Article Information
Article first published online: April 5, 2022

# A Conceptual Replication of a Kindergarten Math intervention Within the Context of a Research-Based Core

## Authors

Ben Clarke, Jessica Turtura, Taylor Lesner, Madison Cook, Keith Smolkowski, Derek Kosty, Christian T. Doabler.

## Full Reference

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. https://doi.org/10.1177/00144029221088938

## Funding

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants **R305K040081** and **R305A080699** to the Center on Teaching and Learning at the University of Oregon and Pacific Institutes for Research.

Abstract

The purpose of this study was to conduct a replication study of a kindergarten mathematics intervention, ROOTS, delivered within the context of a research base core program. In the study, sixty two classrooms were randomly assigned to treatment (ROOTS) or a business as usual control. All classrooms implemented a research based core program (Early Learning in Mathematics). Participants included 163 treatment students and 145 control students nested within classrooms. Key differences between the current replication study and the original study included geographical region, instructional context, and student initial skill. In contrast to the significant positive effects (Hedges' g values of .30 to .38) found in the original study, no significant differences were found between the treatment and control conditions. Pretest skills did not moderate treatment effects. Implications for replication research and evaluating intervention efficacy are discussed.

*Keywords:* math, intervention, replication, number sense

# A Conceptual Replication of a Kindergarten Math intervention Within the Context of a Research-Based Core

Replication research is essential to the scientific process and ensuring the credibility of any findings generated from initial research efforts (Schmidt, 2009). Replication research supports the generalizability of hypotheses generated from initial research (Schmidt, 2009) and serves to not only affirm scientific understanding of the findings replicated, but to expose methodological biases and insufficiencies within the initial study (Makel & Plucker, 2014). Replication strengthens the credibility of previous research findings and informs policy decisions regarding educational practice at multiple levels (NSF & IES 2018). However, replication studies remain undervalued in the field of educational research. While replication work is essential for increasing confidence in the results generated through education research, it is often avoided due to notions that the work is not original, publishable, or prestigious within the research community (Makel & Plucker, 2014). The dire state of replication in education is gaining increased attention and focus. In 2018, the National Science Foundation and the Institute of Education Sciences (NSF & IES 2018) issued an explicit call for increased "reproducibility and replication" of education research (NSF & IES 2018).

The lack of replication is of particular concern within the context of intervention work designed to meet the needs of students with or at-risk for learning disabilities. While there is a growing body of evidence on effective interventions and instructional approaches (Gersten, Beckmann, et al., 2009; Nelson & McMaster, 2019), findings in intervention research must be replicated across a range of educational contexts in order to establish confidence that academic interventions can be used widely in practice, (Makel & Plucker, 2014). Replication allows the research community to examine whether results from previous research stand when settings and contextual factors vary, resulting in a convergence of evidence regarding a treatment or treatment approach (Coyne et al., 2016; Coyne et al., 2013). Replication work also offers unique opportunities to fully explore the robustness of previous research findings and the conditions under which an intervention will work and for whom (Doabler et al., 2016). Despite the importance of replication, few intervention studies have been systematically replicated, in particular within the field of special education (Chhin et al., 2018; Cook et al., 2014; Makel et al., 2016).

There are two types of replication studies: direct replications and conceptual replications (Coyne et al., 2016; Makel et al., 2012; Schmidt, 2009). Direct replications attempt to hold constant all aspects of the treatment, and are conducted using the same methods and under the same conditions as the original research. Direct replications in the field of education research are rare because they are difficult, if not impossible, to conduct given the complicated environments of schools. A more feasible option for educational researchers is to conduct closely aligned conceptual replications (Coyne et al., 2016). Closely aligned conceptual replications typically vary from the original study on one or two elements and are conducted to verify whether findings generalize across settings, conditions, and participants (Schmidt, 2009). If the variations are minimal, conceptual replications can demonstrate similar capacity as direct replications. For example, a closely aligned replication can determine whether treatment effects demonstrated in the original study replicate in a different geographical region (Coyne et al., 2013). Conceptual replications are beneficial for informing both effective educational policy making and usage of interventions as they provide evidence of the generalizability of interventions, which ultimately leads to a greater understanding of their outcomes and efficacy with students of varied sociodemographic backgrounds (NSF & IES 2018). IES, a primary federal funding source for education research, has explicitly called for replication studies in the field since 2004 and the goals of the Education and Special Education Research Grants program under the ESRA call for "replication of interventions with prior evidence of efficacy" (Chhin et al., 2018). Replication studies are of particular value in areas and times of critical academic development (Coyne et al., 2016) such as the during the transition to school and the learning of early mathematics (Gersten, Beckmann, et al., 2009). The current study represents a conceptual replication that tested the efficacy of a kindergarten mathematics intervention program developed and studied with support from IES.

**Mathematics Development and Intervention**

Mathematical proficiency is critical in today's ever-changing, increasingly complex society (Frye et al., 2013; National Research Council, 2001, 2009). Unfortunately, data from the National Assessment of Educational Progress (NAEP) indicates that U.S. students are not achieving mathematical proficiency including those with or at-risk for learning disabilities (2019). Low achievement levels in mathematics are especially alarming in light of findings that math difficulties begin early, are persistent, and predictive of later outcomes. Converging research shows that children's early knowledge of math strongly predicts their later success in math, with children who begin with the lowest achievement levels showing the lowest growth in mathematics across time (Bodovski & Farkas, 2007; Duncan et al., 2007; Hanich et al., 2001; Morgan et al., 2009). Importantly, there is evidence that these long term trajectories can be altered (Morgan et al., 2014) and that early *growth* in mathematical ability can be a stronger predictor of mathematics skill than initial mathematics ability (Watts et al., 2014).

Given the importance of early mathematics and the potential to alter long term trajectories, attention has turned to identifying effective mechanisms for improving student mathematics proficiency including a growing body of research focused on identifying the most effective ways to teach mathematical skills to struggling learners in the early elementary grades. This research suggests that an in-depth understanding of the whole number system is a critical first step in achieving proficiency in more sophisticated mathematics, such as rational numbers and algebra (Gersten, Beckmann, et al., 2009; National Council of Teachers of Mathematics, 2006; National Mathematics Advisory Panel, 2008; National Research Council, 2001). Several researchers have developed, evaluated, and found positive impacts for intervention curricula targeting early number sense and foundational whole number concepts (e.g. Bryant et al., 2008; Clarke et al., 2014; Dyson et al., 2013; Fuchs et al., 2005; Sood & Jitendra, 2013). As a solid evidence base has emerged on early mathematics interventions, greater interest has been placed on understanding the conditions under which interventions work and for whom (Miller et al., 2014). This implies the need for more nuanced studies, including replications.

The purpose of the current study was twofold. First, we wanted to conduct a conceptual replication to demonstrate the importance of replication research in gaining a better understanding intervention impacts. To do so, our second purpose was test the efficacy of a Tier II kindergarten mathematics intervention program (ROOTS) delivered in the context of an evidence-based Tier I core mathematics curriculum (ELM). A previous investigation of the intervention (Clarke et al., 2016) provides initial treatment effects for ROOTS. The initial study, which occurred in the 2009-10 school year, utilized a partially nested randomized controlled trial. A total of 140 students from 29 kindergarten classrooms across two school districts in the Pacific Northwest were randomly assigned to a treatment (ELM + ROOTS) condition ($n$ = 67) or a control (ELM only) condition ($n$ = 73). In the treatment condition, the 50-lesson ROOTS intervention was delivered to small groups of eligible students 3 times per week for 16 to 20 weeks during the second half of the school year. Students in the treatment condition received the ROOTS intervention in addition to core mathematics instruction provided with the ELM curriculum. Control students received only the ELM core mathematics curriculum.

Overall effects of the ROOTS intervention on mathematics achievement were assessed using a mixed model Time × Condition analysis (Murray, 1998) designed to account for students nested within classrooms. Results indicated that students in the ROOTS treatment condition improved from fall to spring at a statistically significant greater rate than students in the control condition on one of the two distal measures (i.e., the Test of Early Mathematics Achievement standard scores), but not the EN-CBM. On both measures, ROOTS students demonstrated substantively important positive effects (What Works Clearinghouse, 2011) with Hedges' g values of .38 for the TEMA standard score and .30 for the EN-CBM.

We consider the current study as a closely aligned replication based on the degree of overlap with the initial study. Similar to the original investigation, the current study tested the same intervention and used the same outcome measures and statistical analyses. Both studies employed a partially nested randomized controlled trial and applied the same criteria for determining students' eligibility for the intervention. In both studies, teachers nominated the five lowest-performing students or those who would benefit most from a small-group math intervention. Intervention dosage levels were the same across studies, with district-employed personnel delivering the ROOTS intervention in small-group formats at the same frequency.

The current study evaluated the impact of the ROOTS intervention in a different geographical region. Clarke et al. (2016) conducted the initial study in the Pacific Northwest, while the current replication study took place in Texas. The student sample and instructional context for the replication study also differed from those in the initial study. For example, one key instructional difference between sites was the duration of the core math block that was provided outside of ROOTS intervention groups. Recognizing these contextual and instructional differences, the new research sites were expected to offer a unique counterfactual. Thus, positive findings from the replication would increase the credibility and generalizability of the intervention's impact on student mathematics outcomes.

The following research questions were addressed as part of the replication study:

1. What is the impact of the Roots program on mathematics achievement of at-risk students?
2. Do Roots students reduce the achievement gap with their non-at-risk peers by making greater gains than their non-at-risk peers?

## Method

### Design

This study involved full-day kindergarten teachers who had participated in a study of the ELM core mathematics program (Clarke et al., 2011). Classrooms were assigned randomly to condition, and then within classrooms teachers selected students whom teachers expected would most benefit from small group instruction. Specifically, classrooms were randomly assigned to treatment or control conditions, blocking on teachers' prior experience with ELM. That is, we randomly assigned teachers with 1 year of ELM experience to ROOTS or control and then randomly assigned teachers new to ELM implementation to ROOTS or control. In schools with multiple classrooms, we also assigned classrooms to condition within school. Blocking, also called stratification, on ELM experience and school experimentally controls for biases that might stem from systematic differences between conditions (e.g., more ROOTS teachers with no prior ELM experience). A total of 62 classrooms were included: 32 in the treatment condition (ELM +

ROOTS) and 30 in the control condition (ELM only). Teachers were asked to nominate the five lowest performing students or those who would most benefit from a small-group math intervention. The nomination process entailed a three-step process. First, to be considered eligible for the intervention, a student had to have a pretest score below the 40th percentile on the Test of Early Mathematics Ability–Third Edition (TEMA-3). The use of the 40th percentile cut score was based on designation of risk status in widely used screening systems (Kaminski et al., 2008) and evaluations of comprehensive core programs (Gamse et al., 2008) including the ELM program (Clarke et al., 2011) From those students who qualified for intervention, teachers were provided with student scores from a battery of curriculum based measures that assessed students' number proficiencies (see "Measures" section). Teachers then selected up to five students who demonstrated low performances on the number sense measures. Teachers nominated 308 students as eligible for small-group instruction, with 163 students in intervention classrooms and 145 students in control classrooms. Forty-two teachers identified five students, one teacher identified three, eleven teachers identified four students, five identified six, and three identified three students, with the deviations from five students split similarly across conditions. Classroom teachers in both conditions provided whole class ELM instruction throughout the year for all students, and treatment and control classrooms provided the same amount of daily mathematics instruction. In intervention classrooms (ELM + ROOTS), the "ROOTS students" received all of the whole-class ELM instruction. On 3 days per week, however, instead of practicing that day's ELM topics independently at the end of the lesson (i.e., math practice worksheets), they received ROOTS instruction. Given that ROOTS was not offered in control classrooms, nominated control students participated in whole class ELM instruction, 5 days per week, including all of the individualized math practice. We controlled for time by delivering the ROOTS instruction during the individual, worksheet-based math practice portion of ELM in treatment classrooms. ROOTS instruction began in January and continued until the end of May. Trained instructional assistants provided ROOTS instruction.

**Participants**

**Instructional Assistants**. A total of 14 instructional assistants (IAs) participated in the study; 11 were female and 6 identified themselves as White, 3 identified themselves as Hispanic, and 5 identified themselves as African American. IAs were included in the study based on time and schedule availability. 9 of the IAs had college degrees, of whom 5 held current teacher certifications in elementary education. With respect to the remaining 5 IAs, 1 held an associate's degree and 4 were high school graduates. 9 of the IAs had completed college level coursework in mathematics. In this sample, 2 of the IAs had 5 or more years' experience, 2 had between 1 and 4 years' experience, and 10 had less than 1 years' experience. It is important to note that all IAs were employed by the participating school districts.

**Students**. All participating schools were pulled from one school district in Dallas, Texas. Within the district, 5% of students were White, 68% Hispanic, 26% Black, and 1% other. 86.8% of students in the district were eligible for free or reduced lunch. Within the ROOTS condition, 50% of students were male, 14% were English learners, and the average age was 66.3 months (SD = 3.9). Among control participants, 55% were male, 18% were English learners, and their average age was 67.1 months (SD = 4.2). Of the 283 students with a TEMA percentile rank at pretest, 53% scored at or below the 10th percentile, with 50% of students in ROOTS classrooms and 56% of students in control classrooms falling below the 10th percentile.

The sample also included 1,315 students who were not eligible for ROOTS, with 658 in intervention classrooms and 657 in control classrooms. Within the intervention classrooms, 49%

of students were male, 17% were English learners, and the average age was 67.5 months (SD = 4.0). Among control participants, 49% were male, 25% were English learners, and their average age was 67.3 months (SD = 3.9). All of these students received ELM instruction, and none of these students participated in ROOTS.

**Measures**

**Fidelity of Implementation**. Online logs completed by the 14 IAs who delivered the ROOTS intervention revealed that groups generally completed all 50 ROOTS lessons during the year. Trained research staff also directly measured implementation fidelity using a standardized observation instrument. The observation instrument was specifically designed to target mathematics activities within each lesson of the ROOTS curriculum. During the observations, observers coded whether IAs taught key design components prescribed within each lesson activity.

All observations were scheduled in advance and observers coded fidelity of implementation data for the duration of the assigned 20-min instructional time periods. Each

ROOTS group was observed 3 times over the course of the study, with approximately 4 to 5 weeks separating each observational round. For each prescribed lesson activity, observers rated implementation fidelity using a 3-point rating scale, where a score of 1.0 represented full implementation, 0.5 represented partial implementation, and 0.0 indicated an activity was not taught. Fidelity scores were computed as the mean across all lesson activities. The mean across the three observations per ROOTS group were used as an overall indicator of implementation fidelity. IAs demonstrated high fidelity scores for prescribed lesson activities with very little variability ($M = .96$, $SD = .07$).

**Instructional Logs**. All participating teachers (treatment and control) completed online instructional logs once every two weeks during the study. Logs were completed approximately 12-15 times during the school year. Teachers reported on their most recent math lesson by noting the lesson format, group size and type, lesson duration, mathematics vocabulary words used, and the specific content and skills taught during the lesson. Each log took approximately 15 minutes to complete.

**Test of Early Mathematics Ability**. TEMA (Pro-Ed, 2007) is a norm-referenced individually administered measure of early mathematics for children ages 3 to 8 years 11 months. The TEMA is designed to identify student strengths and weaknesses in specific areas of mathematics. The TEMA measures both formal mathematics and informal mathematics including skills related to counting, number facts and calculations, and related mathematical concepts. The test authors report alternate-form reliability of .97, and test–retest reliability ranges from .82 to .93. Concurrent validity with other criterion measures of mathematics is reported as ranging from .54 to .91.

**Early Numeracy Curriculum-Based Measurement**. EN-CBM (Clarke & Shinn, 2004) is a set of four measures based on principles of curriculum-based measurement (Shinn, 1989). Each 1-min fluency-based measure assesses an important aspect of early numeracy development including magnitude comparisons and strategic counting. The EN-CBM measures have been validated for use with kindergarten students including established validity with other measures of early mathematics including the Number Knowledge Test and the Stanford Achievement Test (Chard et al., 2008; Chard et al., 2005). The Oral Counting measure requires students to orally rote count as high as possible without making an error. Concurrent and predictive validities range from 46 to .72. The Number Identification measure requires students to orally identify numbers between 0 and 10 when presented with a set of printed number symbols. Concurrent and predictive validities range from .62 to .65. The Quantity Discrimination measure requires students to name which of two visually presented numbers between 0 and 10 is greater. Concurrent and predictive validities range from .64 to .72. The Missing Number measure requires students to name the missing number from a string of numbers (0–10). Students are given strings of three numbers with the first, middle, or last number of the string missing. Concurrent and predictive validities range from .46 to .63.

**Procedures**

**Data Collection**. All measures were individually administered to students. Trained staff with extensive experience in collecting educational research for research projects administered all student measures. All data collectors were required to obtain interrater reliability coefficients of .90 prior to collecting data with students. 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 application. Tests of Teleform scoring procedures of assessment protocols from previous research projects reveal high reliability values (i.e., .99) relative to assessor-scored protocols (.95).

**ROOTS Intervention**. ROOTS is a Tier 2 kindergarten intervention program that was designed to be delivered by IAs in small-group instructional formats (5:1 student to teacher ratio), 3 times per week, for 16 to 20 weeks during the second half of the school year. In contrast to the control condition (ELM only), ROOTS differs on a number of key variables. First, ROOTS is taught in small groups, whereas ELM is taught to the whole classroom. ELM occurs every day and contains 120 lessons. ROOTS occurs 3 days per week and contains 50 lessons. ROOTS exclusively focuses on content associated with whole number understanding. In contrast, ELM covers content in whole number understanding, geometry and measurement and thus is broader in content coverage than ROOTS. The goal of ROOTS is to support students' development of procedural fluency with and conceptual understanding of whole number concepts. The specific focus on whole number aligns with the CCSS (2010) and calls from mathematicians and expert panels for more focused and coherent Tier 1 curricula (NCTM, 2006; NMAP, 2008), and intervention programs designed to meet the needs of students at risk for MLD (Gersten, Beckmann, et al., 2009). ROOTS provides in-depth instruction in whole number concepts by linking the informal mathematical knowledge developed prior to school entry with the formal mathematical knowledge developed in kindergarten. The program includes 50 lessons, approximately 20 min in duration. Each lesson consists of 4 to 5 brief math activities that center on three key areas of whole number understanding: (a) Counting and Cardinality, (b) Number Operations, and (c) Base 10/Place Value. Curricular objectives advance students from an initial understanding of whole number through more sophisticated aspects of whole numbers in kindergarten mathematics. For example, the first half of the curriculum addresses counting objects, identifying numbers, and counting on from a given number. In the second half, lessons focus on beginning computational methods, such as adding one to a number, and place value concepts, such as using base 10 models to compose and decompose teen numbers into one 10 and so many ones.

ROOTS incorporates the principles of explicit and systematic instructional delivery and design that have been empirically validated to improve the mathematics achievement of at-risk learners and students with learning disabilities (Baker et al., 2002; Gersten, Chard, et al., 2009). Explicit delivery principles include modeling and demonstrating what students will learn, providing guided practice opportunities, using visual representations of mathematics, and delivering academic feedback. Systematic instructional design is the "behind-the-scenes" design activities (Kame'enui & Simmons, 1999; Simmons et al., 2007) that attend to the architectural features of a curriculum. Principles of systematic instructional design include prioritizing instruction around critical content, connecting new content with students' background knowledge, selecting and sequencing instructional examples, and scaffolding instruction. ROOTS focuses intensely on the whole number standards identified in the CCSS (2010). When introducing students to new and difficult mathematics concepts and skills, for example, the program initiates instruction with simpler teaching examples. Once students demonstrate initial proficiency with targeted math content, instructional scaffolds are systematically withdrawn to promote learner independence. Finally, the program incorporates positive teaching examples along with a select number of nonexamples to promote students' discrimination skills (Coyne et al., 2011).

**Professional Development**. Participating IAs attended three professional development (PD) workshops focused on the ROOTS curriculum. The initial PD workshop targeted the instructional objectives of Lessons 1 to 25, the critical content of kindergarten mathematics (CCSS, 2010), small group management techniques, and the instructional practices that have been empirically validated to increase student math achievement (e.g., teacher provided academic feedback; Gersten, Chard, et al., 2009). In the second and third workshops a similar format was followed, except that the focus was on the second half of the ROOTS curriculum,

Lessons 26 to 50. Workshops were 4 hr in length and were organized around three principles: (a) active participation, (b) content focused, and (c) coherence. On at least three occasions, IAs also received in classroom coaching from two expert teachers to increase implementation fidelity. Implementation research shows that ongoing coaching enhances teachers' sustained use of new instructional practices (Fixsen et al., 2005).

Two former educators, who were knowledgeable in the science of early mathematics development and instruction, served as coaches during the study. Typical coaching visits included direct observation and post observation feedback focusing on instructional delivery and implementation fidelity. Some IAs received more than three coaching visits if they or the coach felt more support was warranted (e.g., when there were particularly pervasive student behavior problems or the IA struggled with lesson implementation).

**Statistical Analysis**

We assessed intervention effects on each of the primary outcomes with a mixed model (multilevel) Time × Condition analysis (Murray, 1998) to account for students nested within classrooms. Primary analyses included the students in each classroom identified as at risk for math difficulties by the classroom teacher. Because each classroom included only one small group, the classroom and small group are redundant for the sample used in the analysis. The analysis tests differences between conditions on change in outcomes from the fall ($T_1$) to spring ($T_2$) of kindergarten clustered within classroom. The specific model tests time, *T*, coded 0 at pretest and 1 at posttest, condition, *C*, coded 0 for control and 1 for ROOTS, and the interaction between the two with the following composite model.

$$Y_{tjk} = (\gamma_{000} + \gamma_{001}C_k + \gamma_{100}T_{tjk} + \gamma_{101}T_{tjk}C_k) + (\boldsymbol{u}_{00k} + \boldsymbol{u}_{10k}T_{tjk} + \boldsymbol{r}_{0jk} + \boldsymbol{e}_{tjk}).$$

$Y_{tjk}$ represents a score for assessment occasion *t* on individual *j* in classroom *k*. The model includes three predictors: time, $T_{tjk}$, condition, $C_k$, and their interaction. Given the coding of *C* and *T*, the model includes the pretest intercept for the control condition, $\gamma_{000}$, the difference between conditions at pretest, $\gamma_{001}$, the estimate of gains for the control condition, $\gamma_{100}$, and the difference in gains between conditions, $\gamma_{101}$, the primary estimate of intervention efficacy. The model also includes four error variances: the classroom-level intercept, $\boldsymbol{u}_{00k}$, the classroom-level gains, $\boldsymbol{u}_{10k}T_{tjk}$, the student-level intercept, $\boldsymbol{r}_{0jk}$, and the residual, $e_{ij}$. With only two assessments, the variances $\boldsymbol{r}_{1jk}T_{tjk}$ and $\boldsymbol{e}_{tjk}$ are redundant and cannot be simultaneously estimated. The model excludes the $\boldsymbol{r}_{1jk}T_{tjk}$ term (Murray, 1998). The student-level intercept, $\boldsymbol{r}_{0jk}$, is also equivalent to the within-student covariation between pretest and posttest assessments. With 62 classrooms, tests of time by condition used 60 degrees of freedom.

**Model Estimation.** We fit models to our data with SAS PROC MIXED version 9.2 (SAS Institute, 2016) with maximum likelihood (ML) estimation. ML estimation for the Time × Condition analysis uses of all available data to reduce the potential for biased results even in the face of substantial attrition provided the missing data were missing at random (Schafer & Graham, 2002). Collins et al. (2001) demonstrated that sophisticated missing-data approaches, such as ML, do not introduce bias; the assumptions of the approach are relatively benign compared to complete case analysis (Allison, 2009; Schafer & Graham, 2002). In the present study, we did not believe that attrition or other missing data represented 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). 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 family moving).

**Effect Sizes and Interpretation of Results.** Hedges' *g* values were calculated to characterize the magnitude of treatment effects (What Works Clearinghouse, 2017). In response to the recommendations of the American Statistical Association (Wasserstein & Lazar, 2016), we abstained from using bright-line rules, such as from claims of "statistical significance" when $p <$

.05. *P* values have an interpretation as a measure of incompatibility between the observed data and all assumptions of the statistical model including the null hypothesis, $H_0$ (Greenland et al., 2016; Wasserstein & Lazar, 2016). This cumbersome definition neither informs on which assumptions are incorrect nor the importance of the association. To complement *p* values and Hedges's *g* values, we report Akaike weights (Akaike, 1973), which describe the strength of evidence for a one model when comparing it with others. Akaike weights—also called model probabilities—express the probability of a model given a set of competing models and the observed data (Burnham et al., 2011) and can be interpreted as the probability that the same model would be selected with a "replicate data set from the same system" (Burnham et al., p. 30). They quantify the strength of evidence for each hypothesis, represented by a statistical model, given the data and all other hypotheses (models) tested. For each analysis, we compared models for two hypotheses, one with the intervention effect ($H_A$) and one without ($H_0$), and reported the Akaike weight, *w*, for the model with the condition effect ($H_A$). With only two models, the model probability for $H_0$, the model without the condition effect, is $1 - w$. Hence, $w = .75$ suggests that the probability of $H_A$ is .75 while the probability of $H_0$ is .25. This roughly corresponds to a result with $p = .05$, which implies the limited value of "just-significant" results ($p \approx .05$). In such a case, the model for $H_A$ has an approximately 75% chance of being the best-fitting model; in other words, the model for $H_A$ is only three times as likely as the model for $H_0$ given the data.

## Results

Table 1 presents means, standard deviations, and sample sizes for the TEMA standard score, the TEMA percentile, and the EN-CBM by assessment time and condition. Below we present results for tests of attrition effects, ROOTS intervention impact, and additional analyses.

**Joiners and Attrition**

We experienced no classroom-level attrition. Student nonresponse is defined as students with data at posttest but not at pretest (joiners) or data at pretest but not posttest (attrition). The study included 308 students, of which 260 (84.4%) provided data at both assessments, 17 (5.5%) provided data at posttest but not pretest (joiners), 24 (7.8%) provided data at pretest but not posttest (attrition), and 7 (2.3%) provided no data. The analyses exclude students with no data. Because a portion of students missed assessments for temporary reasons (e.g., illness), a portion (unknown) of joiners and attriters were present for the full school year.

**Joiners.** Among students with posttest data, 6.5% were missing pretest data, with 6.0% in the control condition and 7.0% in the intervention condition. Although the differential rate of pretest nonresponse was small, the differential effects at posttest were not. On the posttest TEMA standard score, the 6.5% of students without pretest data differed between ROOTS and control conditions (96.2 versus 86.0; $g = 0.73$) substantially more so than students with pretest data (99.3 versus 88.9; $g = 0.03$). We therefor examined the effects of condition, attrition status, and their interaction on pretest scores within a mixed-model analysis of variance. The interaction between condition and pretest missingness indicated potentially troublesome joiner effects for the posttest TEMA ($t_{60} = –2.21$, p = .0309). The interaction between conditions and pretest missingness for on posttest EN-CBM scores indicated little influence of missingness ($t_{60} = 0.10$, p = .9242).

**Attrition.** Among students with pretest data, the overall attrition rate was 8.5%, with a differential rate of 3.5%, with 6.7% among control students and 10.1% for ROOTS students. On the pretest TEMA standard score, the 8.5% of students without posttest data differed between ROOTS and control conditions (77.6 versus 71.3; $g = 0.40$) more so than students with posttest data (81.6 versus 79.6; $g = 0.12$), but the interaction between condition implied minimal bias ($t_{60} = –0.06$, p = .9540). The interaction between conditions and posttest missingness for on pretest EN-CBM scores indicated little influence of attrition ($t_{60} = –0.32$, p = .7473) even though the effect size for condition differences was larger for students without posttest data ($g = 0.53$) than students who remained in the study ($g = 0.051$).

**Interpretation.** This joiner and attrition analysis should be interpreted in light of three important points. First, rates of attrition convey little about representativeness of the sample. "The proportions of the treatment and control groups that provide information are not particularly important, at least for internal validity. Nor does it matter whether the respondents differ systematically from the nonrespondents" (Foster & Bickman, 1996, p. 698). Second, fewer than 25 students contributed to the effect sizes estimates for students missing data, so the effect sizes have limited precision and should be interpreted with caution. Third, differences between conditions for both joiners and attrition favored the intervention condition, so students missing data at pretest appear to have been replaced by largely similar students at posttest. Maximum likelihood analyses with all available data may balance the effects of nonresponse and help minimize the potential for bias.

**Intervention Effects for ROOTS**

Among those students identified as eligible for ROOTS, we found limited support for differential gains among students provided with ROOTS compared to those in control classrooms on the TEMA standard scores ($g = -0.09$, 95% CI [-0.48, 0.29], $t_{60} = -0.48$, $p = .6312$, $w = .29$) and the EN-CBM total score ($g = 0.12$ [-0.18, 0.42], $t_{60} = 0.79$, $p = .4315$, $w = .33$). The effect sizes were small, and the model estimates for differences in gains between conditions were −1.2 for the TEMA standard scores and 5.9 for EN-CBM. See Table 2 for complete model results. The model probabilities suggest that the hypothesis of condition effects was unlikely to describe the data; the models without condition effects had considerably higher weights of $(1 - w)$ .71 and .67 for the TEMA and EN-CBM, respectively.

Without support for conditions differences, we examined whether pretest scores may have moderated the intervention effects. To do so, we added the interaction between pretest scores and treatment condition to the models. This produced no appreciable moderation effect for the TEMA (estimate = -0.04, 95% CI [-0.20, 0.13], $t_{60}$ = -0.44, $p$ = .6611, $w$ = .28) or the EN-CBM (estimate = -0.074 [-0.328, 0.18], $t_{60}$ = -0.58, $p$ = .5623, w = .29). Here the model probabilities correspond to models with the moderator parameter compared to models without. Hence, they suggest low probabilities (.28 & .29) that the moderation models represent the approximating model for the data compared to models without moderators.

**Transitions between Risk Categories**

To address the practical implications of the effects of the ROOTS curriculum, we cross-tabulated transitions between risk categories from pretest to posttest for students deemed eligible for ROOTS. The boundary between students at high risk compared to students not high risk was the tenth percentile on the TEMA. The tenth percentile was selected because it corresponds roughly to the percent of the student population that is eventually classified as learning disabled in mathematics (Geary, 2004). In this sample, 50.8% of students, 132 of 260, fell into the high-risk category at pretest. Among the 132 high-risk students, 79.5% (105) transitioned to a lower-risk category. The rates were nearly identical across conditions: 79.1% or 14 of 53 control students and 80.0% or 13 of 52 students who received ROOTS. Incidentally, of the 128 students not at high risk at pretest, one student in each condition transitioned to the high-risk category at posttest.

## Discussion

The present study tested the efficacy of the ROOTS intervention delivered as a supplement to an evidence-based core mathematics curriculum (ELM) that has previously demonstrated efficacy for improving math outcomes for at-risk learners (Clarke et al., 2011). The current investigation was a closely aligned conceptual replication of an earlier study (Clarke et al., 2016), which demonstrated a positive impact on the mathematics achievement of at-risk students when ROOTS was delivered in the context of ELM. Additional studies (Clarke et al., 2017; Clarke et al., 2019; Clarke et al., 2020; Doabler et al., 2019; Doabler et al., 2016) have also demonstrated the efficacy of the ROOTS intervention across various geographic locations, populations, cohorts, and instructional contexts. Similar positive findings in the present study would have increased confidence in the generalizability of ROOTS treatment effects and contributed to a growing body of evidence supporting ROOTS as an evidence-based practice. However, the null findings of the present study contradict the results of Clarke et al. (2016), suggesting instead that the impact of the ROOTS program may be dependent upon other variables.

There were several key differences between schools that participated in the current replication versus the initial study by Clarke et al. (2016). Schools in the replication served a less diverse student population with a smaller proportion of students from economically disadvantaged backgrounds. Schools in the Texas replication sample also had lower base rates for risk in mathematics than the Oregon schools in the initial study. In Oregon, 90% of the sample of students eligible for intervention scored in the high-risk range (below the $10^{th}$ percentile) on the TEMA, whereas in Texas, only 51% of eligible students scored in the high-risk range. Relatedly, a larger percentage of students in the replication sample transitioned to a lower-risk category during the study. Of the 132 Texas students scoring in the high-risk range at pretest, nearly 80%, or 105 students, transitioned to a lower risk category by posttest, with similar distribution across control and intervention conditions (79.1% and 80.0%, respectively). Of the 52 Oregon students scoring in the high-risk range at pretest in the initial study, approximately 51% percent were no longer considered high risk at posttest, with 48% of control group students and 54.9% of students receiving ROOTS intervention transitioning to a lower risk category during the study.

These differences in base rates for risk between samples indicate that the Texas replication sample contained many students who were more skilled in mathematics prior to intervention, whereas at-risk students in the initial Oregon sample were a higher-risk group with lower initial skills in mathematics. The difference in initial skill between the samples is further demonstrated by mean standard scores and percentile rank at pretest with a mean score of 68.9 (percentile rank of 4.6) on the TEMA in the initial Oregon sample compared to a mean TEMA score of 80.2 in Texas (percentile rank of 19.1). Critically, multiple studies have found that initial math skill moderates response to the ROOTS intervention, with students with lower initial skills benefitting more from the intervention than at-risk peers with higher initial skills (Clarke et al., 2019; Clarke et al., 2020). This finding has been replicated in Massachusetts and Oregon with multiple cohorts of students. In this lower-risk replication sample, however, initial math skill was not a significant moderator of response to intervention. Taken together, these results suggest that in educational contexts with lower base rates of risk, when strong, explicit and systematic core math instruction is in place, ROOTS may have less of an impact on the outcomes of at-risk learners, regardless of their initial math skill.

As in the initial study, ELM instruction as a business-as-usual control presents a particularly strong counterfactual for the ROOTS intervention. ELM is a Tier 1 mathematics curriculum that was specifically designed to meet the needs of at-risk learners in ways that typical core curricula do not (Clarke et al., 2011; Doabler et al., 2012). The literature has established several instructional design features that contribute to improved outcomes for students with or at risk for math learning difficulties (e.g., Gersten, Beckmann, et al., 2009; NMAP, 2008). However, reviews of core mathematics curricula (e.g. Bryant et al., 2008; Doabler et al., 2012) indicated that these features are rarely built into Tier 1 math programs, suggesting that typical core instruction is unlikely to meet the needs of at-risk learners. ELM, on the other hand, incorporates research-based instructional design features intended to support student learning across a wide range of skill levels with supports tailored to the needs of at-risk students (e.g., extensive teacher scaffolding that is systematically faded as students move toward mastery). ELM also has a strong focus on developing conceptual understanding of whole numbers, integrating this critical content throughout its scope and sequence. This is consistent with recommendations made by the NMAP (2008) and NCTM (2006) that early math instruction target deep understanding of the most critical content, and contributes to substantial alignment between the ELM and ROOTS programs. ELM has previously demonstrated effectiveness in narrowing math achievement gaps between at-risk learners and their peers by accelerating growth among the lowest performing students and increasing rates of at-risk students transitioning to lower-risk status by the end of their kindergarten year (Clarke et al., 2011).

Furthermore, while intervention dosage was identical across studies, schools participating in the replication study devoted more time to core mathematics instruction than those in the initial study. On average, classrooms in the replication study spent an additional 16 minutes per day in core math instruction, with an average core math block of 68.9 minutes (versus 52.9 minutes in the initial study). This represents a substantial difference (exceeding one standard deviation) in core math block duration. Considering the instructional context and student population of the current replication, the null findings suggest that the added benefit of an intensive intervention program like ROOTS in a lower-risk population may be diminished in the context of strong core instructional practices. In the current replication sample, it is plausible that the combination of an evidence-based core math curriculum designed to meet the needs of at-risk learners and a longer core math instructional block was already adequately addressing the needs of many lower-performing students. The lack of treatment effects observed here may reflect greater growth among at-risk control students rather than less growth among students receiving ROOTS. Pinpointing an exact cause for the different outcomes across studies is difficult and speaks to the need to carefully consider and document the counterfactual (Lemons et al., 2014). Doing so will enable a more nuanced understanding of intervention effects and may serve to drive future research.

The divergent findings and contextual factors associated with the present and earlier studies of ROOTS effectiveness highlight the critical role of replication research within the intervention literature. Addressing questions such as how generalizable and robust an intervention's effects are is an important goal of intervention research within a systematic framework of replication (Coyne et al., 2016). The null findings of the present study are valuable because they contribute to a more nuanced understanding of the specific populations for whom and conditions under which the ROOTS intervention may be most effective. There have been calls in the field to move beyond simply examining whether programs are effective and begin to who they are most effective for, for example, by examining individual-level moderating variables, such as students' initial skill (Fuchs & Fuchs, 2018). Studies designed to systematically examine the role of moderating variables would fit within a broader framework for replication work.

The lack of comparable findings across studies of ROOTS also suggest interesting directions for future research addressing how contextual factors impact interactions among different tiers of support and decisions about matching supports to student needs. For example, the null findings here suggest that in a lower risk sample, when strong core instructional practices are in place, implementing an intensive Tier 2 intervention such as ROOTS may not be the best course of action to meet the learning needs of at-risk students. While designed as a Tier 2 intervention, converging evidence suggests that the ROOTS program best meets the educational needs of students with significant skill deficits in mathematics (Clarke et al., 2019; Clarke et al., 2020). The present study further suggests that the learning needs of at-risk kindergarten students with higher initial math skills may be adequately met when core instructional time is maximized via evidence-based instructional supports that facilitate understanding of core content. In such contexts, supplementing core instruction with a less-intensive, more cost-effective Tier 2 intervention may be sufficient to meet the needs of many at-risk students. Alignment between intensity of student needs and provided services is an essential component of MTSS that optimizes use of resources to enhance student outcomes on a systems-level. Future research could address these important issues by comparing intensive intervention programs, such as ROOTS, with less intensive Tier 2 supports across populations with varying base rates for risk in mathematics. This line of research, in conjunction with moderation research addressing what works for whom, could ultimately help guide school level decision-making about which students should receive which services based on their unique profile of individual and contextual factors.

**References**

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika, 60*(2), 255–265. https://doi.org/10.1093/biomet/60.2.255

Allison, P. D. (2009). Missing data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology.* (pp. 72–89). Sage Publications Ltd. https://doi.org/10.4135/9780857020994.n4

Baker, S. K., Gersten, R. M., & Lee, D.-S. (2002). A synthesis of empirical research on teaching mathematics to low- achieving students. *Elementary School Journal, 103*, 51–73. https://doi.org/10.1086/499715

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *Elementary School Journal, 108*, 115–130. https://doi.org/10.1086/525550

Bryant, D. P., Bryant, B. R., Gersten, R. M., Scammacca, N. N., Funk, C., Winter, A., Shih, M., & Pool, C. (2008). The effects of tier 2 intervention on the mathematics performance of first-grade students who are at risk for mathematics difficulties. *Learning Disability Quarterly, 31*, 47–63. https://doi.org/10.2307/20528817

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. https://doi.org/10.1007/s00265-010-1029-6

Chard, D. J., Baker, S. K., Clarke, B., Jungjohann, K., Davis, K. L. S., & Smolkowski, K. (2008). Preventing early mathematics difficulties: The feasibility of a rigorous kindergarten mathematics curriculum. *Learning Disabilities Quarterly, 31*, 11–20. https://doi.org/10.2307/30035522

Chard, D. J., Clarke, B., Baker, S. K., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30*, 3–14. https://doi.org/10.1177/073724770503000202

Chhin, C. S., Taylor, K. A., & Wei, W. S. (2018). Supporting a culture of replication: An examination of education and special education research grants funded by the Institute of Education Sciences. *Educational Researcher, 47*(9), 594–605. https://doi.org/10.3102/0013189X18788047

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. https://doi.org/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*(2), 152–165. https://doi.org/10.1177/0022219414538514

Clarke, B., Doabler, C. T., Smolkowski, K., Turtura, J., Kosty, D., Kurtz-Nelson, E., Fien, H., & Baker, S. K. (2019). Exploring the relationship between initial mathematics skill and a kindergarten mathematics intervention. *Exceptional Children, 85*(2), 129–146. https://doi.org/10.1177/0014402918799503

Clarke, B., Doabler, C. T., Strand Cary, M., Kosty, D. B., Baker, S. K., Fien, H., & Smolkowski, K. (2014). Preliminary evaluation of a tier-2 mathematics intervention for first grade students: Utilizing a theory of change to guide formative evaluation activities. *School Psychology Review, 43*, 160–177. https://doi.org/10.1080/02796015.2014.12087442

Clarke, B., Doabler, C. T., Turtura, J., Smolkowski, K., Kosty, D., Sutherland, M., Kurtz Nelson, E., Fien, H., & Baker, S. K. (2020). Examining the efficacy of a kindergarten mathematics intervention by group size and initial skill: Implications for practice and policy. *The Elementary School Journal, 121*(1), 125–153. https://doi.org/10.1086/710041

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. https://doi.org/10.1080/02796015.2004.12086245

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. https://doi.org/10.1086/659033

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*, 330–351. https://doi.org/10.1037/1082-989X.6.4.330

Common Core State Standards Initiative. (2010). Common core standards for mathematics. Retrieved 12/15/10, from http://www.corestandards.org/the-standards/mathematics

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2014). CEC’s standards for classifying the evidence base of practices in special education. *Remedial and Special Education*, Advanced online publication. https://doi.org/10.1177/0741932514557271

Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: A framework of systematic, conceptual replications. *Remedial and Special Education, 37*, 244–253. https://doi.org/10.1177/0741932516648463

Coyne, M. D., Kame'enui, E. J., & Carnine, D. (2011). *Effective teaching strategies that accommodate diverse learners* (4th ed.). Pearson Education.

Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Simmons, L. E., Kwok, O.-M., Kim, M., Fogarty, M., Oslund, E. L., Taylor, A. R., Capozzoli-Oldham, A., Ware, S., Little, M. E., & Rawlinson, D. A. M. (2013). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children, 80*, 25–44. https://doi.org/10.1177/001440291308000101

Doabler, C. T., Cary, M. S., Jungjohann, K., Clarke, B., Fien, H., Baker, S., Smolkowski, K., & Chard, D. (2012). Enhancing core mathematics instruction for students at risk for mathematics disabilities. *TEACHING Exceptional Children, 44*(4), 48–57. https://doi.org/10.1177/004005991204400405

Doabler, C. T., Clarke, B., Kosty, D., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2019). Examining the impact of group size on the treatment intensity of a tier 2 mathematics intervention within a systematic framework of replication. *Journal of Learning Disabilities, 52*(2), 168–180. https://doi.org/10.1177/0022219418789376

Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a tier 2 mathematics intervention: A conceptual replication study. *Exceptional Children, 83*(1), 92–110. https://doi.org/10.1177/0014402916660084

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446. https://doi.org/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. https://doi.org/10.1177/0022219411410233

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI publication No. 231). University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Foster, E. M., & Bickman, L. (1996). An evaluator's guide to detecting attrition problems. *Evaluation Review, 20*, 695–723. https://doi.org/10.1177/0193841X9602000603

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

Fuchs, D., & Fuchs, L. S. (2018). On the importance of moderator analysis in intervention research: An introduction to the special issue. *Exceptional Children, 85*(2), 126–128. https://doi.org/10.1177/0014402918811924

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*, 493–513. https://doi.org/10.1037/0022-0663.97.3.493

Gamse, B. C., Bloom, H. S., Kemple, J. J., Jacob, R. T., Boulay, B., Bozzi, L., Caswell, L. J., Horst, M., Smith, W. C., St. Pierre, R. G., Unlu, F., Herlihy, C., Zhu, P., Greene, D., Kim, J., LaLiberty, D., Lam, K., Maree, K., McCormick, R., Miura, J., Unterman, R., & Wong, E. (2008). *Reading First impact study* (Interim report No. NCEE 2008-4016). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Departement of Education. http://files.eric.ed.gov/fulltext/ED501218.pdf

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*, 4–15. https://doi.org/10.1177/00222194040370010201

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). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_math_pg_042109.pdf

Gersten, R. M., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*, 1202–1242. https://doi.org/10.3102/0034654309334431

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. *European Journal of Epidemiology, 31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93*, 615–627. https://doi.org/10.1037/0022-0663.93.3.615

Kame'enui, E. J., & Simmons, D. C. (1999). Beyond effective practices to schools as host environments: Building and sustaining a school-wide intervention model in beginning reading for all children. *Australasian Journal of Special Education, 23*, 100–127. https://doi.org/10.1080/1030011990230204

Kaminski, R. A., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. (2008). Best practices in using dynamic indicators of basic early literacy skills for formative assessment and evaluation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. V, pp. 1181–1204). NASP.

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher, 43*, 242–252. https://doi.org/10.3102/0013189x14539189

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

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*, 304–316. https://doi.org/10.3102/0013189x14545513

Makel, M. C., Plucker, J. A., Freeman, J., Lombardi, A., Simonsen, B., & Coyne, M. (2016). Replication of special education research: Necessary but far too rare. *Remedial and Special Education, 37*, 205–212. https://doi.org/10.1177/0741932516646083

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: how often do they really occur? *Perspectives on Psychological Science, 7*(6), 537–542. https://doi.org/10.1177/1745691612460688

Miller, A. C., Fuchs, D., Fuchs, L. S., Compton, D., Kearns, D., Zhang, W., Yen, L., Patton, S., & Kirchner, D. P. (2014). Behavioral attention: A longitudinal study of whether and how it influences the development of word reading and reading comprehension among at-risk readers. *Journal of Research on Educational Effectiveness, 7*, 232–249. https://doi.org/10.1080/19345747.2014.906691

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306–321. https://doi.org/10.1177/0022219408331037

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. https://doi.org/10.1111/jcpp.12204

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

National Assessment Educational Progress. (2019). *Results from the 2019 mathematics and reading assessments* (NCES No. 2020012). National Center for Education Statistics, U.S. Department of Education, Institute of Education Sciences. https://www.nationsreportcard.gov/mathematics?grade=4

National Council of Teachers of Mathematics. (2006). *Curriculum Focal Points for prekindergarten through grade 8 mathematics: A quest for coherence*. National Council of Teachers of Mathematics. https://www.nctm.org/store/Products/Curriculum-Focal-Points-for-Prekindergarten-through-Grade-8-Mathematics-A-Quest-for-Coherence/

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. US Department of Education.

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

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. National Academies Press. http://www.nap.edu/catalog.php?record_id=12519

National Science Foundation, The Institute of Education Sciences, & U.S. Department of Education. (2018). *Companion guidelines on replication & reproducibility in education research: A supplement to the common guidelines for education research and development*. National Science Foundation, The Institute of Education Sciences, U.S. Department of Education,. https://ies.ed.gov/pdf/CompanionGuidelinesReplicationReproducibility.pdf

Nelson, G., & McMaster, K. L. (2019). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. *Journal of Educational Psychology, 111*(6), 1001–1022. https://doi.org/10.1037/edu0000334

Pro-Ed. (2007). *Test of early mathematics ability- Third edition (TEMA-3)*. ProEd.

SAS Institute. (2016). *Base SAS® 9.4 procedures guide: Statistical procedures* (5th ed.). SAS Institute, Inc.

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

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*, 90–100. https://doi.org/10.1037/a0015108

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. Guilford.

Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M., Stoolmiller, M., Santoro, L., Smith, S., Beck, C., & Kaufman, N. K. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design of instructional specificity. *Journal of Learning Disabilities, 40*, 331–347. https://doi.org/10.1177/00222194070400040401

Sood, S., & Jitendra, A. K. (2013). An exploratory study of a number sense program to develop kindergarten students' number proficiency. *Journal of Learning Disabilities, 46*, 328–346. https://doi.org/10.1177/0022219411422380

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352–360. https://doi.org/10.3102/0013189x14553660

What Works Clearinghouse. (2011). *Procedures and standards handbook. Version 2.1*. Institute of Education Sciences. http://ies.ed.gov/ncee/wwc/references/idocviewer/doc.aspx

What Works Clearinghouse. (2017). *Standards handbook version 4.0*. Institute of Education Science, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Table 1

*Descriptive Statistics for Mathematics Measures by Condition and Assessment Time*

| | | Intervention | | Control | |
|---|---|---|---|---|---|
| Measure | | Pretest | Posttest | Pretest | Posttest |
| TEMA Standard Score | M | 81.22 | 99.07 | 79.05 | 98.11 |
| | (SD) | (17.29) | (12.35) | (16.77) | (13.49) |
| | N | 148 | 143 | 135 | 134 |
| TEMA Percentile | M | 20.82 | 48.90 | 17.14 | 47.34 |
| | (SD) | (24.59) | (24.54) | (22.86) | (26.51) |
| | N | 148 | 143 | 135 | 134 |
| EN-CBM | M | 49.99 | 155.84 | 46.65 | 146.31 |
| | (SD) | (44.32) | (46.06) | (41.30) | (52.53) |
| | N | 149 | 143 | 135 | 134 |

*Note*. TEMA = Test of Early Mathematics Ability. The TEMA percentiles were provided for descriptive purposes only; all analyses used the TEMA standard score.

Table 2

*Fixed Effect and Variance Component Estimates from the Test of Condition on Mathematics Outcomes, with Hedges' g Values for the Time by Condition Effect*

| Effect or Statistic | | TEMA Standard Score | EN-CBM |
|---|---|---|---|
| Model probability (*w*) | | .26 | .33 |
| Fixed effects | Intercept | 78.55 (2.05) | 45.18 (6.36) |
| | Time | 19.30 (1.80) | 100.79 (5.37) |
| | Condition | 2.63 (2.85) | 4.34 (8.84) |
| | Time × Condition | -1.20 (2.50) | 5.91 (7.46) |
| Variances | Classroom-Level Intercept | 68.86 (18.96) | 690.58 (185.21) |
| | Classroom-Level Gains | 27.96 (8.83) | 258.78 (79.89) |
| | Student-Level Intercept | 41.59 (9.47) | 435.90 (86.98) |
| | Residual (Error) | 87.05 (8.63) | 736.15 (73.41) |
| Intraclass correlation | | .24 | .26 |
| Hedges' *g* | Time × Condition | -0.09 | 0.12 |
| 95% CI | | [-0.48, 0.29] | [-0.18, 0.42] |
| *P* value | Time × Condition; | .6312 | .4315 |
| Degrees of freedom | | 60 | 60 |

*Note.* Table entries show parameter estimates with standard errors in parentheses except for model probabilities (Akaike weights), intraclass correlation coefficients (ICCs), Hedges's *g* values, and *p* values. The model probabilities indicate the likelihood of the model that contains the condition effect compared to the model without the condition effect given the data. Time is coded 0 for T1 and 1 for T2. Condition is coded 0 for control and 1 for ROOTS. All tests fixed effects used 27 *df*. In gain-score models, the residual is equivalent to the variance of the student-level gains. The ICC is defined as the variance for the classroom-level gains divided by the sum of the classroom-level and student-level (residual) gains (Murray, 1998).