Research Article

Hammill Institute on Disabilities

Journal of Learning Disabilities
2021, Vol. 54(2) 97–110

DOI: 10.1177/0022219420972185
journaloflearningdisabilities.sagepub.com
SAGE

# Kindergarteners at Risk for Severe Mathematics Difficulties: Investigating Tipping Points of Core Mathematics Instruction

**Christian T. Doabler, PhD[1], Ben Clarke, PhD[2], Derek Kosty, PhD[3], Steven A. Maddox, MEd[1], Keith Smolkowski, PhD[3], Hank Fien, PhD[2], Scott K. Baker, PhD[2,4], and Georgia L. Kimmel, BS[1]**

## Abstract

A concerning number of students enter kindergarten facing an intractable variation of mathematics difficulties (MD). This study investigated the impact of an explicit, core kindergarten mathematics program on the mathematical outcomes of kindergartners who demonstrated risk for severe MD at kindergarten entry and examined whether these students improved from a category of high MD risk (i.e., <10th percentile) to a lower risk of MD (i.e., norm-referenced performance at or above the 10th, 20th, 30th, and 40th percentiles) between the fall and spring of kindergarten. Differential response to the program based on the classroom-level proportion of students with severe MD was also explored. A total of 795 kindergarteners with severe MD from 122 classrooms were included in the analyses. Results suggested students with severe MD in treatment classrooms improved from fall to spring at a greater rate than their control peers. Treatment students also demonstrated higher rates of improvement from below the 10th percentile to a performance at or above the 20th, 30th, and 40th percentiles across the school year. No evidence of differential efficacy of the program by the classroom-level proportion of students with severe MD was found. Implications for using explicit mathematics programs to thwart the onset of severe MD among academically vulnerable students are discussed.

## Keywords

mathematics, multitiered systems of support, intervention, elementary, age

Recently, the National Science & Technology Council, Committee on STEM Education (2018) tasked the field with increasing diversity, equity, and inclusion in science, technology, engineering, and mathematics (STEM). At the forefront of the council's charge is promoting strong foundations in all areas of STEM so that students can achieve overall academic success throughout public school, obtain meaningful postsecondary experiences, and contribute to the nation's STEM workforce. In mathematics, this attention toward STEM literacy is timely as compelling research suggests that a considerable number of students enter kindergarten lacking mathematical readiness and that such skill deficits are consequential for later mathematics learning (Bodvoski & Farkas, 2007; Duncan et al., 2007; Morgan et al., 2009). For example, examining longitudinal data from nearly 8,000 students in the Early Childhood Longitudinal Study—Kindergarten (ECLS-K), Morgan et al. (2009) found that students entering and exiting kindergarten with performances below the 10th percentile on a nationally normed mathematics assessment had a 70% chance of experiencing persistent difficulties in mathematics throughout the later grades. These data clearly suggest that unless educational research and practice focuses on the instructional needs of these struggling learners during the kindergarten year, many will experience low growth trajectories in mathematics and face an eventual pathway of mathematical failure.

[1]The University of Texas at Austin: The Meadows Center for Preventing Educational Risk, Austin, TX
[2]University of Oregon, Eugene, OR
[3]Oregon Research Institute, Eugene, OR
[4]Southern Methodist University, Dallas, TX, USA

**Corresponding Author:**
Christian T. Doabler, PhD, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin, 1 University Station, D5300, Austin, TX 78712, USA.
Email: cdoabler@austin.utexas.edu

The current study focuses specifically on the mathematics achievement of students who begin their kindergarten year performing below the 10th percentile on a standardized mathematics assessment. Aligned with existing research (i.e., Fletcher et al., 2019; Morgan et al., 2009), we considered those students with beginning-of-kindergarten mathematical performances at the bottom 10% being at risk for "severe" mathematics difficulties (MD). Students who exhibit such profound learning problems at the start of kindergarten are at high risk for experiencing a more persistent and intractable type of MD. From a perspective of a multitiered system of support (MTSS; Fuchs & Vaughn, 2012), students who display severe MD in kindergarten are academically vulnerable. Consequently, they have a strong propensity for needing Tier 3 intervention supports and having a later identification of learning disabilities in mathematics (Compton et al., 2012).

Among the considerable number of kindergarten students who are at risk for severe MD, many come from economically and educationally disadvantaged households, and thus do not obtain support at home or in preschool to develop adequate readiness for kindergarten mathematics (Barnes et al., 2016). And unlike many of their peers, including even those who perform in the low average range at the start of kindergarten (e.g., at or below the 25th percentile), students at risk for severe MD enter kindergarten with deep knowledge gaps around number and numeration. Consequently, this lack of "number sense" (Berch, 2005) affects their ability to learn and understand the kindergarten mathematics curriculum. Despite compelling evidence that severe MD in kindergarten leads to long-term difficulties in mathematics (Morgan et al., 2009), few, if any, methodologically rigorous studies have investigated whether and to what extent these at-risk learners respond to the core mathematics instruction delivered in kindergarten classrooms.

## Research on Core Mathematics Instruction in Kindergarten Classrooms

Core (Tier 1) mathematics instruction is integral at all grade levels. In kindergarten, however, core mathematics instruction is of critical importance because it serves as a valuable window of mathematical learning for all students (Doabler et al., 2015; Jordan et al., 2009; Morgan et al., 2009). For many students, the core instruction delivered during the kindergarten year represents their initial exposure to the foundational concepts and skills of early mathematics, such as the principle of cardinality, magnitude comparisons among numerical quantities, and basic number combinations (Geary et al., 2018). Moreover, because the time and resources put toward beginning reading in kindergarten classrooms often result in limited time for mathematics instruction (La Paro et al., 2009), core instruction may represent the totality of mathematics instruction for kindergarten students at risk for severe MD. Therefore, the core mathematics instruction delivered in kindergarten classrooms has important implications for all kindergarten students (Doabler et al., 2015). Not only is core mathematics instruction charged with allowing typical achieving kindergarten students to learn and progress successfully, it is also responsible for establishing a positive trajectory of mathematical learning among those kindergarteners who are at risk for severe MD beginning at the start of the kindergarten year.

Against that backdrop, researchers have begun to investigate different instructional approaches for improving the efficacy of core mathematics instruction in kindergarten classrooms. Explicit instruction is one approach that shows promise for a range of learners in kindergarten classrooms. For example, Sood and Jitendra (2013) investigated the impact of an explicitly designed, Tier 1 kindergarten mathematics program focused on promoting students' knowledge of early number sense concepts. Results suggested significant differences in student mathematics achievement favoring the number sense intervention program relative to the comparison condition. Reported effect sizes (Hedges' $g$) ranged from 0.55 to 1.44.

In a more recent randomized controlled trial, Clarke et al. (2015) investigated the efficacy of the Early Learning in Mathematics (ELM) program. ELM is a yearlong, core (Tier 1) kindergarten mathematics program that is delivered in whole-class settings and aimed at supporting students' development of mathematical proficiency with concepts of whole numbers, measurement, and geometry identified in the Common Core State Standards for Mathematics (CCSS-M, Common Core State Standards Initiative, 2010). The program's lessons are invariably grounded in empirically validated principles of explicit mathematics instruction (Coyne et al., 2011; Gersten et al., 2009). In this way, the program systematically facilitates explicit instructional interactions around critical mathematics content. Such interactions are composed of teachers offering overt demonstrations of new mathematical content, facilitating guided and independent practice opportunities for students to develop mathematical proficiency, and providing timely academic feedback to confirm student responses and address potential misconceptions.

Participating in the ELM Efficacy Trial were 2,598 kindergarten students from 129 kindergarten classrooms (Clarke et al., 2015). In the study, Clarke et al. randomly assigned the 129 kindergarten classrooms, blocking on schools, to either treatment or control conditions. Classrooms in the treatment condition implemented the ELM program, whereas control classrooms continued to use standard district practices (business-as-usual). Results indicated that ELM classrooms did not significantly differ from control classrooms ($g$ = 0.11). Findings did, however, reveal evidence of a moderation effect, suggesting that students who began the kindergarten year below the 25th percentile on the Test of Early Mathematics Ability—Third Edition (TEMA-3; Ginsburg & Baroody, 2003) demonstrated the strongest gains across the school year. Clarke et al. also reported that treatment students who tested above the 25th percentile on the pretest distribution of the TEMA-3 remained "on track" for developing mathematics proficiency relative to their control peers (>25th percentile).

While findings from Clarke et al. (2015) suggested a pattern of differential response among students considered at "some" risk for MD (i.e., <25th percentile), it did not fully establish the impact of the ELM program on the mathematics achievement of kindergartners at risk for severe MD (i.e., <10th percentile). For example, it did not examine whether ELM enabled accelerated learning gains for these students such that their mathematical performance improved beyond a threshold of severe MD by the end of kindergarten. Moreover, Clarke et al. (2015) did not investigate the composition of kindergarten classrooms to determine whether and to what extent the proportion of kindergarten students at risk for severe MD in a kindergarten classroom might moderate the effects of the ELM program. Recognizing that schools have few resources available to provide instructional support in kindergarten mathematics beyond Tier 1 educational settings, it seems paramount to examine and unpack the potential benefits of explicit, core mathematics instruction for kindergartens who enter school facing severe MD.

## Explicit Instruction and Tipping Points of Efficacy

Despite the potential benefits of explicitly designed core mathematics instruction (Gersten et al., 2009), it is likely that this instructional approach may reach a point at which its effects on student mathematics achievement begin to slow or completely level off (Doabler et al., 2019). In kindergarten classrooms, this "tipping point" (Institute of Medicine, 2010; Lenton et al., 2008) may be a function of the proportion of students who perform well below average in a given classroom. Thus, a large proportion of students with severe MD in a given classroom may stifle a teacher's ability to deliver effective core mathematics instruction, such as facilitating the necessary number of mathematics verbalizations required to build proficiency with early mathematical concepts.

Interestingly, these tipping points can be present even when teachers are equipped with empirically validated, explicitly designed core academic programs. In the area of reading, for example, Vaughn et al. (2017) found that once the classroom-level proportion of middle school English learners with reading difficulties exceeded 10%, the effects of an explicit, evidence-based reading comprehension program began to diminish.

The work of Vaughn and colleagues (2017) on tipping points based on classroom composition was particularly cogent in that it helped shed light on a blank spot in the empirical literature on mathematics intervention research. We know of no mathematics intervention studies that have investigated whether the classroom-level percentage of kindergarten students at risk for severe MD at the start of the school year influences the yearlong impact of core mathematics instruction that teachers provide in kindergarten classrooms. This is surprising, given that core mathematics instruction in kindergarten plays a pivotal role in supporting students' development of mathematical proficiency (Morgan et al., 2009).

Similar to the findings of Vaughn et al. (2017), large proportions of students with severe MD in kindergarten classrooms may mitigate teachers' capacity to deliver highly effective core mathematics instruction even when they are equipped with empirically validated, core mathematics programs. When the classroom-level proportion of students with severe MD crests a particular threshold (e.g., >.25), teachers may be forced to slow the pace of instruction, reteach concepts, and limit the number of instructional interactions for students to engage in critical mathematics content. For example, students may have limited time to work with concrete manipulatives and engage in meaningful mathematical discourse. Thus, establishing the point at which the effects of core mathematics instruction on student mathematics achievement slow or completely level off as a function of the classroom-level proportion of kindergarten students at risk for severe MD could help pinpoint when additional classroom-level supports (e.g., Tier 2 or 3) are necessary to accelerate the mathematical learning of students who enter kindergarten significantly "off track" for developing mathematical proficiency. This study aimed to identify these particular thresholds.

## Purpose of the Study

The current study sought to expand the work of Clarke et al. (2015) by specifically examining the impact of the ELM mathematics program on the mathematical outcomes of students who entered kindergarten at risk for severe MD (i.e., <10th percentile on the TEMA-3). In addition, we were interested in the notion of whether a well-designed and delivered core mathematics program could accelerate students' mathematical performance to a point where students who face severe MD at the beginning of kindergarten get "on track" for developing mathematics proficiency by the end of the school year. It was hypothesized that the rate of improvement across the school year for students in ELM classrooms at risk for severe MD at pretest would exceed that of students at risk for severe MD in control classrooms.

Progress on a norm-referenced mathematics outcome measure from a category of at risk for severe MD (<10th percentile) at the start of kindergarten to at or above the 10th, 20th, 30th, and 40th percentiles at the end of kindergarten served as targeted indices of improvement. Finally, we examined whether and to what extent the classroom-level proportion of kindergarten students with at risk for severe MD moderated the effects of the ELM program. As with other core instructional programs (e.g., Vaughn et al., 2017), the efficacy of ELM may begin to depreciate once the proportion of students at risk for severe MD reaches a particular magnitude. Three research questions were addressed:

**Research Question 1:** What is the effect of the ELM program on the mathematics achievement of students at risk for severe MD at the start of the school year?
**Research Question 2:** To what extent does ELM accelerate student mathematics performance above various MD thresholds (i.e., norm-referenced performance at or above the 10th, 20th, 30th, and 40th percentiles)?
**Research Question 3:** Does the classroom-level proportion of students at risk for severe MD moderate the effect of ELM on the outcomes of students who face severe MD at the beginning of kindergarten?

## Method

### Research Design and Database

This study analyzed data collected during a federally funded efficacy trial aimed at investigating the impact of the ELM kindergarten mathematics program (Clarke et al., 2015). One study of the larger ELM Efficacy Trial was conducted during the 2008 to 2009 and 2009 to 2010 school years in two different geographical regions of the United States (i.e., Pacific Northwest and South Central). Blocking on schools, 129 kindergarten classrooms were randomly assigned to either treatment (ELM; $n = 68$) or control (district-approved kindergarten mathematics instruction; $n = 61$) conditions. In the aggregate, the original sample included 2,598 kindergarten students attending 129 classrooms in 46 schools, most of which were eligible for Title 1 funding.

The current study sought to extend the line of research involving the ELM program by utilizing a subset of the 2,598 kindergartners from Clarke et al. (2015) to focus specifically on students at risk for severe MD. As such, the present study included only those treatment and control classrooms that enrolled students who tested below the 10th percentile at pretest on the TEMA-3. From the original sample of 129 kindergarten classrooms, 7 classrooms (2 treatment, 5 control) were dropped because they did not include students with severe MD at the start of kindergarten. In total, the analytical sample for the current study included 122 kindergarten classrooms with 2,454 students, of which 32% ($n = 795$) were considered as facing severe MD based upon scoring at or below the norm-referenced 10th percentile on the TEMA-3 upon entering their kindergarten year. Data analyzed in the current study included student mathematics outcome data collected from the 795 students at risk for severe MD, documenting their gains in mathematics achievement from the beginning to the end of kindergarten.

### Teacher and Student Sample

The 122 classrooms (66 treatment, 56 control) were from 45 schools located in three Pacific Northwest school districts and four South Central school districts. Teachers in treatment classrooms delivered the ELM program, whereas control classrooms continued to offer standard district kindergarten mathematics instruction (business-as-usual). Of the 122 classrooms, 96 were located in public schools, 9 were in charter public schools, and 17 were in private schools. All charter and private school classrooms were located in the South Central school districts. Public school classrooms were located in schools eligible for Title 1 funding. Table 1 provides descriptive information about the classrooms by condition. The majority of classrooms (87%) provided a full-day kindergarten program versus a half-day program. All half-day classrooms were located in the Pacific Northwest school districts. Teachers delivered mathematics instruction in English. Average class size for treatment and control classrooms was $M = 21.3$ ($SD = 3.6$) and $M = 20.1$ ($SD = 3.7$), respectively. A total of 123 teachers taught in the 122 participating classrooms. One control classroom had two teachers, each working a half-day schedule. All teachers participated for the duration of the ELM Efficacy Trial.

The 795 kindergarten students at risk for severe MD at the start of kindergarten were nested in the 122 classrooms. Students were determined as facing severe MD based upon a pretest performance below the 10th percentile on the TEMA-3. Of the 795 students, 447 and 348 were in treatment and control classrooms, respectively. The number of students at risk for severe MD in treatment classrooms ranged from 1 to 18, while the range in control classrooms was 1 to 15. As shown in Table 1, students facing severe MD in both conditions were comparable in terms of gender, race, and percentage identified for special education. Treatment and control classrooms also had equivalent proportions of students at risk for severe MD at pretest.

### ELM Kindergarten Mathematics Program

ELM is a core kindergarten mathematics program that consists of four quarterly teacher manuals, each containing 30 daily lessons. Mathematics content is systematically introduced, reviewed, and extended through ELM's explicit

**Table 1.** Descriptive Information of Students and Classrooms by Condition.

| Variable | ELM | Control |
|---|---|---|
| Student characteristics | | |
| Number of students, *n* | 447 | 348 |
| Male, *n* (%) | 225 (50) | 163 (47) |
| Hispanic, *n* (%) | 277 (74) | 218 (76) |
| Eligible for special education, *n* (%) | 39 (9) | 35 (10) |
| Classroom characteristics | | |
| Number of classrooms, *n* | 66 | 56 |
| Number of students at risk for severe mathematics difficulties at pretest, *M* (*SD*) | 6.8 (3.9) | 6.2 (3.5) |
| Proportion of students at risk for severe mathematics difficulties at pretest, *M* (*SD*) | 0.38 (0.21) | 0.38 (0.22) |

*Note.* ELM = Early Learning in Mathematics.

instructional design framework. For example, each manual offers scripted guidelines to support teachers in demonstrating key mathematics content, delivering timely academic feedback, and facilitating frequent student practice opportunities, including structured verbal interactions between teachers and students, and among students, around key mathematics content. Such practice opportunities are systematically designed to help students build mathematical proficiency, and develop mathematical language and vocabulary that students would not otherwise acquire on their own. ELM promotes such vocabulary development by allowing students to (a) hear and practice the correct pronunciation of key mathematical terms, (b) understand the meaning of targeted words in multiple contexts, and (c) verbally apply their understanding of the target words by using them in complete sentences. To build conceptual understanding, lessons incorporate frequent opportunities for students to work with visual representations of mathematical ideas, such as three-dimensional (3D) shapes, counting blocks, and number lines.

Mathematics domains targeted in ELM include (a) counting and cardinality, (b) operations and algebraic thinking, (c) number and operations in base 10, (d) measurement and data, and (e) geometry. Daily lessons last approximately 45 min in duration and include (a) whole-class and small-group activities focused on new mathematical content, (b) judicious review of previously learned material, and (c) worksheet activities that provide students extended practice with previously taught concepts and skills. Problem-solving activities are introduced every five lessons to help students practice newly acquired problem-solving skills and engage in complex mathematical problems, such as collecting categorical data and representing the data on a graph. The 66 treatment teachers implemented the ELM program 5 days per week in general education classrooms.

*Professional development.* Treatment teachers received four professional development workshops related to the program implementation. Each workshop lasted 6 hr and corresponded with the ELM quarterly teacher manuals. For example, the first workshop was conducted prior to the start of the school year and focused on Lessons 1 to 30. Each workshop centered on evidence-based principles of mathematics instruction and the instructional design and delivery features of the ELM program. Workshops also offered ELM teachers opportunities to practice with sample lessons and receive feedback from the ELM curriculum team.

*Treatment fidelity.* The ELM Efficacy Trial planned to assess fidelity of implementation via real-time direct observation three times per project year (i.e., fall, winter, and spring) in each of the 66 treatment classrooms. Of the total fidelity observations scheduled, 90% were conducted. Across the 2 years, a total of 179 curriculum-specific fidelity checks were conducted in the treatment classrooms. Trained project staff conducted the fidelity observations. In each project year, observers received training across three sessions, including an initial training and two follow-up ones to help minimize observer drift and increase interobserver reliability. Training focused on kindergarten mathematics instruction and procedures associated with the ELM direct observation system, including the fidelity of the implementation arm. Prior to observing independently, observers were required to complete a video reliability checkout and a real-time classroom checkout with a trained research team member. On 54 occasions, two observers collected data simultaneously in the same classrooms to assess interobserver reliability.

For each activity within an ELM lesson, teachers' adherence to the curriculum was rated on a scale ranging from 0 (did not implement), 0.5 (partial implementation), to 1.0 (full implementation). Observers based the level of implementation on the extent to which teachers applied five instructional features for each observed ELM activity. These features included (a) addressing the targeted learning objectives, (b) following the teacher scripting, (c) using the prescribed mathematics visual representations, (d) offering the intended student practice opportunities, and (e) providing timely academic feedback. A rating of "partial" implementation indicated that teachers' use of these features during the activity was observed approximately 50% of the time of implementation, whereas a rating of "full" was present approximately 80% or more of the time. For example, observers would note a "full" level of implementation rating for an activity called the "More and Less" game (ELM Lesson 70), if the teacher (a) explicitly stated the rules of the game, (b) offered a brief demonstration of the game using a student volunteer as an example game partner, (c) directly taught the three targeted vocabulary words ("more," "less," and "same"), and (d) provided necessary academic feedback as students played the game. As reported in Clarke et al. (2015), overall, ELM teachers implemented the curriculum with moderate levels of implementation fidelity ($M = 0.87$, $SD = 0.14$) and no evidence of contamination between ELM and comparison classrooms was observed.

## Control "Business-as-Usual" Classrooms

Mathematics instruction in the 56 control classrooms consisted of standard district (i.e., business-as-usual) mathematics practices. Teachers in control classrooms used a variety of instructional materials, including teacher-developed activities and a number of commercially available programs. Surveys indicated that mathematics materials used in control classrooms varied within participating districts and schools. The most widely used programs were *Everyday Mathematics, Houghton Mifflin, Scott Foresman, Texas Mathematics*, and *Bridges in Mathematics*. According to Agodini and Harris (2010), these programs would be considered more "student centered." In that respect, they task teachers with facilitating whole-class discussions to introduce new concepts and skills rather than using overt explanations and demonstrations, posing more open-ended questions instead of more structured ones, and promoting more conceptual knowledge rather than a blend of conceptual understanding and procedural fluency. Student-centered programs are also less likely to teach for learning mastery than programs that employ a more explicit instructional approach. Despite the student-centered billing, these same control classrooms were found to utilize, at least on occasion, features of explicit instruction. For example, direct observations revealed that control classrooms, when contrasted with ELM classrooms, provided a comparable rate of teacher demonstrations and a similar proportion of student practice followed by teacher corrective feedback; however, they offered statistically lower rates of group and individual student practice opportunities (Doabler et al., 2014). Past research also noted that ELM and control classrooms did not differ by instructional quality (Doabler et al., 2014). Outside of the delivery of instruction, control classrooms varied in terms of their instructional focus. Some control classrooms emphasized whole number concepts, while others focused primarily on patterning and particular aspects of geometry and measurement. A variety of different mediums were employed to deliver instruction in the control classrooms, including whole-class and center-based activities.

## Student Mathematics Outcome Measures

All participating students were administered two mathematics outcome measures at the start (fall) and end (spring) of their kindergarten school year. Trained staff administered the measures, with data collection meeting acceptable reliability criteria (i.e., implementation fidelity of .95 or higher). Our analyses focused on students' pretest and posttest performances, with pretesting occurring just prior to the start of kindergarten for participating students and post-testing taking place at the end of the school year.

*TEMA-3.* The TEMA-3 (Ginsburg & Baroody, 2003) is a 72-item norm-referenced, general outcome measure purported to measure students' procedural and conceptual knowledge of early number sense. The TEMA-3 assesses mathematical understanding at the formal and informal levels for children ranging in age from 3 to 8 years 11 months. Internal consistency reliabilities of the measure exceed .92, and alternate-form and test–retest reliabilities exceed .80. Concurrent validity coefficients with four commonly used assessments of mathematics ranged from .55 to .91. For classroom-level reliability, the intraclass correlation coefficient (ICC) for classrooms for the pretest TEMA-3 was .26, and the average classroom reliability of pretest TEMA-3 across all 122 classrooms was .85. The TEMA-3 provides age norms to calculate standard scores and percentile ranks.

*Early Numeracy Curriculum-Based Measurement (EN-CBM).* EN-CBM (Clarke & Shinn, 2004) is a set of four fluency-based measures (1-min each) of early number sense: oral counting, number identification, quantity discrimination, and strategic counting with strings of numbers. The Oral Counting measure requires students to orally rote count as high as possible and the discontinue rule applies after the first counting error. The Number Identification measure requires students to orally identify numbers between 0 and 10 when presented with a set of printed number symbols. Quantity Discrimination requires students to name which of the two visually presented numbers between 0 and 10 is greater. 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. In this study, the total score on the EN-CBM, as computed as the sum across the four measures, was used in the analyses. Prior research reported a predictive validity coefficient of $r = .81$ between an EN-CBM total score and the TEMA-3 (Clarke et al., 2015).

### Statistical Analysis

For Research Question 1, we assessed intervention effects on the TEMA-3 and EN-CBM raw scores of students with severe MD using a mixed-model (multilevel) Time × Condition analysis (Murray, 1998) to account for the intraclass correlation associated with students nested within classrooms, the level of random assignment. The analysis tested differences between conditions on change in outcomes from the fall to spring of kindergarten, with gains for individual students clustered within classrooms. The statistical model included time, condition, and the Time × Condition interaction, with time coded 0 in the fall and 1 in the spring of kindergarten and condition coded 0 for control and 1 for ELM. Analyses were based on 122 classrooms that included at least one student with severe MD at the start of the school year.

For Research Question 2, we examined rates by study condition at which norm-referenced student TEMA-3 scores increased from below the 10th percentile to at or above the 10th, 20th, 30th, and 40th percentiles between the fall and spring of kindergarten. Specifically, we created four binary variables indicating whether student performance increased beyond each threshold. These variables form non-mutually exclusive categories, as a student who scored at or above the 40th percentile also scored at or above the 20th by definition. We tested condition differences in these binary indicators of achievement gains using contingency table analyses and chi-square tests.

Finally, for Research Question 3, we explored differential response to the ELM program as a function of the classroom-level proportion of students with severe MD. We expanded the statistical model described for Research Question 1 to include this potential moderator and its interaction with condition, time, and the Time × Condition term, resulting in a three-way interaction, all corresponding two-way interactions, and individual (conditional) effects. The three-way interaction of the moderator, time, and condition provided an estimate of whether condition effects varied by the proportion of students at risk for severe MD.

*Model estimation.* We fit multilevel statistical models to our data using SAS PROC MIXED version 14.2 (SAS Institute Inc, 2016) and 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), although nonrandom missingness "is often not sufficient to affect the internal validity of an experimental study to any practical extent" (Graham, 2009, p. 568). In the present study, we did not believe that 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 families moving).

The models assume independent and normally distributed observations. 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).

*Effect sizes.* To interpret condition differences, we computed effect sizes, the Hedges' $g$ for continuous measures and Cox's $d$ for dichotomous measures, using What Works Clearinghouse (WWC, 2020) procedures.

## Results

Herein, we summarize demographic information and present results for the comparison between study conditions (ELM versus control) among the subgroup of students with severe MD at the start of the school year. We then present results for condition differences moderated by the classroom-level proportion of students at risk for severe MD.

### Descriptive Results and Baseline Equivalence

Demographic characteristics are reported in Table 1. ELM and control classrooms did not meaningfully differ on proportions of students who were male, Hispanic, or eligible for special education services ($d < 0.10$). Table 2 reports the descriptive statistics for TEMA-3 and EN-CBM scores by assessment time and intervention condition.

### Attrition

The overall rate of missing TEMA-3 data at posttest was 12.8%, and the difference in rates of missingness between conditions was 1.0%. "The proportions of the treatment and control groups that provide information are not particularly important, at least for internal validity" (Foster & Bickman, 1996, p. 698), so we tested the potential for *differential attrition effects* to identify potential threats to internal validity. To do so, we conducted a mixed-model analysis of variance designed to test whether attrition differentially affected condition differences for outcome variables. Specifically, the analyses tested the association between pretest outcome and (a) study condition (ELM versus control), (b) attrition status, and (c) the interaction between the two (Graham & Donaldson, 1993). We found no interaction between attrition and condition that predicted baseline outcome large enough to suggest that attrition threatened internal validity ($p = .5215$).

**Table 2.** Descriptive Statistics for TEMA-3 and EN-CBM Scores by Time and Condition.

| | | Fall | | Spring | |
|---|---|---|---|---|---|
| Measure or statistic | | ELM | Control | ELM | Control |
| TEMA-3 | *M* | 8.67 | 8.87 | 25.92 | 23.59 |
| | (*SD*) | (4.29) | (3.99) | (8.42) | (7.54) |
| | *n* | 386 | 295 | 380 | 306 |
| EN-CBM | *M* | 29.94 | 28.54 | 125.87 | 113.65 |
| | (*SD*) | (22.92) | (23.74) | (47.74) | (45.68) |
| | *n* | 443 | 342 | 379 | 306 |

*Note*. TEMA-3 = Test of Early Mathematics Ability—Third Edition; EN-CBM = Early Numeracy Curriculum-Based Measurement; ELM = Early Learning in Mathematics.

**Table 3.** Efficacy and Moderation Results from Mixed-Model Time × Condition Analyses of Fall-to-Spring Gains in TEMA-3 and EN-CBM Scores.

| | | Efficacy results | | Moderation results | |
|---|---|---|---|---|---|
| Effect or statistic | | TEMA-3 | EN-CBM | TEMA-3 | EN-CBM |
| Fixed effects | Intercept | 8.84*** (0.49) | 29.10*** (2.69) | 8.79*** (0.49) | 28.57*** (2.76) |
| | Time | 15.02*** (0.63) | 86.21*** (3.39) | 14.87*** (0.65) | 86.09*** (3.49) |
| | Condition | −0.15 (0.66) | 1.43 (3.63) | −0.22 (0.66) | 1.27 (3.73) |
| | Time × Condition | 2.23** (0.84) | 8.63 (4.59) | 2.43** (0.87) | 9.38~ (4.74) |
| | Moderator (class-level proportion of students with significant MD) | | | −2.29 (2.33) | −0.04 (118.00) |
| | Moderator × Condition | | | −1.16 (3.21) | −3.56 (17.82) |
| | Moderator × Time | | | −2.67 (2.97) | −2.70 (15.84) |
| | Moderator × Time × Condition | | | 3.75 (4.13) | 15.77 (22.41) |
| Variances | Classroom intercept | −0.12 (1.14) | 27.48 (36.53) | −0.27 (1.12) | 27.05 (37.07) |
| | Classroom gains | 5.50*** (1.37) | 162.68*** (40.62) | 5.48*** (1.39) | 164.55*** (41.48) |
| | Student | 13.12*** (1.69) | 382.28*** (47.13) | 12.97*** (1.68) | 380.70*** (47.10) |
| | Residual | 22.68*** (1.44) | 743.38*** (42.98) | 22.75*** (1.44) | 744.02*** (43.04) |
| ICC (ρ) | Classroom gains | .195 | .180 | .194 | .181 |
| Hedges' *g* | Time × Condition | 0.28 | 0.18 | | |
| *p* values | Time × Condition | .0095 | .0623 | | |
| | Moderator × Time × Condition | | | .3657 | .4831 |

*Note.* Table entries show parameter estimates with standard errors in parentheses. Tests of fixed effects used 120 degrees of freedom. TEMA-3 = Test of Early Mathematics Ability—Third Edition; EN-CBM = Early Numeracy Curriculum-Based Measurement; MD = mathematics difficulties; ICC = intraclass correlation coefficient.
$^{**}p < .01$. $^{***}p < .001$. $\sim p < .10$.

## *ELM Efficacy for Students With Severe MD*

We tested the hypothesis that students at risk for severe MD in ELM classrooms would experience greater gains in TEMA-3 and EN-CBM scores during kindergarten than students at risk for severe MD in control classrooms. The *g* and *p* values reported in the left columns of Table 3 represent the tests of ELM efficacy. Students at risk for severe MD in ELM classrooms made greater gains from fall to spring than students in the control condition on the TEMA-3 ($g = 0.28$, 95% confidence interval [CI] = [0.07, 0.49]) and EN-CBM ($g = 0.18$, 95% CI [−0.01, 0.38]).

Contingency table analyses of norm-referenced TEMA-3 scores revealed that higher rates of students in ELM classrooms scored at or above the 20th (54% vs. 41%; $\chi^2$ [1, 795] = 11.88, $p = .0006$, $d = 0.32$, odds ratio [OR] = 1.64), 30th (44% vs. 32%; $\chi^2$ [1, 795] = 11.95, $p = .0005$, $d = 0.31$, OR = 1.67), and 40th (29% vs. 20%; $\chi^2$ [1, 795] = 9.95, $p = .0016$, $d = 0.30$, OR = 1.71) percentiles at the end of the school year than their control peers. Rates at which student TEMA-3 scores increased to at or above the 10th percentile did not differ between ELM and control conditions (68% vs. 64%, respectively; $\chi^2$ [1, 795] = 1.31, $p = .2519$, $d = 0.11$, OR = 1.19).

### *Moderation by Class-Level Proportion of Students With Severe MD*

The right column of Table 3 presents tests of differential response to ELM as a function of the classroom-level proportion of students at risk for severe MD at the start of the school year. Tests of moderation require additional fixed effects of the moderator and its interaction with condition, time, and the Time × Condition term. This three-way interaction provided no statistical evidence of differential efficacy of the ELM program by the classroom-level proportion of students at risk for severe MD for the TEMA-3 ($p = .3657$) and EN-CBM outcomes ($p = .4831$).

## Discussion

Today's kindergarten classrooms are becoming increasingly diverse with students who face severe difficulties in mathematics at the start of their kindergarten year. While many of these knowledge gaps may be attributable to students' lack of informal learning opportunities in mathematics prior to school entry, the implications for core mathematics instruction are visible.

The purpose of this study was to extend the work of Clarke et al. (2015) by examining the impact of a core, explicitly designed mathematics program on the mathematical outcomes of students who entered kindergarten with severe MD. Three research questions were addressed.

### *Efficacy of the ELM Mathematics Program*

First, we tested the efficacy of the ELM core mathematics program on the mathematics achievement of students who performed below the 10th percentile on the TEMA-3 at the start of kindergarten. We hypothesized that these students would reap positive benefit from ELM based on its explicit design. Our findings confirmed this hypothesis. Results suggested ELM had a substantively positive effect for students at risk for severe MD on the TEMA-3, producing an effect size of 0.28 (Hedges' $g$). Encouragingly, we found a moderate effect ($g = 0.18$) on the EN-CBM; however, the 95% CI around the point estimate included zero [−0.01, 0.38]. Overall, these findings begin to complement the growing line of research that suggests explicit, core mathematics instruction is beneficial for students who struggle to develop mathematical proficiency in the early grades (Agodini & Harris, 2010; Sood & Jitendra, 2013). Thus, delivering explicit mathematics programs at the preventive instructional tier in an MTSS model may result in thwarting the onset of severe MD among the most academically vulnerable students.

Because the current study involved only one core mathematics program, additional research is needed to investigate the impact of other core programs on the mathematics outcomes of students with severe MD. Over the past 10 years, significant development efforts have been made in the area of mathematics. For example, to date, the WWC (n.d.) has examined more than 20 elementary mathematics programs (i.e., kindergarten to fifth grade). While only a few have demonstrated positive or potentially positive effects on student mathematics achievement, future research is still warranted that examines whether these core programs and those that are still in the development pipeline improve the mathematics achievement of students with severe MD.

### *Rates of Improvement Beyond Thresholds of Severe MD*

Our second research question examined whether ELM accelerated student mathematics performance above various MD thresholds (i.e., performance at or above the 10th, 20th, 30th, and 40th percentiles). It was hypothesized that the rate of improvement across the school year for students at risk for severe MD in ELM classrooms would exceed that of students facing severe MD in control classrooms. Findings indicated the odds of improving from below the 10th percentile at the start of kindergarten to at or above the 20th, 30th, and 40th percentiles at the end of kindergarten were nearly two times higher for students in ELM classrooms than their control peers.

Findings from our second research question suggest systematically designed and explicitly delivered core mathematics instruction can accelerate the mathematical performance of students at risk for severe MD beyond a threshold of high risk by the end of the school year. While there is chance that ELM did not fully meet the instructional needs of these students, our results provide preliminary support at least for well-designed and delivered core mathematics instruction positioning those students with some of the most significant mathematical needs on a positive learning trajectory in mathematics. This is critical as being "on track" for mathematical success at the end of kindergarten significantly increases the likelihood of experiencing positive mathematics outcomes in the later grades (Claessens & Engel, 2013).

We contend that the implications of this finding are significant for schools, particularly those looking to adopt core elementary mathematics programs (Doabler et al., 2018). In many early elementary classrooms, mathematics intervention supports are often at a premium. This is typically due to extra resources being devoted toward accelerating reading outcomes among at-risk learners. Consequently, core mathematics instruction is left to serve as the lone source of instruction in early mathematics. For schools, this highlights the critical importance of selecting and implementing a well-designed core mathematics program, particularly one that incorporates explicit and systematic instructional design principles. As evidenced in the current study, an explicit core mathematics program has the capacity to move students from a category of high risk to on track for developing mathematics proficiency. While early elementary classrooms will likely still require intervention supports beyond Tier 1 mathematics, core programs like the ELM curriculum can substantially alleviate the pressure of mathematics intervention services to act as a backstop to mathematics failure. Moreover, validated core programs can help the field, particularly school administrators, save precious intervention resources and deploy them in ways that are targeted and cost-effective for at-risk learners.

### *Moderation by Class-Level Proportion of Students With Severe MD*

For our third research question, we investigated whether the classroom-level proportion of students with severe MD moderated the effects of core mathematics instruction delivered in kindergarten classrooms. A growing line of empirical research suggests that explicitly designed, core mathematics instruction in kindergarten classrooms offers a plausible mechanism to promote positive student outcomes for students, particularly students who enter school at risk for long-term MD (Clarke et al., 2011; Sood & Jitendra, 2013). Yet, as in the area of reading (e.g., Vaughn et al., 2017), the effectiveness of empirically validated, core mathematics program is likely susceptible to reaching a tipping point. That is, as the classroom-level proportion of kindergarten students at risk for severe MD increases in a given classroom, it may suppress a teacher's capacity to deliver effective core mathematics instruction.

While we hypothesized that ELM would demonstrate a high threshold for the classroom-level percentage of students at risk for severe MD, given the program's explicit and systematic instructional design, it was also expected that the program would exhibit an observable tipping point. That is, we figured that once the classroom proportion of students at risk for severe MD exceeded a certain magnitude, the effect of ELM would begin to diminish. Interestingly, our findings indicated otherwise. Nonsignificant moderation effects for the proportion of students with severe MD were reported for the TEMA-3 and EN-CBM measures, suggesting effects of the ELM program were comparable, regardless of classroom composition of students' initial skill level in mathematics.

These findings can be interpreted in at least two ways. On one hand, it may be that ELM's explicit instructional design was able to support teachers in successfully managing classrooms that contained students with severe MD. When core mathematics programs are engineered to embrace a systematic and explicit instructional framework, they have particular design features that differentiate them from other programs, such as ones that utilize more student-centered instructional approaches. One key design feature of explicit mathematics programs is the incorporation of scaffolds or temporary supports to promote a high success rate with new and complex mathematics content. For example, scaffolds in the ELM program include carefully sequenced instructional examples and judicious review of previously learned content. Another design feature that sets explicit mathematics programs apart from other programs is their capacity to support teachers in (a) offering overt demonstrations and explanations of new mathematical content, (b) providing specific academic feedback to confirm student responses and address potential misconceptions, and (c) facilitating important practice opportunities for students to demonstrate their mathematical thinking and reasoning, such as mathematics verbalizations. In the ELM program, such practice opportunities not only allow struggling learners to build mathematical proficiency but also to collaborate and learn from their typically achieving peers. Combined, these design features may allow kindergarten teachers to meet the instructional needs of students who are at high risk for failure in mathematics.

On the other hand, far less is known about the extent and severity of MD for kindergarten students than reading disabilities among middle school students (Vaughn et al., 2017). Therefore, one could argue that while students in the current sample appeared to demonstrate symptoms of risk for "severe" MD at the start of kindergarten (i.e., performing $<$10th percentile), their learning difficulties are less entrenched than reading disabilities at the middle school level. Thus, it may be easier to reach and address the varying needs of kindergartners at risk for severe MD than the needs of older students who have faced a longer line of academic failure and frustration. Consequently, it may prove difficult to identify a visible tipping point of explicit programs in the early elementary grades (Doabler et al., 2019). Regardless, we encourage the field to continue investigating this line of tipping point research.

For example, while the current study focused specifically on students experiencing severe MD, future research should consider investigating whether other student-related factors at the classroom level serve as potential moderators of explicit mathematics programs. For example, the number of students identified with behavior disorders in a given classroom may influence a teacher's ability to effectively implement an evidence-based core mathematics program. In these situations, teachers may spend more time redirecting or attending to nonacademic, off-task behaviors than delivering the program with fidelity. Future research that includes larger student samples with diverse categories of disabilities may be needed to gain a deeper understanding for how the student composition of mathematics classrooms influences core instruction.

### *Limitations*

When interpreting the findings from this study, a number of limitations should be considered. First, the study included only 795 participating students. Tests of moderation may require a larger sample for including classroom-level moderating variables, such as the proportion of students at risk for severe MD in a kindergarten classroom. Relatedly, the misidentification of students with “some risk” for MD at the time of pretest (i.e., false positives) may have increased the size of our analytic sample. However, a consistent educational finding is that a considerable number of U.S. students enter kindergarten lacking a level of number sense knowledge required for early success in mathematics (Duncan et al., 2007; Morgan et al., 2009). As such, we feel confident that the observed pretest results were an accurate representation of students’ mathematical performance at the start of the kindergarten year.

Similar to cutoffs used in existing research (e.g., Morgan et al., 2009), we operationalized severe MD as an initial performance less than the 10th percentile on a nationally recognized mathematics outcome measure. However, to provide teachers with actionable recommendations for improving the effectiveness of their core mathematics instruction, we, as a field, need to agree upon what constitutes categories of MD, particularly severe MD. Fields such as the medical field have established cutoffs for conditions as varied as hypertension, obesity, diabetes, and thyroid disease (Nettina, 2019). Agreed-upon cutoffs of MD would increase consistency in the way researchers and educators conceptualize, operationalize, and provide needed supports in early mathematics. Moreover, they would help improve methodologies for investigating rates of improvement in mathematics, such as whether kindergarten students with severe MD transition to a less at-risk category of MD across the school year. Relatedly, our study focused exclusively on immediate outcomes and thus did not include a longitudinal follow-up assessment (Watts et al., 2019). Fadeout effects are a common finding of intervention research in the field of education (Bailey et al., 2017). As such, future studies are required to determine whether explicit, core mathematics programs produce lasting impacts on the long-term learning trajectories of students with severe MD.

Our study also reported moderate levels of implementation fidelity. While these fidelity levels are comparable to those reported in other large-scale, cluster-randomized controlled trials (Smith et al., 2016; Wanzek et al., 2014), they may have affected this study’s results. For example, stronger implementation fidelity may have revealed greater treatment effects and higher acceleration of student mathematics performance. Conversely, teachers may have elected to skip particular aspects of ELM because of a lack of fit to their teaching style or the needs of their students. Regardless of the reason, stronger efforts are needed to improve implementation fidelity of ELM in future research. The current study also lacked information on the number of ELM lessons delivered (i.e., dosage). While research staff had active presence in participating classrooms throughout the school years, it is unclear whether students received the prescribed dose of core mathematics instruction.

In addition, we have no knowledge of whether students, both treatment and control, received intervention supports in mathematics beyond the instruction delivered in their core educational settings. While we cannot rule out if such supplemental supports existed in the present study or whether they had any influencing role on our findings, it is important to note the larger ELM Efficacy Trial (Clarke et al., 2015) purposefully recruited schools that typically did not offer supplemental mathematics instruction in kindergarten to accommodate research aims planned for later in the project. Research also suggests that a relatively small amount of time during the school day is allocated to mathematics instruction in kindergarten classrooms (La Paro et al., 2009). Furthermore, relative to early literacy instruction, schools are less likely to have in place the same types of support in beginning mathematics, particularly at the kindergarten level. Such supports include validated intervention materials, measurement systems to screen and progress monitor at-risk students, professional development to improve current practices, and implementation support provided by specialists or coaches. Thus, the likelihood is high that students in the current study received no supplemental instruction beyond Tier 1.

Finally, the current study could draw professional scrutiny, given its investigation of student outcome data from a previously published study. To mitigate these criticisms (American Psychological Association, 2020; Drotar, 2010), we contend there are several compelling, scientific reasons that justify the present study, despite its overlap with our prior work (Clarke et al., 2015). First, the space constraints of scholarly journals often restrict research teams from integrating the full myriad of a priori research hypotheses into an original article. We encountered this same issue in Clarke et al. (2015) and believe our peers in the education science field have faced similar challenges. Second, and above all, relative to our original work (Clarke et al., 2015), we contend the current study has a distinct purpose that targets unique and important research questions. The present study also addresses several blank spots in the literature. There is little empirical evidence on whether and to what extent students who are clearly at significant risk for long-term MD at the start of their schooling respond to previously validated mathematics programs. Moreover, few large-scale studies in the area of mathematics intervention research, including our own, have explored classroom-level tipping points and whether validated core mathematics programs can accelerate learning gains for at-risk students beyond a threshold of severe MD. Taken together, we contend the present study makes a distinct and important scientific contribution to the knowledge base over and above our previous work.

### Implications for Practice

Relative to other fields, such as climate science (Lenton et al., 2008; Russill & Nyssa, 2009), investigations of tipping points in education are still in its infancy. However, recent research has begun to unveil classroom-level variables that contribute to the point at which validated core programs begin to produce diminishing returns (Vaughn et al., 2017). While the current study did not reveal a tipping point for ELM based on the class-level proportion of students at risk for severe MD, we encourage researchers to continue this line of work as it has important implications for how schools can maximize the effectiveness of evidence-based, core mathematics programs.

For example, consider Program A, a core kindergarten mathematics program that has a solid evidentiary base yet demonstrates a tipping point once the classroom proportion of students at risk for severe MD exceeds 20%. Theoretically, under an MTSS model (Fuchs & Vaughn, 2012), a school would use a Tier 2 intervention to supplement Program A. However, extra resources are often scarce, and consequently, many schools lack the support systems needed to boost the effectiveness of core mathematics instruction. Therefore, schools need cost-effective ways to thwart the point at which the impact of validated core mathematics programs, such as Program A, begins to diminish. Below, we briefly make two practical recommendations.

One approach is to monitor fidelity of implementation of core mathematics programs as effective program delivery is a plausible way to mitigate the diminishing returns of evidence-based programs. To fit local contexts, teachers will likely adapt validated core mathematics programs. While fidelity and program adaptability often co-occur (Durlak & DuPre, 2008), such adaptations or modifications may affect program effectiveness. As such, some measure that documents the preservation of the essential components of core mathematics programs, such as Program A, is likely necessary. A second recommendation is for schools to examine the procedures they use to assemble students in classrooms. When administrators and teachers are deciding which students should be placed in particular classrooms, as much information as possible about students' prior and current performance in mathematics should be gathered. Such information may prove invaluable for maximizing the effectiveness of validated core mathematics programs.

## Conclusion

The research suggesting that a considerable number of U.S. students face severe MD in kindergarten is clear and compelling. In light of this evidence, the average kindergarten classroom likely has little support for students' mathematics development beyond the core curriculum. Given that core mathematics instruction delivered in kindergarten classrooms may represent the totality of a student's exposure to foundational mathematics content, it is therefore critical that it positively affects the development of mathematical proficiency of all students, including those who face severe MD. We believe that core mathematics programs that are purposefully designed to meet the instructional needs of the full range of learners, such as those programs that incorporate explicit and systematic design principles, serve as a valuable first line of defense in affecting the mathematics achievement of students with severe MD.

**Authors' Note**

Scott K. Baker is no longer affiliated with Southern Methodist University, Dallas, TX, USA.

**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: Drs. Ben Clarke, Hank Fien, and Scott K. Baker are eligible to receive a portion of royalties from the University of Oregon's distribution and licensing of certain Early Learning in Mathematics (ELM)-based works. Potential conflicts of interest are managed through the University of Oregon's Research Compliance Services.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080699 awarded to the Center on Teaching and Learning at the University of Oregon.

**ORCID iD**

Georgia L. Kimmel https://orcid.org/0000-0001-7649-3073

### References

Agodini, R., & Harris, B. (2010). An experimental evaluation of four elementary school math curricula. *Journal of Research on Educational Effectiveness*, *3*, 199–253. https://doi.org/10.1080/19345741003770693

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). https://doi.org/10.1037/0000165-000

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*, 7–39. https://doi.org/10.1080/19345747.2016.1232459

Barnes, M. A., Klein, A., Swank, P., Starkey, P., McCandliss, B., Flynn, K., Zucker, T., Huang, C. W., Fall, A. M., & Roberts, G. (2016). Effects of tutorial interventions in mathematics and attention for low-performing preschool children. *Journal of Research on Educational Effectiveness*, *1*, 155–178. https://doi.org/10.1080/19345747.2016.1191575

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

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

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*, 1–29.

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

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.

Clarke, B., Smolkowski, K., Baker, S. K., Fien, H., Doabler, C. T., & Chard, D. J. (2011). The impact of a comprehensive Tier 1 kindergarten program on the achievement of students at-risk in mathematics. *The Elementary School Journal*, *111*, 561–584.

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. http://www.corestandards.org/Math/

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L., Cho, E., Barquero, L. A., & Bouton, B. (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. https://doi.org/10.1177/0022219412442151

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

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

Doabler, C. T., Gearin, B., Baker, S., Stoolmiller, M., Kennedy, P., Clarke, B., Nelson, N. J., Fien, H., & Smolkowski, K. (2019). Student practice opportunities in core mathematics instruction: Exploring for a goldilocks effect for kindergartners with mathematics difficulties. *Journal of Learning Disabilities*, *52*, 271–283. https://doi.org/10.1177/0022219418823708

Doabler, C. T., Nelson-Walker, N., Kosty, D., Fien, H., Baker, S. K., Smolkowski, K., & Clarke, B. (2014). Examining teachers' use of evidence-based practices during core mathematics instruction. *Assessment for Effective Intervention*, *39*, 99–111.

Doabler, C. T., Smith, J. L., Nelson, N., Clarke, B., Berg, T., & Fien, H. (2018). A guide for evaluating the mathematics programs used by special education teachers. *Intervention in School and Clinic*, *54*, 97–105. https://doi.org/10.1177/1053451218765253

Drotar, D. (2010). Editorial: Guidance for submission and review of multiple publications derived from the same study. *Journal of Pediatric Psychology*, *35*(3), 225–230. https://doi.org/10.1093/jpepsy/jsp134

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

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350. https://doi.org/10.1007/s10464-008-9165-0

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention* (2nd ed.). Guilford Press.

Foster, E. M., & Bickman, L. (1996). An evaluator's guide to detecting attrition problems. *Evaluation Review*, *20*, 695–723.

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

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early conceptual understanding of cardinality predicts superior school-entry number-system knowledge. *Psychological Science*, *29*, 191–205. https://doi.org/10.1177/0956797617729817

Gersten, R. M., Chard, D., 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

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability—Third Edition (TEMA-3)*. PRO-ED.

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

Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology*, *78*(1), 119–128.

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*(3), 338–352.

Institute of Medicine. (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5*. The National Academies Press. https://doi.org/10.17226/12999

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early Math Matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*, 850–867. https://doi.org/10.1037/a0014939

La Paro, K. M., Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., Clifford, R., Oscar, B., Howes, C., & Burchinal, M. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms.

*Early Education and Development*, *20*, 657–692. https://doi.org/10.1080/10409280802541965

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schnellnhuber, H. J. (2008). Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 1786–1793.

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

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/10.3102/0162373714536608

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

National Science & Technology Council, Committee on STEM Education. (2018). *Chartering a course for success: America's strategy for STEM education*. White House.

Nettina, S. M. (2019). *Lippincott manual of nursing practice* (11th ed.). Wolters Kluwer Health.

Russill, C., & Nyssa, Z. (2009). The tipping point trend in climate change communication. *Global Environmental Change*, *19*, 336–344.

SAS Institute Inc. (2016). *SAS/STAT® 14.2 user's guide*. 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.

Smith, J., Smolkowski, K., Baker, S., Fien, H., & Kosty, D. (2016). Examining the efficacy of a multitiered intervention for at-risk readers in grade 1. *The Elementary School Journal*, *116*, 549–573. https://doi.org/10.1086/686249

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

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

Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A. M. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology*, *109*, 22–34. https://doi.org/10.1037/edu0000069

Wanzek, J., Vaughn, S., Kent, S. C., Swanson, E. A., Roberts, G., Haynes, M., & Solis, M. (2014). The effects of team-based learning on social studies knowledge acquisition in high school. *Journal of Research on Educational Effectiveness*, *7*, 183–204. https://doi.org/10.1080/19345747.2013.836765

Watts, T. W., Bailey, D. H., & Li, C. (2019). Aiming further: Addressing the need for high-quality longitudinal research in education. *Journal of Research on Educational Effectiveness*, *12*, 648–658. https://doi.org/10.1080/19345747.2019.1644692

What Works Clearinghouse. (2020). *Procedures handbook* (Version 4.1). Institute of Education Sciences, U.S. Department of Education.

What Works Clearinghouse. (n.d.). *What works in math?* https://ies.ed.gov/ncee/wwc/FWW/Results?filters=,Math