. doi:10.1177/2332858417706899

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*, 152–165. doi:10.1177/0022219414538514

Clarke, B., Doabler, C., Smolkowski, K., Kurtz Nelson, E., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a Tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness*, *9*, 607–634. doi:10.1080/19345747.2015.1116034

Clarke, B., Gersten, R. M., Dimino, J., & Rolfhus, E. (2011). *Assessing Student Proficiency of Number Sense (ASPENS)*. Longmont, CO: Cambium Learning Group, Sopris Learning.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review*, *33*, 234–248.

Codding, R. S., & Lane, K. L. (2015). A spotlight on treatment intensity: An important and often overlooked component of intervention inquiry. *Journal of Behavioral Education*, *24*, 1–10. doi:10.1007/s10864-014-9210-z

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. Retrieved from http://www.corestandards.org/the-standards/mathematics

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., . . . Bouton, B. D. (2012). Accelerating chronically unresponsive children to Tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities*, *45*, 204–216. doi:10.1177/0022219412442151

Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Simmons, L. E., Kwok, O.-M., Kim, M., . . . Rawlinson, D. A. M. (2013). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children*, *80*, 25–44. Retrieved from http://cec.metapress.com/content/F4321275232V11WX

Dasgupta, A., Lawson, K. A., & Wilson, J. P. (2010). Evaluating equivalence and noninferiority trials. *American Journal of Health-System Pharmacy*, *67*, 1337–1343. doi:10.2146/ajhp090507

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*, 219–232. doi:10.1177/001440298505200303

Denton, C. A., Fletcher, J. M., Taylor, W. P., Barth, A. E., & Vaughn, S. (2014). An experimental evaluation of guided reading and explicit interventions for primary-grade students at-risk for reading difficulties. *Journal of Research on Educational Effectiveness*, *7*, 268–293. doi:10.1080/19345747.2014.906010

Doabler, C. T., Clarke, B., & Fien, H. (2012). *ROOTS Assessment of Early Numeracy Skills (RAENS)* [Unpublished measurement instrument]. Eugene, OR: Center on Teaching and Learning.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446. doi:10.1037/0012-1649.43.6.1428

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, *46*, 166–181. doi:10.1177/0022219411410233

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia*, *54*, 304–331. doi:10.1007/s11881-004-0015-y

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide* (Practice Guide No. NCEE 2014-4005). Washington, DC: National Center for Education and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/early_math_pg_111313.pdf

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

Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice*, *18*, 172–186. doi:10.1111/1540-5826.00073

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities*, *45*, 257–269. doi:10.1177/0022219412442167

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2013). Intervention effects for students with comorbid forms of learning disability: Understanding the needs of nonresponders. *Journal of Learning Disabilities*, *46*, 534–548. doi:10.1177/0022219412468889

Fuchs, L. S., Malone, A. S., Schumacher, R. F., Namkung, J., Hamlett, C. L., Jordan, N. C., . . . Changas, P. (2016). Supported self-explaining during fraction intervention. *Journal of Educational Psychology*, *108*, 493–508. doi:10.1037/edu0000073

Fuchs, L. S., Sterba, S. K., Fuchs, D., & Malone, A. S. (2016). Does evidence-based fractions intervention address the needs of very low-performing students? *Journal of Research on Educational Effectiveness*, *9*, 662–677. doi:10.1080/19345747.2015.1123336

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*, 195–203. doi:10.1177/0022219412442150

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*, 1539–1552. doi:10.1037/a0025510

Gersten, R. M. (2016). What we are learning about mathematics interventions and conducting research on mathematics interventions. *Journal of Research on Educational Effectiveness*, *9*, 684–688. doi:10.1080/19345747.2016.1212631

Gersten, R. M., Beckmann, S., Clarke, B., Foegen, A., March, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools* (Practice Guide Report No. NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_math_pg_042109.pdf

Gersten, R. M., & Chard, D. J. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education*, *33*, 18–28. doi:10.1177/002246699903300102

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability–Third Edition (TEMA-3)*. Austin, TX: ProEd.

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

Hannan, P. J., & Murray, D. M. (1996). Gauss or bernoulli? A Monte Carlo comparison of the performance of the linear mixed-model and the logistic mixed-model analyses in simulated community trials with a dichotomous outcome variable at the individual level. *Evaluation Review*, *20*, 338–352. doi:10.1177/0193841x9602000306

Harcourt Educational Measurement. (2002). *Stanford Achievement Test (SAT-10)*. San Antonio, TX: Author.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage.

Jordan, N. C., Glutting, J., & Ramineni, C. (2008). A number sense assessment tool for identifying children at risk for mathematical difficulties. In A. Dowker (Ed.), *Mathematical difficulties: Psychology and intervention* (pp. 45–57). San Diego, CA: Academic Press.

Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology*, *94*, 586–597. doi:10.1037/0022-0663.94.3.586

Kromrey, J. D., & Dickinson, W. B. (1996). Detecting unit of analysis problems in nested designs: Statistical power and type i error rates of the f test for groups-within-treatments effects. *Educational and Psychological Measurement*, *56*, 215–231. doi:10.1177/0013164496056002003

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

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2 ed.). New York, NY: Wiley.

Miller, B., Vaughn, S., & Freund, L. (2014). Learning disabilities research studies: Findings from NICHD funded projects. *Journal of Research on Educational Effectiveness*, *7*, 225–231. doi:10.1080/19345747.2014.927251

Morgan, P. L., Hillemeier, M. M., Farkas, G., & Maczuga, S. (2014). Racial/ethnic disparities in ADHD diagnosis by kindergarten entry. *Journal of Child Psychology and Psychiatry*, *55*, 905–913. doi:10.1111/jcpp.12204

National Center for Education Statistics. (2015). *The nation's report card: 2015 mathematics and reading assessements. National achievement level results, 8th grade*. Washington, DC: National Center for Education Statistics. Retrieved from http://www.nationsreportcard.gov/reading_math_2015/#reading/acl?grade=8

National Center on Response to Intervention. (2010). *Essential components of RTI: A closer look at response to intervention*. Washington, DC: U.S. Office of Special Education Programs, National Center on Response to Intervention. Retrieved from http://www.rti4success.org/

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

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: Mathematics Learning Study Committee.

National Science Board. (2008). *Science and engineering indicators 2008* (Vol. 2). Arlington, VA: National Science Foundation.

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

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*, 437–448.

Roberts, C., & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials*, *2*, 152–162. doi:10.1191/1740774505cn076oa

SAS Institute. (2016). *SAS/STAT (Version 14.2)* [User's guide]. Cary, NC: Author. Retrieved from http://support.sas.com/documentation/index.html

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037//1082-989X.7.2.147

Starkey, P., & Klein, A. (2008). Sociocultural influences on young children's mathematical knowledge. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 253–276). Charlotte, NC: Information Age.

Toll, S. W. M., & Van Luit, J. E. H. (2013). Accelerating the early numeracy development of kindergartners with limited working memory skills through remedial education. *Research in Developmental Disabilities*, *34*, 745–755. doi:10.1016/j.ridd.2012.09.003

Van Belle, G. (2008). *Statistical rules of thumb* (2nd ed.). Hoboken, NJ: Wiley.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research &*

Practice, *18*, 137–146. doi: 10.1111/1540-5826.00070

Vaughn, S., Thompson, S. L., Kouzekanani, K., & Dickson, S. (2003). The effects of three grouping formats on the reading performance of monolingual and English language learners with reading problems. *Journal of Educational Psychology*, *24*, 301–315.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*, 541–561.

Woodward, J. (2016). Commentary on intensive interventions: What are the limits of highly structured curriculum for at-risk students? *Journal of Research on Educational Effectiveness*, *9*, 678–683. doi:10.1080/19345747.2016.1212630

Authors' note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A120304 to the Center on Teaching and Learning at the University of Oregon. The opinions expressed are those of the authors and do not represent views of the institute or the U.S. Department of Education.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Scott Baker, Ben Clarke, Chris Doabler, and Hank Fien 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.

Manuscript received April 2018; accepted August 2018.